In This Issue:

In This Issue

Perspectives

Editorials:

Robert E. Becker
PTSD: A Disorder and a Reaction

Reviews and Overviews:

David L. Penn, Evan J. Waldheter, Diana O. Perkins, Kim T. Mueser, and Jeffrey A. Lieberman
Psychosocial Treatment for First-Episode Psychosis: A Research Update
Robyn Honea, Tim J. Crow, Dick Passingham, and Clare E. Mackay
Regional Deficits in Brain Volume in Schizophrenia: A Meta-Analysis of Voxel-Based Morphometry Studies

Introspections:

Edmund C. Levin
Messages From "Fran"
Images in Psychiatry:

Rajko Igic
Anton Pavlovich Chekhov (1860–1904)

New Research

Articles:

Christopher G. Fairburn, Zafra Cooper, Helen A. Doll, and Beverley A. Davies
Identifying Dieters Who Will Develop an Eating Disorder: A Prospective, Population-Based Study

Association Between the Insulin-Like Growth Factor 2 Gene (IGF2) and Scores on the Eating Attitudes Test in Nonclinical Subjects: A Family-Based Study

Pamela K. Keel, David J. Dorer, Debra L. Franko, Safia C. Jackson, and David B. Herzog
Postremission Predictors of Relapse in Women With Eating Disorders

Joanna Holliday, Kate Tchanturia, Sabine Landau, David Collier, and Janet Treasure
Is Impaired Set-Shifting an Endophenotype of Anorexia Nervosa?

Markus Heinrichs, Dieter Wagner, Walter Schoch, Leila M. Soravia, Dirk H. Hellhammer, and Ulrike Ehleert
Predicting Posttraumatic Stress Symptoms From Pretraumatic Risk Factors: A 2-Year Prospective Follow-Up Study in Firefighters

Rael D. Strous, Mordechai Weiss, Irit Felsen, Boris Finkel, Yuval Melamed, Avraham Bleich, Moshe Kotler, and Dori Laub
Video Testimony of Long-Term Hospitalized Psychiatrically Ill Holocaust Survivors

John Briere, Catherine Scott, and Frank Weathers
Peritraumatic and Persistent Dissociation in the Presumed Etiology of PTSD

Armen K. Goenjian, David Walling, Alan M. Steinberg, Ida Karayan, Louis M. Najarian, and Robert Pynoos
A Prospective Study of Posttraumatic Stress and Depressive Reactions Among Treated and Untreated Adolescents 5 Years After a Catastrophic Disaster
Zahava Solomon, Rami Shklar, and Mario Mikulincer

**Frontline Treatment of Combat Stress Reaction: A 20-Year Longitudinal Evaluation Study**

Chiharu Tamagaki, Göran C. Sedvall, Erik G. Jönsson, Gaku Okugawa, Håkan Hall, Stefan Pauli, and Ingrid Agartz

**Altered White Matter/Gray Matter Proportions in the Striatum of Patients With Schizophrenia: A Volumetric MRI Study**

Beth E. Snitz, Angus MacDonald, III, Jonathan D. Cohen, Raymond Y. Cho, Theresa Becker, and Cameron S. Carter

**Lateral and Medial Hypofrontality in First-Episode Schizophrenia: Functional Activity in a Medication-Naive State and Effects of Short-Term Atypical Antipsychotic Treatment**

Jonathan K. Wynn, Gregory A. Light, Bruno Breitmeyer, Keith H. Nuechterlein, and Michael F. Green

**Event-Related Gamma Activity in Schizophrenia Patients During a Visual Backward-Masking Task**

Silke Bachmann, Christina Bottmer, and Johannes Schröder

**Neurological Soft Signs in First-Episode Schizophrenia: A Follow-Up Study**

Dace S. Svikis, Nathan Berger, Nancy A. Haug, and Roland R. Griffiths

**Caffeine Dependence in Combination With a Family History of Alcoholism as a Predictor of Continued Use of Caffeine During Pregnancy**

Rajesh Narendran, W. Gordon Frankle, Richard Keefe, Roberto Gil, Diana Martinez, Mark Slifstein, Lawrence S. Kegeles, Peter S. Talbot, Yiyun Huang, Dah-Ren Hwang, Leyla Khenissi, Thomas B. Cooper, Marc Laruelle, and Anissa Abi-Dargham

**Altered Prefrontal Dopaminergic Function in Chronic Recreational Ketamine Users**

Heather A. Berlin, Edmund T. Rolls, and Susan D. Iversen

**Borderline Personality Disorder, Impulsivity, and the Orbitofrontal Cortex**

---

**Brief Reports:**

Molly G. Willer, Paul Thuras, and Scott J. Crow

**Implications of the Changing Use of Hospitalization to Treat Anorexia Nervosa**
Andreas Ströhle, Christian Feller, Marlies Onken, Frank Godemann, Andreas Heinz, and Fernando Dimeo
The Acute Antipanic Activity of Aerobic Exercise

Sanjay J. Mathew, Jonathan M. Amiel, Jeremy D. Coplan, Heidi A. Fitterling, Harold A. Sackeim, and Jack M. Gorman
Open-Label Trial of Riluzole in Generalized Anxiety Disorder

Johannes Hamann, Rudolf Cohen, Stefan Leucht, Raymonde Busch, and Werner Kissling
Do Patients With Schizophrenia Wish to Be Involved in Decisions About Their Medical Treatment?

Sukhwinder S. Shergill, Gabrielle Samson, Paul M. Bays, Chris D. Frith, and Daniel M. Wolpert
Evidence for Sensory Prediction Deficits in Schizophrenia

Stuart W. Twemlow and Peter Fonagy
The Prevalence of Teachers Who Bully Students in Schools With Differing Levels of Behavioral Problems

Palaniappan Muthappan, Heidi Forster, and David Wendler
Research Advance Directives: Protection or Obstacle?

Communications and Updates

Letters to the Editor:

LAURENCE BORRAS, E.L. CONSTANT, A. EYTAN, and PHILIPPE HUGUELET
Hypertension and Aripiprazole

LOUAI BILAL, CARLOINE TSAI, JAMES J. GASPER, and J. CHARLES NDLELA
Parkinsonism With Intramuscular Ziprasidone

MARLOES G. POSTEL, COR A.J. de JONG, and HEIN A. de HAAN
Does E-Therapy for Problem Drinking Reach Hidden Populations?

GARY REMINGTON and SHITIJ KAPUR
Remission: What’s in a Name?

NANCY C. ANDREASEN
Dr. Andreasen Replies

EMMANUELLE CORRUBLE and PATRICK HARDY
Why the Hamilton Depression Rating Scale Endures
RASMUS W. LICHT and PER BECH
Why the Hamilton Depression Rating Scale Endures

CHING-LIN HSIEH and CHENG-HIS HSIEH
Why the Hamilton Depression Rating Scale Endures

BERNARD J. CARROLL
Why the Hamilton Depression Rating Scale Endures

PER BECH, NINA ENGELHARDT, KENNETH R. EVANS, MICHAEL GIBERTINI,
AMIR H. KALALI, KENNETH A. KOBAK, JOSHUA D. LIPSITZ, JANET B.W.
WILLIAMS, JAY D. PEARSON, and MARGARET ROTHMAN
Why the Hamilton Depression Rating Scale Endures

R. MICHAEL BAGBY, DEBORAH R. SCHULLER, ANDREW G. RYDER, and
MARGARITA B. MARSHALL
Dr. Bagby and Colleagues Reply

NATHALIE T. GODART, ZOÉ REIN, FABIENNE PERDEREAU, FLORENCE
CURT, and PHILIPPE JEAMMET
Predictors of Premature Termination of Anorexia Nervosa Treatment

D. BLAKE WOODSIDE
Dr. Woodside Replies

PAUL MACKIN and PETER GALLAGHER
Brain-Derived Neurotrophic Factor in Patients With Remitted Depression

ALEXANDER NEUMEISTER and HUSSEINI K. MANJI
Drs. Neumeister and Manji Reply

CARLA de BRUIJN and WIM van den BRINK
Validity of the One-Criterion Threshold for an Alcohol Abuse Diagnosis

MARC A. SCHUCKIT
Dr. Schuckit Replies

CHRISTOPHER G. IVANY
Explanatory Pluralism and Patchy Reductionism

WILLIAM G. CAMPBELL and KEITH I. PEARCE
Explanatory Pluralism and Patchy Reductionism

KENNETH S. KENDLER
Dr. Kendler Replies
Book Forum:

ARNOLD WERNER  
American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer  

WILLIAM A. FROSCH  
Frank Lloyd Wright  

ROBERT MICHELS  
In My Brother’s Shadow: A Life and Death in the SS  

DAVID V. FORREST  
A Sweeper Up After Artists: A Memoir  

PAUL ROAazen  
A Compulsion for Antiquity: Freud and the Ancient World  

ARMANDO R. FAVAZZA  
I’m Your Father, Boy: A Family Memoir of Barbados  

RODERIC GORNEY  
My Life So Far  

PETER D. KRAMER  
Last Night: Stories • The Dog of the Marriage: Stories  

JUSTIN SIMON  
Amador  

GLEN O. GABBARD  
Skinny Dip  

JAMES R. MERIKANGAS  
Transmission: A Novel  

DINAH MILLER  
The Interview Room: A Novel  

ALAN A. STONE  
The Story of the Stone (The Dream of the Red Chamber, vol. 1: The Golden Days)  

WILLIAM EDWIN FANN  
Stalin: In the Court of the Red Tsar  
DAVID V. FORREST

American Mania: When More Is Not Enough

RICHARD D. CHESSICK

History of Italian Renaissance Art, 5th ed.
Predicting Eating Disorders

Many people diet, but few develop eating disorders. To learn who is most at risk, Fairburn et al. (p. 2249) surveyed 2,992 young women who were dieting and contacted them again four times over the next 2 years. They completed the same eating disorder questionnaire each time. Participants with possible anorexia nervosa, bulimia nervosa, or other eating disorder were interviewed in person. Diagnoses were substantiated for 104 of the subjects, or 3.5%. Participants with and without eating disorders were distinguished by baseline body mass index plus five baseline questionnaire items: frequency of binge eating, frequency of vomiting/laxative use, secret eating, fear of losing control over eating, and desire to have an empty stomach. A simple screening instrument assessing these variables in the original group of dieters would have identified about 70% of the subjects who later developed eating disorders. Incorporating this instrument into routine health assessments might help identify high-risk dieters.

What Sparks PTSD in Firefighters?

It is difficult to determine predisposing factors for posttraumatic stress disorder (PTSD) in retrospective studies because the illness can influence patients’ perception or recollection of pretrauma characteristics. Heinrichs et al. (p. 2276) circumvented this problem by testing healthy probationary firefighters just after basic training and again 6, 9, 12, and 24 months later. At 24 months, 16% of the men met the criteria for PTSD. No psychological or biological variable except body weight changed significantly over time. However, high scores for PTSD symptoms at 2 years were predicted by greater baseline hostility and lower self-efficacy, i.e., self-perceived competence. Firefighters with these two high-risk characteristics at job entry showed a general increase in all psychological symptoms (e.g., PTSD, depression, anxiety, alexithymia) over the 2 years. Hostility and self-efficacy may therefore be informative measures in populations at high risk for trauma-related disorders and could be addressed through coping skills training.

Looking Back at Frontline Treatment

Treatment of acute combat stress reaction on the front lines is preferred by many armies. In the 1982 Lebanon War, Israeli soldiers with combat stress reaction who were treated on the front lines were more likely to return to duty and less likely to develop posttraumatic stress disorder (PTSD) than were soldiers treated back in Israel. How enduring were these benefits? Solomon et al. (p. 2309) administered a questionnaire to the soldiers 20 years after the war. The 79 veterans who had received frontline treatment had a PTSD rate of 30%, compared with 41% among those treated in the rear echelon and 14% among veterans without combat stress reaction. Those who received frontline treatment also had fewer other psychiatric symptoms, less loneliness, and better interpersonal functioning than those treated in the rear. The benefits of frontline treatment for acute stress reaction thus extend far into soldiers’ futures.

Low Gamma Waves in Schizophrenia

Perception of visual forms requires integration of basic visual elements. Oscillations in the gamma frequency range (30–70 Hz) seem to be key to the brain’s ability to bind visual features together. Gamma abnormalities have been found in patients with schizophrenia and could contribute to their perceptual aberrations. Wynn et al. (p. 2330) measured gamma activity in the brains of schizophrenia patients and healthy participants as they identified visual targets with and without visual masks (acting as distractors) shown immediately afterward. Without these masks, the patients did not differ significantly from the comparison subjects in correct identifications or gamma activity. With masking, however, the patients scored lower, produced less gamma activity, and did not show predominance of the right hemisphere. As the right hemisphere is preferentially responsible for processing ambiguous visual information, this deficit seems an important clue to the neurophysiological abnormalities underlying schizophrenia.

Similarities of Borderline Personality to Orbitofrontal Brain Injury

Damage to the orbitofrontal cortex can result in loss of inhibition. Impulsivity is also a characteristic of people with borderline personality disorder, who have shown low orbitofrontal metabolism and volume in some studies. To clarify possible orbitofrontal involvement in borderline personality disorder, Berlin et al. (p. 2360) administered psychological and cognitive tests to patients with borderline personality disorder, patients with orbitofrontal damage, comparison patients with damage to other parts of the prefrontal cortex, and healthy subjects. The patients with borderline personality disorder were more emotional and neurotic, and less extroverted and conscientious, than all other groups. However, they were similar to patients with orbitofrontal damage in having more impulsivity and anger, less happiness, and more inappropriate behaviors than the other two groups. The possible involvement of the orbitofrontal cortex in some deficits in borderline personality disorder may suggest new treatment approaches.
PTSD: A Disorder and a Reaction

“I have done that,” says my memory.
“I could not have done that,” says my pride, and remains inexorable.
Finally my memory yields.

—F. Nietzsche (1)

When I awoke in bed at King County Hospital in Seattle, I assured myself that I was just dreaming. It was my fourth month of internship. One of my patients, a chronic alcoholic with liver failure, had persistent bleeding from enlarged esophageal veins. In spite of our best efforts, his condition was progressively worsening. I awoke upset from dreaming that had I taken his pillow and smothered him. My thought was that we needed the bed for another patient. Even today, this memory remains disturbingly real. I have no grounds to suggest that I contributed to his later death, yet there is still present a sense of failure that I need not dispel. This dream provides a shorthand reminder of ways I might fail in medicine.

In retrospect, I would have been diagnosed with posttraumatic stress disorder (PTSD); however, at that time, all psychiatric diagnoses were reactions applied to people grouped by similarities but not typed with disorders. I did not have far to look for sources for my reactions, which I still find of more concern than the syndrome I developed.

Eighteen months before this dream, after completing my third year of medical school, I accepted a position as summer “extern” at a rural community hospital in Massachusetts. It was my first hands-on experience of medicine outside the shelter of school. As was common in the 1950s, the hospital doctors were only available on call from home; therefore, because I was present when the need arose, I delivered babies, managed patients with heart attacks, provided the first interventions for serious accident victims brought in from the nearby Massachusetts Turnpike, and tried to help until a doctor arrived. Most were rewarding experiences. One woman whose baby I delivered as she arrived at the emergency room named the baby after me. I looked forward to being qualified as a doctor.

One night the nursing supervisor who managed the emergency room whenever there was a patient called me to see a young woman. She had been run over by her boyfriend’s car after she lay in its path to keep him from leaving her. She was in shock from internal bleeding. There were tire marks across her abdomen, running from just under her ribs on the right side over her pelvis on the left. In moments such as this, we do more things at once than we can write about. I drew a blood sample as I started an intravenous infusion to replace her lost blood volume. I asked if the doctor on call had been asked to come in—he had—and the surgeon as backup. No, I was told. Only a doctor could call a surgeon. We agreed to ask the surgeon to come in. The nurse also agreed to call the laboratory technician so we could cross-match the blood we had on hand, the X-ray technician, and staff for the operating room. She agreed to do these things while I worked with the patient. Shortly afterward, the doctor and surgeon arrived, examined the patient, who seemed to be maintaining her breathing and blood pressure sufficiently that she would reach the operating room alive, and then withdrew, asking me to continue what I was doing for the patient.
Time passed—all too much time, it seemed—without blood being delivered to the patient, the X-ray technician appearing, or for the preparations to take the patient to surgery to begin. I asked the nursing supervisor where the blood was. “The doctors canceled the orders and had me call everyone and tell them it wasn’t necessary for them to come in.” I struggled to find words to respond. At the time, I sensed that we both struggled to corral our feelings and thoughts. This situation seemed an impossible break with what I had just learned in 3 years of studying medicine. I remember asking the nurse to stay with the still-conscious patient so I could go and talk to the doctors. I found them in an adjacent room and asked, “What will we do?”

“We don’t want to treat this kind of patient here….It won’t be necessary for us to get involved….There is nothing we can do….You should stay with the patient to see what happens. There’s no point in sending her on to Worcester; she won’t make it.” All of these phrases have never left my mind since that night.

I saw no alternative except to do as I was instructed. I did not say, “We can try,” but I still hear the words in my mind when I think of those moments. “We can try,” seems important in medicine. At the time I spoke silently to myself, “I am a student. They certainly know more than I do about these situations.” Saying this was neither satisfying nor reassuring. I went back to the patient and kept the fluids going. In about 30 minutes, she became unconscious and then died. After she was no longer breathing and there were no heart sounds, I asked the nurse to call a doctor to pronounce her legally dead.

The state medical examiner agreed that I could attend the autopsy. We found a 3-inch laceration in her liver but only minor bleeding, one fracture posterior and one anterior in the left pelvis that had been drawn closed by muscular contractions, and a tear in the internal iliac artery over the posterior fracture site. The patient died from blood loss into the abdomen through the torn artery. “Why didn’t they open her abdomen and sew this up?” the medical examiner asked me. “I don’t know,” I said, relating what I was told and defensively noting that I was a student.

In many ways, I felt responsible, even though no responsibility was suggested. I wanted to shed tears, but I would not allow myself to do so. Tears, I thought, are not in a doctor’s formulary. Years later, I understood differently when another resident told me how Yale Chief of Medicine Paul Beeson had wept when he could not save his resident, a young doctor, from a drug-resistant infection.

In the shadow of this summer, my fourth year of medical school was less clear for me than the others had been. To avoid some clinics, I took an afternoon job teaching an organic chemistry laboratory at Roger Williams College in Montreal, went through the motions, watched my class ranking fall, and went to my internship. I avoided lessons in clinical medicine that I did not want to learn.

For the first 2 months of my internship, on a psychiatry rotation at the University of Washington’s University Hospital, I talked with patients and teachers and rediscovered a place for my interests in patients and the diseases that affect them. I spoke to no one of the previous summer. Nor do I remember those events to be of great significance years later in my personal analysis, perhaps because there are no memories to recover and no doubts to resolve. The events and my reactions remain too clear. My conflicts draw on deeper personal roots that I followed in analysis, and I still view my reactions as aversions to failings in moral courage, medical skills, medical training, and what it means to be a physician that I witnessed that day and did not wish to see again. Whatever psychiatric disorder might have been present in me seems shallow and out of focus compared to the circumstances I chose not to abide.

In the third month of internship, I was back in trouble. The internal medicine resident who supervised my work found me lacking. Unfortunately, the feelings were polarizing and reciprocal, so I mainly did his bidding while waiting for better days. He said one morning, “Get a stomach tube in that stroke in bed 3 and get her out of here this afternoon to a nursing home so we can get some better teaching cases.” That she was learn-
ing to walk and eat with her left hand did not matter; I passed the tube into her as she watched and gagged, unable to speak from her stroke. I made a call to social workers, she was shipped out, and the bed was free.

While taking orders, I think I released in myself a part of the medical training I hoped I could avoid. I did not realize the changes until my dream a few weeks later pointed out the realities. As Robert Petersdorf, the new chief of medicine at the university’s King County Hospital, reinforced for me as he questioned us in rounds: knowing the latest discoveries reflects well on one—a game I was adept at playing from medical school—while being distressed over how we use technology does not lead to a place in a residency program. I chose to arm myself with skills so that I would never have to stand by unprepared, but I feared myself all too willing to lose my soul for success. In those transition years from school to specialty training, memories, nightmares, loss of interest, avoidance, fear that I would fail in some ways and not in others, indecision, and distractions introduced to avoid conflict would today be diagnosed as PTSD, but they each remain dynamic interactions with events for me.

Later in my internship, I was back on an internal medicine ward at the University Hospital in Seattle. We had a patient who was losing blood into his gut through his intestinal walls. Testing showed no unaffected areas, surgeons saw no feasible interventions, and internists could reach no diagnosis that suggested a way to stop the blood loss. The blood bank notified us that at the current rate of use they would be out of suitable blood by midday Sunday. Today was Friday. We presented the situation to the attending physician, Wade Volweiler. He listened, went over the findings with us, and walked in to see the patient. He examined the patient and asked him to excuse us from the room. In the hall he said, “We have to tell the patient that there probably will be nothing more we can do but try. I would like to give him the decision if it is all right with you—unless one of you has some other ideas we should discuss.” We did not. We went back into the room.

Dr. Volweiler asked the patient if he might sit on the bed to talk with him. At that time, we were trained not to sit on the patient’s bed out of respect. This break with protocol signaled to me a different, informal, personal, shared moment in the relationship between doctor and patient. His sitting on the bed still expresses to me that someday each of us will be in that bed. Dr. Volweiler explained the situation and said gently but directly to the patient that he thought he would want to know that we saw little likelihood of his surviving, so the patient could arrange to speak with his family before the apparently inevitable moment of crisis. He pointed out that the patient had a choice: he could receive whatever blood was available, or he could choose a time to stop adding new blood and leave some blood for others who might need it. The patient thought for what seemed longer than it was then asked if he could arrange to be with his family Saturday morning and to stop the addition of blood at noon Saturday. I was off duty that weekend. When I came in Monday morning, a new patient was in his bed. I still feel pain, tears, an unreasonable senselessness, helplessness, loss, and injury to unreasonable pride or expectation for what we might do as doctors, but I mainly respected this patient’s courage and dignity—each a reaction to this traumatic stress. I also feel a debt for a gift that made my life uniquely richer: I learned that even at the limits of science, in extremes of suffering and sorrow, we can be decent with each other and with ourselves, and I saw how to be different. Years later, sadly, when I stated to a chief of medicine who had trained at the University of Washington how I owed a debt to Wade Volweiler for showing me what it means to be a physician, he described my mentor’s role in later years as not mainstream and marginalized to his office. I thought it sad that this chief could not share my perceptions.

Ten years ago, I read a story about William Osler, the much-venerated model for physicians during my training at McGill University. During our medical school years, W.W. Francis had shared many personal memories of William but not this story, perhaps because it occurred in Oxford. William was asked to see a young child with diphtheria who
was not eating and seemed headed toward death. William was preparing for the graduation ceremony at Oxford, so he put on his academic robes and stopped to see the child in this dramatic regalia. The child was delighted at the sight of this colorful doctor. After completing his assessment of the child’s condition, William ordered some soft food and fed it spoonful by spoonful to the child. After the child finished eating, William went to graduation. Each of the following days, William called on the child in the same academic robes and continued to feed the child until the diphtheria inflammation resolved sufficiently that the child ate on his own.

In his earlier years, William made medicine at McGill University scientific but also left us a legacy of how to work with people. In this Hippocratic tradition, people do not become invisible in the presence of disease. I doubt that technology will ever replace the efficacies of some Hippocratic traditions of practice or at least of William’s embodiments of them. He combined scholarship and knowledge with concern for the person as the conditions of practice. For me, this means somehow understanding PTSD as both a disorder and a reaction.

Events of our times reawaken these old conflicts of disorder with reaction for me but with each way of thinking valid only in context with the other. I feel I am back in that emergency room some 50 years ago when I hear the words of soldiers in Iraq who are labeled by the media, with medicine’s approval, as having PTSD because they exclaim, “I don’t want to die here!” and panic, cannot forget shooting someone who did not stop when approaching a checkpoint, or cannot stop thinking about saving colleagues who seem certain to die. I sense in them what I found in myself and many times over in practice—a disorder only at the conclusion of an out-of-control cascade begun by natural human reactions. No one wants patients to die when medicine could try to save them; no one wants to die in Iraq. Before I bury the complexities of many individuals’ diverse reactions behind labels of disorders, I want to ask each person one-on-one if he or she believes each of us will want to take every measure to survive and help others survive. Could it be that this soldier thought himself—or was actually caught up—in circumstances in which others betrayed his trust in them? Did he—unlike that patient who waited for help that never came to the emergency department—flee the helpless inevitability imposed by others who had already fled their personal responsibilities for the anonymous safety of roles and “we”? Before I fixate on disorders, I want to discuss with the person about to be labeled or to be self-labeled, whether by placing others in positions where they have no choices, do not leaders incur responsibility for what happens on their watch if others follow orders? If these words do not reassure a soldier who follows orders and commits no crime, I expect I would stridently argue with him or her that a soldier protects a checkpoint as his duty, much as we physicians occupy our stations with senses of duty. The mistakes that ensue from mandated procedures are the failings of planners. How unjust I found labels of disorder to be in the past when simple inquiry revealed facts to be otherwise and vastly more complex. Why should I expect these problems of diagnosis to be any different now?

Public policy labels the many different human stories and reactions of veterans as disorders with alacrity. This labeling serves systems willing to offer only waits of many months to see doctors who then have time only for another meeting months later and only drugs—not themselves—to fill the gaps. In my years of practice, I never saw one disorder or type. Perhaps because I had the time to inquire, I saw many different people with many different reactions and many different routes to healing as best they could the memories of personal traumas. Psychiatry’s now-out-of-fashion reactive diagnoses stayed alive for me over these years as reminders that diversity, not uniformity, characterizes the lives of organisms. Now government policies encourage us to imprint lives as disordered and to fail veterans with the conveniences we find by homogenizing diversity into disorders. Thousands of different stories become one diagnosis. I fear that psychia-
try conspires in this because over the last 50 years, overcommitted to the homogenizing perspectives of disorders, we have lost our grasp on how to understand an individual.

Only biochemicals have predictable and regular disorders. Only people have diversely individualized diseases. Only biochemicals show unvarying patterns in their reactions. Only people show uniqueness both genetically and historically and provide the diversity that characterizes life through lives that dynamically build in time, never to repeat the past in exactly the same ways. This diversity displays how we as whole beings differ from our molecular parts.

Disorders imprint a lifeless uniformity on experience. Disorders are appropriate for derangements of molecules but meaningless for individuals. Today veterans are as much victims of labels and stereotypes as the patient in the emergency room or the stroke victim in bed 3 were. Then people disappeared behind stereotypes of “not wanted here”; now people disappear into a stereotyped “disorder.” We are physicians to both people with molecular disorders and unique individuals, just as Darwin made us biologists of molecular and whole-organism processes. Darwin found it necessary to think in new and different ways to capture how unique organisms differ from their molecules. Why can psychiatry not support us in these difficult balances of molecules and living? Disorders describe our molecular abnormalities, and reactions preserve us as individuals—always diverse, always personal, always uniquely encountering life, always needing to balance impersonal explanations with personal understandings.

Reference

1. Nietzsche F: Beyond good and evil, in Basic Writings of Nietzsche. Edited by Kaufmann W. New York, Modern Library, 1992, number 68

ROBERT E. BECKER, M.D.

Address correspondence and reprint requests to Dr. Becker, 2008 Mahre Dr., Park City, UT 84098; rebecker@xmission.com (e-mail).
Psychosocial Treatment for First-Episode Psychosis: A Research Update

David L. Penn, Ph.D.
Evan J. Waldheter, M.A.
Diana O. Perkins, M.D.
Kim T. Mueser, Ph.D.
Jeffrey A. Lieberman, M.D.

Objective: This article reviews research on psychosocial treatment for first-episode psychosis.

Method: PsycINFO and MEDLINE were systematically searched for studies that evaluated psychosocial interventions for first-episode psychosis.

Results: Comprehensive (i.e., multielement) treatment approaches show promise in reducing symptoms and hospital readmissions, as well as improving functional outcomes, although few rigorously controlled trials have been conducted. Individual cognitive behavior therapy has shown modest efficacy in reducing symptoms, assisting individuals in adjusting to their illness, and improving subjective quality of life, but it has shown minimal efficacy in reducing relapse. Some controlled research supports the benefits of family interventions, while less controlled research has evaluated group interventions.

Conclusions: Adjunctive psychosocial interventions early in psychosis may be beneficial across a variety of domains and can assist with symptomatic and functional recovery. More randomized, controlled trials are needed to evaluate the effectiveness of these interventions, particularly for multielement, group, and family treatments. 

Psychotic disorders, particularly schizophrenia, are the most disabling of all mental illnesses. In fact, schizophrenia is included among the world’s top 10 causes of disability-adjusted life-years (1). The majority of individuals with schizophrenia have a poor long-term outcome (2–4), which results in great personal suffering and societal cost. The largest expenditure for mental health in the United States is for schizophrenia (5), with an annual cost of $32.5 billion (6–8). Most of this cost can be attributed to repeated hospitalizations following relapse (9).

In an effort to improve the long-term outcome for individuals with schizophrenia, research has focused on early identification and intervention for psychosis. This approach to secondary prevention has been bolstered by findings that the sooner antipsychotic treatment is initiated after the emergence of psychosis, the better the clinical response (for example, see reference 10; see references 11 and 12 for reviews and references 13–16 for exceptions). In addition, most clinical and psychosocial deterioration in schizophrenia occurs within the first 5 years of the onset of the illness (11), suggesting that this is a critical period for treatment initiation (17, 18). Thus, pharmacological and psychosocial treatment delivered during this critical period has been hypothesized to have a stronger impact than comparable treatment provided later in the illness (17). Finally, there is a growing risk of treatment-resistant symptoms with each subsequent relapse (19–24). This is consistent with findings that show progressive loss of brain gray matter associated with recurrent episodes, suggesting that each relapse may reduce the individual’s capacity to respond to subsequent treatment (25, 26). Early intervention may therefore reduce the risk of relapse and long-term disability associated with chronic schizophrenia (27–29).

Pharmacological Treatment of First-Episode Psychosis

Most individuals with first-episode psychosis are responsive to antipsychotic medication (30). Remission of psychotic symptoms occurs in 50% of individuals with first-episode psychosis within the first 3 months after initiation of treatment with antipsychotic medication (24, 31, 32), 75% within the first 6 months (32), and up to 80% at 1 year (31, 33–35).

The beneficial effects of antipsychotic medication on first-episode psychosis are tempered by the following issues: 1) individuals with first-episode psychosis are particularly sensitive to the side effects of antipsychotics, such as weight gain (36, 37), 2) medication adherence is variable, with 6–12-month adherence rates in the 33%–50% range (38, 39), 3) up to 20% of individuals with first-episode psychosis show persistent psychotic symptoms (40), and 4) over 50% of individuals with first-episode psychosis report significant depression and/or anxiety secondary to the traumatic nature of psychosis (41–43).

In addition, despite initial symptom reduction, there is poor functional recovery following a first psychotic episode. Tohen et al. (32) found that although approximately 75% of individuals with first-episode psychosis showed symptom remission at 6 months, most (79.8%) failed to...
show functional recovery during the same time period (see also reference 35). Individuals with first-episode psychosis tend to have impairments in general social functioning (44, 45), quality of life (46, 47), and occupational functioning (48) despite clinical recovery. These functional impairments are present up to 5 years after illness onset even when optimal pharmacological treatment is provided (49).

**Psychosocial Interventions for First-Episode Psychosis**

Clearly, pharmacotherapy alone is not sufficient to prevent relapses or assure functional recovery from acute psychosis. Thus, there is a growing interest in psychosocial interventions as a means of facilitating recovery from an initial episode of psychosis and reducing the long-term disability associated with schizophrenia (50). Work in this area is still in its infancy, however. Treatment guidelines for first-episode psychosis, which include therapeutic engagement, targeting psychological and social adjustment, developing an active relapse prevention plan, and identifying barriers to treatment (42, 51, 52), are based on clinical experience and not controlled research evaluating standardized psychosocial programs. There is a need for updated guidelines, informed by a rigorous review of available research.

According to Edwards and colleagues (53–55), the literature on psychosocial interventions for first-episode psychosis can be conceptualized as constituting two broad categories: 1) studies evaluating comprehensive (i.e., multielement) interventions, which typically include community outreach/early detection efforts, in- and outpatient individual, group, and/or family therapy, and case management, in addition to pharmacological treatment, and 2) studies evaluating specific (i.e., single-element) psychosocial interventions (e.g., individual cognitive behavior therapy). In this article we review the extant literature on psychosocial interventions for early psychosis in light of these two categories.

**Search Strategy**

A comprehensive search of the PsycINFO and MEDLINE databases (January 1983 to October 2004) was conducted by using the following terms: 1) “first-episode schizophrenia” and “psychosocial treatment” (or “therapy or treatment”), 2) “first-episode psychosis” and “psychosocial treatment” (or “therapy or treatment”), and 3) “early psychosis” and “psychosocial treatment” (or “therapy or treatment”). The results were evaluated for relevance, and only the studies evaluating psychosocial interventions for first-episode psychosis were selected for review. Specifically, we selected papers that quantitatively evaluated multielement interventions, individual cognitive behavior and supportive therapy approaches, and group and family interventions. The designs of the studies reported in the selected articles included experimental/randomized-controlled (i.e., comparing outcomes in randomized groups), quasi-experimental (i.e., comparing outcomes in nonrandomized groups), and single-group (i.e., evaluating change over time in one group of individuals receiving treatment). Studies that compared subgroups of patients within a particular intervention or program (e.g., patients with short durations of untreated psychosis versus patients with long durations of untreated psychosis) were excluded. Finally, to ensure that our search was as comprehensive and current as possible, we also conducted independent searches for recent publications by leading psychosocial researchers in the field of early psychosis (e.g., Addington, Birchwood, Edwards, Jackson, Lewis, Linszen, Malla, McGorry, Morrisson, Tarrier). The findings of all of the selected studies are summarized in Table 2 (multielement studies) and Table 3 (single-element studies).

---

**TABLE 1. Characteristics of Selected Comprehensive (i.e., Multielement) Treatment Programs for Early Psychosis**

<table>
<thead>
<tr>
<th>Program</th>
<th>Intake Age Range (years)</th>
<th>In- and Outpatient Services</th>
<th>Atypical Antipsychotic Treatment*</th>
<th>Individual Cognitive Behavior and Supportive Therapy</th>
<th>Individual Supportive Therapy Only</th>
<th>Group Therapy</th>
<th>Family Therapy</th>
<th>Case Managementb</th>
<th>Community Outreach/Early Detection Efforts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early psychosis prevention and intervention centre, Melbourne, Victoria, Australia</td>
<td>15–25</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Prevention and early intervention program for psychosis, London, Ont., Canada</td>
<td>16–50</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Early psychosis program, Calgary, Alta., Canada</td>
<td>16–45</td>
<td>Outpatient only</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Early treatment and identification of psychosis project, Norway and Denmark</td>
<td>18–65</td>
<td>Outpatient only</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*All programs initially prescribe low doses of atypical antipsychotics as first-line pharmacological treatment.

bMost programs adhere to an assertive case management model, in which the case manager coordinates all treatment for the client, serves as primary contact for the program, and may also assist with vocational and/or housing needs.
PSYCHOSOCIAL TREATMENT

TABLE 2. Summary of Studies Evaluating the Effectiveness of Comprehensive (i.e., Multielement) Treatment for Early Psychosis

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Description</th>
<th>Design</th>
<th>Intervention</th>
<th>Comparison Group(s)</th>
<th>Follow-Up Period (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>McGorry et al., 1996 (56)</td>
<td>102</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Early Psychosis Prevention and Intervention Centre (EPPIC)</td>
<td>Before EPPIC (historical control)</td>
<td>12</td>
</tr>
<tr>
<td>Power et al., 1998 (57)</td>
<td>231 (longitudinal data on 120)</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Single group</td>
<td>EPPIC</td>
<td>Before EPPIC (historical control)</td>
<td>12</td>
</tr>
<tr>
<td>Carbone et al., 1999 (58)</td>
<td>250</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>EPPIC</td>
<td>Before EPPIC (historical control)</td>
<td>12</td>
</tr>
<tr>
<td>Malla et al., 2001 (59)</td>
<td>41</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Prevention and Early Intervention Program for Psychosis (PEPP)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Malla et al., 2002 (60)</td>
<td>85 (longitudinal data on 53)</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>PEPP</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Malla et al., 2002 (61)</td>
<td>66</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>PEPP</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Addington and Addington, 2001 (62)</td>
<td>93</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Calgary Early Psychosis Program</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Addington et al., 2003 (33)</td>
<td>180</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Calgary Early Psychosis Program</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Addington et al., 2003 (46)</td>
<td>177</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Calgary Early Psychosis Program</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Addington et al., 2004 (63)</td>
<td>238</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Calgary Early Psychosis Program</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Mintz et al., 2004 (64)</td>
<td>180</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Calgary Early Psychosis Program</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Larsen et al., 2001 (65)</td>
<td>109</td>
<td>Nonaffective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Early Treatment and Identification of Psychosis (TIPS) project</td>
<td>Before TIPS (historical control)</td>
<td>None</td>
</tr>
<tr>
<td>Cullberg et al., 2002 (66)</td>
<td>297</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Integrated treatment (Parachute Project)</td>
<td>Before Parachute Project (historical control); inpatient treatment as usual (prospective control)</td>
<td>12</td>
</tr>
<tr>
<td>Nordenoft et al., 2002 (67, 68)</td>
<td>341</td>
<td>Nonaffective first-episode psychosis</td>
<td>Randomized, controlled trial</td>
<td>Integrated treatment (OPUS Project)</td>
<td>Outpatient treatment as usual (prospective control)</td>
<td>12</td>
</tr>
</tbody>
</table>

a “Better” denotes that patients in the intervention program did significantly better than the comparison group(s) in studies with an experimental or quasi-experimental design or that there was significant improvement over time in studies with a single-group design. “No group differences” denotes no significant difference between the intervention and comparison groups in studies with an experimental or quasi-experimental design or that there was no change over time in studies with a single-group design.

b Nonaffective first-episode psychoses were schizophrenia spectrum disorders. Affective first-episode psychoses were mood disorders with psychotic features.

c The elements of the EPPIC, PEPP, Calgary, and TIPS interventions are shown in Table 1.

d Care before the multielement program typically consisted of standard inpatient services, limited outpatient services, limited emphasis on phase-specific psychosocial treatment, and limited outreach and early detection efforts.

e Measures were the Brief Psychiatric Rating Scale, the Scale for the Assessment of Positive Symptoms, and the Positive and Negative Syndrome Scale.

f Measures were the Scale for the Assessment of Negative Symptoms and the Positive and Negative Syndrome Scale.

g Measures were the Quality of Life Scale, the Wisconsin Quality of Life Index, and the Global Assessment of Functioning Scale.

Multielement Interventions

Multielement programs offer a comprehensive array of specialized in- and outpatient services designed for individuals experiencing first-episode psychosis, and they emphasize both symptomatic and functional recovery. Further, many of the issues that are particularly problematic among young individuals experiencing psychosis (e.g., substance abuse, suicidality, engagement with the mental health system) are addressed through a variety of targeted therapeutic approaches. Table 1 provides general information about several multielement programs and their respective components (for a full description of these and additional programs, see reference 55).

The Early Psychosis Prevention and Intervention Centre in Australia is an exemplar of multielement care for first-episode psychosis among young individuals experiencing psychosis with a focus on comprehensive, integrated, and evidence-based treatments.
episode psychosis and consists of a mobile assessment and treatment team; a 16-bed inpatient unit; in- and outpatient case management (including housing and vocational assistance); individual, group, and family therapy; pharmacological management (emphasizing low doses of atypical antipsychotic medication as first-line treatment); and specialized treatment for individuals with persistent psychotic symptoms. The Prevention and Early Intervention Program for Psychosis and the Calgary Early Psychosis Program are additional examples of established early intervention centers (55). There have also been several large-scale efforts to evaluate the effectiveness of multi-element treatment approaches for early psychosis delivered in the context of existing systems of care. For example, the Early Treatment and Identification of Psychosis project is a prospective, longitudinal 5-year study investig-

<table>
<thead>
<tr>
<th>Positive Symptoms</th>
<th>Negative Symptoms</th>
<th>Relapse/Hospitalizations</th>
<th>Social Functioning/Quality of Life</th>
<th>Other/Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>No group differences</td>
<td>Better</td>
<td>—</td>
<td>—</td>
<td>Reduced trauma associated with psychosis and hospitalization</td>
</tr>
<tr>
<td>Better</td>
<td>No difference</td>
<td>Better</td>
<td>—</td>
<td>63% in remission at follow-up; reduction in aggression and self-harm behaviors</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Better</td>
<td>Only EPPIC patients with midrange duration of untreated psychosis (1–6 months) had better outcomes on Quality of Life Scale</td>
</tr>
<tr>
<td>Better</td>
<td>Better</td>
<td>Better</td>
<td>—</td>
<td>70% in remission at follow-up</td>
</tr>
<tr>
<td>Better</td>
<td>Better</td>
<td>—</td>
<td>Better</td>
<td>74% in remission at follow-up; improvements in cognitive functioning</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Reductions in hallucinogen, cannabis, and alcohol use (among heavy users)</td>
</tr>
<tr>
<td>Better</td>
<td>No difference</td>
<td>—</td>
<td>—</td>
<td>Improvements in depression</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Reduction in parasuicidal behavior; patients who attempted suicide before program made no further attempts</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Improvements in insight</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>TIPS more successful at early identification of psychosis; shorter median duration of untreated psychosis in TIPS group (4.5 weeks versus 26 weeks)</td>
</tr>
<tr>
<td>No group differences</td>
<td>No group differences</td>
<td>Better (Parachute Project)</td>
<td>Better (Parachute Project and prospective control)</td>
<td>High level of consumer satisfaction in Parachute Project group</td>
</tr>
<tr>
<td>Better</td>
<td>Better</td>
<td>—</td>
<td>Better</td>
<td>Less hopelessness in OPUS group; suicidal ideation and attempts reduced in both groups</td>
</tr>
</tbody>
</table>

h Trauma symptoms were measured by the Structured Interview for PTSD. Aggression/self-harm symptoms were measured by the Health of the Nation Outcome Scale and medical records. Cognitive functioning was measured by the WAIS-III, Wechsler Memory Scale, 3rd ed., Wisconsin Card Sorting Test, National Adult Reading Task, Paced Auditory Serial Addition Task, Continuous Performance Test, and Word Fluency Test. Substance use was measured by case manager ratings. Depression was measured by the Calgary Depression Rating Scale for Schizophrenia. Insight was measured by item G12 from the Positive and Negative Syndrome Scale. Hopelessness was measured by the Schedules for Clinical Assessment in Neuropsychiatry interview (version 2.0).

The Parachute Project entailed collaboration between multiple clinics in Sweden, with treatment consisting of comprehensive outpatient services (low-dose atypical antipsychotics, individual and family therapy), available overnight crisis home (in place of inpatient ward), case management, and continuity of care. The historical control consisted of conventional antipsychotic medications and supportive therapy, and the prospective control consisted of inpatient treatment as usual (low-dose atypical antipsychotic medications, no phase-specific psychosocial treatment).

The OPUS Project group received comprehensive outpatient services (low-dose atypical antipsychotics, family psychoeducation and therapy, social skills training) and assertive community treatment. The control group received standard treatment at community mental health centers.
## TABLE 3. Summary of Studies Evaluating the Effectiveness of Specific (i.e., Single-Element) Treatments for Early Psychosis

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Description</th>
<th>Design</th>
<th>Intervention</th>
<th>Comparison Condition(s)</th>
<th>Treatment Length/Follow-Up Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haddock et al., 1999</td>
<td>21</td>
<td>Nonaffective early psychosis (first episode or &lt;5 years since first episode)</td>
<td>Randomized, controlled trial</td>
<td>Individual cognitive behavior therapy (CBT)</td>
<td>Supportive counseling</td>
<td>5 weeks; booster sessions over 4 months</td>
</tr>
<tr>
<td>Lewis et al., 2002</td>
<td>309</td>
<td>Nonaffective early psychosis (83% with first episode)</td>
<td>Randomized, controlled trial</td>
<td>Individual CBT</td>
<td>Supportive counseling, routine care</td>
<td>5 weeks; booster sessions over 3 months</td>
</tr>
<tr>
<td>Tarrier et al., 2004</td>
<td>225</td>
<td>Nonaffective early psychosis (83% with first episode)</td>
<td>Randomized, controlled trial</td>
<td>Individual CBT</td>
<td>Supportive counseling, routine care</td>
<td>18-month follow-up of Lewis et al. study (70)</td>
</tr>
<tr>
<td>Jackson et al., 1998</td>
<td>80</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Individual CBT at Early Psychosis Prevention and Intervention Centre (EPPIC)</td>
<td>EPPIC services without CBT (refusers); EPPIC inpatient care only, no post-discharge services (control)</td>
<td>12 months (median=19 sessions)</td>
</tr>
<tr>
<td>Jackson et al., 2001</td>
<td>51</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Individual CBT at EPPIC</td>
<td>EPPIC services without CBT (refusers); EPPIC inpatient care only, no post-discharge services (control)</td>
<td>12-month follow-up of 1998 Jackson et al. study (72) 8–10 sessions; 6-month follow-up</td>
</tr>
<tr>
<td>Power et al., 2003</td>
<td>56</td>
<td>Nonaffective or affective first-episode psychosis with acute suicidality</td>
<td>Randomized, controlled trial</td>
<td>Individual CBT targeting suicidality at EPPIC</td>
<td>EPPIC services without CBT (control)</td>
<td>Multiple groups per week; 6-month follow-up</td>
</tr>
<tr>
<td>Jolley et al., 2003</td>
<td>21</td>
<td>Nonaffective early psychosis (first or second episode)</td>
<td>Randomized, controlled trial</td>
<td>Individual CBT</td>
<td>Routine care</td>
<td>6 months (mean=11 sessions)</td>
</tr>
<tr>
<td>Wang et al., 2003</td>
<td>251</td>
<td>Nonaffective first-episode psychosis</td>
<td>Randomized, controlled trial</td>
<td>Individual CBT</td>
<td>Routine care</td>
<td>2-year follow-up</td>
</tr>
<tr>
<td>Abiston et al., 1998</td>
<td>95</td>
<td>Nonaffective or affective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>EPPIC group program</td>
<td>EPPIC services without groups</td>
<td>Multiple groups per week; 6-month follow-up</td>
</tr>
<tr>
<td>Miller and Mason, 2001</td>
<td>77</td>
<td>Nonaffective first-episode psychosis</td>
<td>Quasi-experimental</td>
<td>Group therapy</td>
<td>Individual therapy</td>
<td>Once per week for 2 years</td>
</tr>
<tr>
<td>Lecomte et al., 2003</td>
<td>5</td>
<td>Nonaffective first-episode psychosis</td>
<td>Single group</td>
<td>Group CBT</td>
<td></td>
<td>Twice per week for 3 months</td>
</tr>
<tr>
<td>Linszen et al., 1996</td>
<td>76</td>
<td>Nonaffective early psychosis</td>
<td>Randomized, controlled trial</td>
<td>Behavioral family therapy (and individual therapy)</td>
<td>Individual therapy only</td>
<td>12 months</td>
</tr>
<tr>
<td>Lefor et al., 2001</td>
<td>73</td>
<td>Nonaffective early psychosis</td>
<td>Randomized, controlled trial</td>
<td>Behavioral family therapy (and individual therapy)</td>
<td>Individual therapy only</td>
<td>5-year follow-up of study by Linszen et al. (80)</td>
</tr>
<tr>
<td>Lefor et al., 2002</td>
<td>73</td>
<td>Nonaffective early psychosis</td>
<td>Randomized, controlled trial</td>
<td>Behavioral family therapy (and individual therapy)</td>
<td>Individual therapy only</td>
<td>5-year follow-up of study by Linszen et al. (80)</td>
</tr>
<tr>
<td>Zhang et al., 1994</td>
<td>83</td>
<td>Nonaffective first-episode psychosis</td>
<td>Randomized, controlled trial</td>
<td>Family therapy</td>
<td>Routine care</td>
<td>18 months</td>
</tr>
<tr>
<td>Lehtinen, 1993</td>
<td>81</td>
<td>Nonaffective early psychosis</td>
<td>Quasi-experimental</td>
<td>Family-oriented treatment</td>
<td>Individual-oriented treatment (historical cohort)</td>
<td>5-year follow-up</td>
</tr>
</tbody>
</table>

---

a “Better” denotes that patients in the intervention program did significantly better than the comparison group(s) in studies with an experimental or quasi-experimental design or that there was significant improvement over time in studies with a single-group design. “No group differences” denotes no significant difference between the intervention and comparison groups in studies with an experimental or quasi-experimental design or that there was no change over time in studies with a single-group design.

b Nonaffective first-episode psychoses were schizophrenia spectrum disorders. Affective first-episode psychoses were mood disorders with psychotic features.

c The elements of the EPPIC intervention are shown in Table 1. Psychosocial treatments were always adjunctive to pharmacological treatment unless otherwise noted.

d Routine care was primarily medication management.

e Measures were the Brief Psychiatric Rating Scale, Psychotic Symptom Rating Scales, Positive and Negative Syndrome Scale, and chart notes.

f Measure was the Scale for the Assessment of Negative Symptoms.

g Relapse was variably defined as change in patient management (per medical records), hospital admission, and score on Life Chart Schedule.

h Measures were the Quality of Life Scale, Global Assessment of Functioning Scale score, and Life Chart Schedule.

i Measures were the Beck Depression Inventory, Explanatory Model Scale (insight/beliefs about illness), Integration/Sealing Over Measure (adaptation to illness), Suicide Ideation Questionnaire, Suicide Intent Scale, Reasons for Living Inventory, Beck Hopelessness Scale, Self-Esteem Scale, and Self-Report Problem-Solving Rating Scale.

---

<table>
<thead>
<tr>
<th>Positive Symptoms</th>
<th>Negative Symptoms</th>
<th>Relapse/Hospitalizations</th>
<th>Social Functioning/Quality of Life</th>
<th>Other/Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>No group differences</td>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>CBT group improved nonsignificantly faster; auditory hallucinations responded better to CBT than to supportive counseling</td>
</tr>
<tr>
<td>No group differences</td>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>Auditory hallucinations responded better to CBT than to supportive counseling</td>
</tr>
<tr>
<td>No group differences</td>
<td>CBT better than control</td>
<td>No group differences</td>
<td>CBT better than control</td>
<td>Patients receiving CBT did better than refusers and control subjects in adaptation to illness; CBT was better than control for insight/attitudes toward treatment</td>
</tr>
<tr>
<td>No group differences</td>
<td>No group differences</td>
<td>No group differences</td>
<td>No group differences</td>
<td>Patients receiving CBT did better than refusers in adaptation to illness</td>
</tr>
<tr>
<td>No group differences</td>
<td>No group differences</td>
<td>—</td>
<td>Better</td>
<td>Targeted CBT was better than control for hopelessness; both groups improved on suicidal ideation and attempts</td>
</tr>
<tr>
<td>No group differences</td>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>CBT group spent less time in hospital</td>
</tr>
<tr>
<td>Better</td>
<td>—</td>
<td>Better</td>
<td>—</td>
<td>CBT was better than routine care for insight, treatment adherence</td>
</tr>
<tr>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>No group differences</td>
<td>At baseline, group participants had lower premorbid functioning and nonsignificantly more negative symptoms; group treatment was associated with prevention of illness-related deterioration</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Group therapy associated with better treatment adherence (i.e., fewer dropouts)</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Group therapy associated with high treatment satisfaction and decrease in psychotic symptoms</td>
</tr>
<tr>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>—</td>
<td>Family therapy associated with slightly higher relapse rate (nonsignificant difference) among families with low expressed emotion</td>
</tr>
<tr>
<td>—</td>
<td>No group differences</td>
<td>No group differences</td>
<td>—</td>
<td>Family therapy group spent less time in hospitals; 65% of all patients relapsed at least once in 5 years</td>
</tr>
<tr>
<td>—</td>
<td>No group differences</td>
<td>—</td>
<td>—</td>
<td>No differential effect of family therapy on expressed emotion</td>
</tr>
<tr>
<td>Better (in patients not admitted to hospital)</td>
<td>—</td>
<td>Better</td>
<td>Better (in patients not admitted to hospital)</td>
<td>Family group spent less time in hospital</td>
</tr>
<tr>
<td>Better</td>
<td>—</td>
<td>Better</td>
<td>—</td>
<td>Family group spent less time in hospital</td>
</tr>
</tbody>
</table>

---

1 The program of cognitively oriented psychotherapy for early psychosis (COPE) consisted of individual therapy in conjunction with other EPPIC services. It promoted adjustment to illness, recovery, and stigma reduction and targeted associated depression and anxiety.

2 This program, known as LifeSPAN, was conducted in conjunction with other EPPIC services and emphasized distress management, problem solving, self-esteem, hopelessness, warning signs, and aftercare planning.

3 Content areas included vocational skills, creative expression, social and recreational skills, health promotion, and personal development.

4 Behavioral family therapy emphasized communication skills training and reduction of high expressed emotion.

5 Family therapy consisted of family groups and individual family therapy sessions, and emphasized psychoeducation, identification of warning signs, stress management, importance of attributing maladaptive behavior to illness, communication skills training, and reduction of high expressed emotion.

6 In the family-oriented treatment, family therapy was primary, with emphasis on crisis intervention, systemic factors, life difficulties, and short-term treatment. In the individual-oriented treatment, individual dynamic therapy was primary, with focus on intrapsychic factors and long-term treatment.
gating whether early detection and treatment of psychosis can lead to better long-term outcomes (85). A quasi-experimental design is comparing outcomes among individuals with nonaffective first-episode psychosis at three sites: 1) Rogaland, Norway, 2) Oslo, Norway, and 3) Roskilde, Denmark. All three sites offer multielement care, including individual supportive therapy, family work, case management, and pharmacological treatment; however, only the Rogaland site includes specialized early detection and community outreach efforts. Additional efforts to evaluate multielement models of care include the Parachute Project in Sweden, a collaboration among multiple psychiatric clinics to adapt and implement comprehensive first-episode services (i.e., low-dose atypical antipsychotic treatment, case management, individual and family therapy, and access to overnight crisis homes as an alternative to hospitalization) (66), and the OPUS Project in Denmark, which evaluated the effectiveness of comprehensive, integrated care across a variety of treatment modalities (i.e., low-dose atypical antipsychotic treatment, assertive community treatment, family psychoeducation, and social skills training) (67, 68). Indeed, the multielement model of care for early psychosis has been in existence for only a little over a decade but has already garnered significant research support across a variety of programs.

Examination of Table 2 reveals that only one randomized, controlled trial of a multielement intervention has been conducted thus far (67, 68). While additional programs are currently being evaluated in randomized, controlled designs, e.g., at the Early Psychosis Prevention and Intervention Centre (55), the majority of published research in this area has utilized quasi-experimental and single-group designs to evaluate program effectiveness. Thus, the findings should be viewed with caution. Nevertheless, data emerging from these interventions have been encouraging.

A review of Table 2 indicates that multielement interventions for early psychosis have been associated with symptom reduction and/or remission (33, 56, 57, 59–61, 68), improved quality of life and social functioning (46, 56, 58, 59, 61, 68), improved cognitive functioning (61), reduced duration of untreated psychosis (65), low rates of inpatient admissions (56, 60, 66), improved insight (64), high level of satisfaction with treatment (66), less time spent in the hospital (56, 66), decreased substance abuse (62), fewer self-harm behaviors (57, 63, 67), and reduced trauma secondary to psychosis and hospitalization (56). It should be noted that the foregoing results primarily refer to 1-year outcomes; longer-term benefits conferred by multielement programs have not been widely reported. Finally, a recent study suggests that these comprehensive and specialized first-episode services are likely to yield superior outcomes (e.g., shorter duration of untreated psychosis, fewer inpatient admissions, less time in the hospital) when compared with generic mental health services (86).

Single-Element Interventions

Single-element studies have evaluated the effectiveness of specific psychosocial interventions, rather than assessing the effects of a comprehensive, multielement intervention as a whole. That is, these studies sought to measure the relative utility of various adjunctive psychosocial interventions in the treatment of early psychosis. These interventions were offered in addition to pharmacological treatment and, in some cases, other services as well (e.g., case management). Examination of Table 3 reveals that several randomized, controlled trials have been conducted with respect to individual cognitive behavior therapy in early psychosis, but less controlled research has evaluated group and family interventions. Findings from many of these studies have been promising, and the results are discussed in more detail in the following sections.

Individual Therapy

Individual therapy for first-episode psychosis has been examined both for facilitating recovery from acute psychosis and for improving longer-term outcome following remission of acute psychosis. With respect to the former, the Study of Cognitive Reality Alignment Therapy in Early Schizophrenia was a large, multisite randomized, controlled trial of cognitive behavior therapy for recent-onset acute psychosis. On the basis of a pilot study by Haddock et al. (69), Lewis and colleagues (70) randomly assigned 309 individuals who had either a first (83%) or second psychiatric admission for psychosis to 5 weeks of cognitive behavior therapy and routine care, supportive counseling and routine care, or routine care alone. While all groups improved over the course of treatment, the group receiving cognitive behavior therapy improved nonsignificantly faster. Further, auditory hallucinations improved significantly faster in that group than in the group receiving supportive counseling. There were no significant group differences, however, in symptoms at the end of treatment. At 18-month follow-up, Tarrier and colleagues (71) found that both cognitive behavior therapy and supportive counseling were significantly better than routine care in reducing symptoms. Further, auditory hallucinations responded better to cognitive behavior therapy than to supportive counseling. However, there were no group differences in relapse rates, with high overall rates of relapse across the total study group. Tarrier et al. hypothesized that the short duration of treatment, a failure of treatment effects to generalize outside the hospital, potential exposure to environmental stressors after discharge, and the tendency for relapse to accumulate over time in first-episode psychosis may explain the lack of an impact on relapse conferred by cognitive behavior therapy or supportive counseling. Nevertheless, these results suggest that individual therapy (i.e., cognitive behavior therapy or supportive counseling) may have beneficial long-term effects on symptoms in early psychosis.
Promising results have also been reported with respect to cognitive behavior therapy conducted during the period of recovery following reduction of acute psychotic symptoms. Jackson and colleagues (72) conducted a prospective study of cognitively oriented psychotherapy for early psychosis with 80 individuals in the Early Psychosis Prevention and Intervention Centre program who were experiencing nonaffective or affective first-episode psychosis. Cognitively oriented psychotherapy for early psychosis promoted adjustment to the illness, helped individuals resume developmental tasks, and focused on overall recovery, in addition to targeting secondary morbidity (i.e., depression, anxiety). Forty-four individuals received cognitively oriented psychotherapy as part of their outpatient care, 21 refused but received all of the center’s other services, and 15 individuals received inpatient care only with no additional services following discharge (they were designated the control group). At the end of treatment, the patients receiving cognitively oriented psychotherapy performed significantly better than the control group on measures of insight and attitudes toward treatment, adaptation to illness, quality of life, and negative symptoms, but they significantly outperformed the refusal group only with respect to adaptation to illness. There were no significant differences in relapse rates among the three groups. At 1 year following treatment, the group receiving cognitively oriented psychotherapy maintained significantly better adaptation to their illness than the refusal group; however, the group differences were not maintained for the other outcomes, and there were no group differences in relapse rate or time to readmission (73). These findings are based on a quasi-experimental design and need to be interpreted with caution; nevertheless, the results suggest that individual cognitive behavior therapy may be beneficial in assisting patients to adjust to the experience of psychosis following remission of first-episode symptoms.

Individual cognitive behavior approaches have been developed to target specific challenges facing patients experiencing first-episode psychosis, such as suicidality, substance abuse, and persistent symptoms. In a study focusing on suicidal ideation and behavior in early psychosis, Power and colleagues (74) randomly assigned 56 suicidal individuals with nonaffective or affective first-episode psychosis in the Early Psychosis Prevention and Intervention Centre program to either LifeSPAN therapy plus the center’s other services or regular services without LifeSPAN therapy. LifeSPAN therapy is based on cognitively oriented psychotherapy for early psychosis as well as cognitive therapy for suicide, and it emphasizes distress management, problem-solving skills, identification of warning signs, and development of an aftercare plan. In addition, low self-esteem and feelings of hopelessness are specifically targeted. In this study, both groups improved on ratings of suicidal ideation and number of suicide attempts. However, the results showed an advantage for LifeSPAN therapy on self-reported hopelessness and quality of life at both 10 weeks posttreatment and 6-month follow-up. Power et al. concluded that adding cognitive behavior therapy to treatment for first-episode psychosis may lead to significant improvements in factors associated with suicide, such as hopelessness.

Edwards and colleagues at the Early Psychosis Prevention and Intervention Centre have developed cognitive behavior interventions targeting substance use and persistent psychotic symptoms (87, 88). One intervention focuses on reducing problematic cannabis use in individuals with first-episode psychosis and consists of psychoeducation, motivational interviewing, goal setting, and discussion about goal achievement and relapse prevention. A randomized, controlled trial comparing the cannabis and psychosis intervention with psychoeducation alone was conducted, and the preliminary results suggested that cannabis use in both groups decreased, with no clear advantages for the cannabis and psychosis intervention over psychoeducation alone (89). Edwards and colleagues have also developed “systematic treatment of persistent psychosis,” given that approximately 20% of individuals with first-episode psychosis may experience persistent psychotic symptoms (40). This therapy is based on the cognitively oriented psychotherapy for early psychosis at the Early Psychosis Prevention and Intervention Centre and is designed to facilitate recovery in patients experiencing persistent positive symptoms. A randomized, controlled trial evaluating the relative and combined effects of clozapine and systematic treatment of persistent psychosis in the treatment of individuals with persistent symptoms is currently being conducted at the Early Psychosis Prevention and Intervention Centre (88).

Other randomized, controlled studies of individual cognitive behavior therapy for first-episode psychosis have demonstrated the following benefits over routine care: fewer days spent in the hospital (75), reduced psychotic symptoms, fewer hospital admissions, increased insight, and better treatment adherence (76). The foregoing findings suggest that individual cognitive behavior therapy may provide some benefits in the treatment of first-episode psychosis, especially in the areas of symptom reduction, adaptation to one’s illness, and improvements in subjective quality of life. Most studies have not shown individual therapy to be effective in reducing relapses or rehospitalizations. Finally, the long-term findings are mixed; the follow-up data reported thus far have demonstrated some long-term benefits associated with individual therapy (e.g., references 71 and 73) but also suggest that some of the initial gains made in treatment may not persist over time (e.g., reference 73).

**Group and Family Treatment**

Unlike individual therapy, group treatment for first-episode psychosis does not appear to have been examined for efficacy in randomized, controlled studies. Quasi-experimental research has demonstrated benefits of group ther-
apy with respect to prevention of illness-related deterioration and disability, especially for individuals with poor premorbid functioning (77). Additional uncontrolled studies have shown improved treatment adherence (78) and increased treatment satisfaction (79) associated with group participation. However, given the uncontrolled nature of these studies, these findings need to be interpreted with caution.

Family therapy for first-episode psychosis has been more systematically investigated. Linszen and colleagues (80) randomly assigned 76 outpatients to 12 months of behavioral family therapy (focusing on communication and problem-solving skills training) plus individual-oriented treatment (focusing on relapse prevention and psychoeducation) or individual-oriented treatment without family therapy. Both groups had recently been discharged after 3 months of inpatient treatment emphasizing integrated psychosocial and pharmacological treatment, and they were currently receiving outpatient medication management. After 1 year, there was no differential effect of the family therapy on relapse; the two groups had similar relapse rates, and the overall relapse rate was low (i.e., 16%). Five-year follow-up (81, 82) also indicated no added benefit of family treatment over individual treatment for relapse rates, and it showed that 65% of the patients in the total group with nonchronic symptoms relapsed at least once over the course of 5 years. In addition, this study showed no differential effect of family treatment on social functioning or expressed emotion. However, individuals who received family therapy spent significantly less time in hospitals and/or shelters.

Other research on family therapy for early psychosis has demonstrated more positive results. For example, Zhang and colleagues (83) randomly assigned 83 outpatients with first-episode psychosis to 18 months of family therapy and routine care or to routine care alone. The family therapy intervention consisted of family groups and individual family therapy sessions, and it emphasized psychoeducation, identification of warning signs, stress management, the importance of attributing maladaptive behavior to the illness (rather than to personality or “laziness”), communication skills training, and reduction of high expressed emotion (i.e., decreasing familial criticism, hostility, and overinvolvement). There was contact with the families at least every 3 months, and families who did not attend appointments were visited in their homes. The results showed that the family intervention was associated with a significantly lower rate of hospital readmissions and fewer days spent in the hospital. Indeed, the authors concluded that the patients not receiving the family intervention were 3.5 times as likely to be readmitted to the hospital during the study period as the patients who did receive family therapy. This effect remained even after differences in medication compliance were controlled for. Further, the patients receiving family therapy who were not readmitted to the hospital demonstrated significant improvements in positive symptoms and social functioning. Additional research has shown similar favorable outcomes associated with family treatment, such as fewer hospital admissions, less time spent in the hospital, and symptom reduction (84).

Thus, while some research has indicated that family interventions in early psychosis are beneficial with respect to reducing relapse and improving clinical and functional status (e.g., reference 83), other findings have not been as encouraging (e.g., reference 80). More empirical work needs to be done before any firm conclusions can be made.

Finally, Drury and colleagues (90, 91) specifically evaluated the effects of a multimodal treatment approach combining individual and group cognitive behavior therapy with family therapy in the treatment of recent-onset acute psychosis. In a randomized, controlled trial, the combination treatment, compared with basic support and recreational activities, yielded faster and greater improvements of positive symptoms, reduced recovery time by 25%-50%, and led to improvements in insight, dysphoria, and “low-level” psychotic thinking (e.g., suspiciousness). In a 5-year follow-up, Drury et al. (92) found enduring positive effects for the combination therapy group relative to the control group; however, these benefits were predominantly observed in individuals who had experienced at most one relapse over the course of follow-up. The long-term benefits in this subgroup included fewer positive symptoms, less delusional conviction and thought disorder, and better subjective “control over illness.” While these findings are positive, this study has been criticized for methodological flaws in its design, such as nonblinded assessments (93) and baseline differences in medication doses between the two groups (94).

Discussion

The findings reviewed suggest that adjunctive psychosocial interventions for patients experiencing early psychosis are beneficial across a variety of domains and can assist with symptomatic and functional recovery. Research on multielement interventions indicates that following an initial episode of psychosis, these comprehensive treatment approaches may positively influence short-term outcomes, such as clinical status and social functioning, as well as time spent in the hospital and likelihood of hospital readmission. However, as noted in another recent review of this area (53), most of the research on multielement programs is based on quasi-experimental designs using historical (56, 58, 65, 66) or prospective (66) comparison groups or on single-group designs, which track the progress of one group over a specified period of time (33, 46, 57, 59–64). Indeed, there is still a paucity of randomized, controlled research in this area; thus, these findings need to be interpreted with caution. Other methodological issues making interpretation of these results challenging include subject heterogeneity (e.g., affective versus nonaffective first-episode psychosis) and varying defini-
tions for certain outcomes, such as relapse and remission, across studies.

One important caveat regarding multielement interventions is that the scope of these programs makes them difficult to implement on a widespread basis, particularly in countries whose public health care systems do not support the necessary infrastructure or do not recognize best treatment practices for early psychosis (95). Indeed, given the wide range of services offered in these programs, it would be helpful to isolate the “effective ingredients” when evaluating a program’s utility. Understanding which elements are critical can help inform program development in areas currently lacking such specialized first-episode treatment services. Thus, the current findings in this area point to two important future research directions: 1) a greater number of randomized, controlled designs to provide a more stringent test of the efficacy of multielement programs and 2) utilization of research designs that will allow one to deconstruct the key ingredients of these programs and to determine the specific types of patients for whom these services are particularly beneficial. Single-element studies can be quite helpful in this regard.

With respect to research on single-element interventions, support for individual cognitive behavior therapy in early psychosis is modest yet encouraging, especially regarding symptom improvements (particularly positive symptoms), adaptation to one’s illness, and increased subjective quality of life (e.g., references 71–74). In addition, the majority of studies evaluating individual cognitive behavior therapy have used randomized, controlled designs. However, individual therapy has not been shown to be effective in reducing relapse or rehospitalization in early psychosis, and some findings suggest that early treatment gains may not be maintained over time.

No firm conclusions can yet be drawn from the literature on group and family therapies for this population. Group therapy is a widely used treatment modality for early psychosis, but to our knowledge, no randomized, controlled trials have been conducted. Research findings on family therapy in early psychosis have been mixed, with some studies documenting benefits with respect to symptoms, social functioning, and likelihood of rehospitalization (e.g., reference 83) and other studies having less favorable results (e.g., reference 80). One possible interpretation of these findings is that family interventions are indeed beneficial to individuals with early psychosis although they may not add significant benefit above and beyond concurrent individual therapy. Additional well-controlled research is needed to clarify the impact of family and group therapy in first-episode psychosis.

In general, while the research done to date on specific (i.e., single-element) psychosocial treatments for early psychosis is promising, there are few robust findings. Many of the aforementioned single-element studies were conducted in the context of large multielement programs (e.g., references 72–74); it is therefore difficult to yield robust additive effects of a specific intervention, when such a large degree of improvement is likely due to the impact of the program as a whole. Further, as with the literature on multielement treatments, significant obstacles to drawing broader conclusions with respect to specific psychosocial treatments for first-episode psychosis include the paucity of well-controlled studies, as well as methodological variation among studies (e.g., study group composition, definitions for remission and relapse).

Future work should take an increasingly integrative approach to psychosocial treatment, drawing on a variety of empirically validated treatment approaches to address the variety of challenges that individuals with first-episode psychosis experience (e.g., positive and negative symptoms, medication adherence, substance use, functional impairments). Indeed, future studies should place more emphasis on measuring functional recovery (i.e., social, work, and school functioning, recreation, and social relationships [96]) both during and after treatment. Despite demonstrated short-term benefits, the ability of psychosocial interventions delivered early in psychosis to effect long-term improvement, particularly with respect to social/occupational disability, is still unknown. Additional longitudinal research is needed to shed light on this question.

Some findings suggest that many of the initial benefits achieved in treatment may not be maintained over time in patients with first-episode psychosis (97). This may be addressed through greater efforts to improve ongoing engagement with available services (which is a significant challenge in the field of early psychosis [e.g., reference 98]) and to lengthen the duration of active interventions, if necessary. Studies of individuals with chronic schizophrenia suggest that longer-term treatments are often associated with more favorable outcomes (99). In addition, naturalistic studies of psychological treatments for a variety of non-psychotic disorders have demonstrated that patients tend to show greater degrees of improvement with longer periods of treatment (100). Clinicians and researchers alike should utilize these findings to inform the delivery of psychosocial interventions in early psychosis. Ideally, these efforts will be successful at improving both short- and long-term outcomes, thus reducing the morbidity and mortality so often associated with this devastating illness.

Received Jan. 19, 2004; revision received Jan. 3, 2005; accepted Jan. 6, 2005. From the Department of Psychology and the Department of Psychiatry, University of North Carolina; the Department of Psychiatry, Dartmouth Medical School, Hanover, N.H.; and the Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York. Address correspondence and reprint requests to Dr. Penn, Department of Psychology, CB#3270, Davie Hall, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599-3270; dpenn@email.unc.edu (e-mail).

References


http://ajp.psychiatryonline.org
PSYCHOSOCIAL TREATMENT


29. Wyatt RJ, Henter I: Rationale for the study of early intervention. Schizop hr Bull 2001; 51:69–76


76. Wang C, Li Y, Zhao Z, Pan M, Feng Y, Sun F, Du B: [Controlled study on long-term effect of cognitive behavior intervention on first episode schizophrenia.] Chinese Ment Health J 2003; 17:200–202 (abstract in English)


Regional Deficits in Brain Volume in Schizophrenia: A Meta-Analysis of Voxel-Based Morphometry Studies

Robyn Honea, B.Sc.
Tim J. Crow, M.B., Ph.D.
Dick Passingham, Ph.D.
Clare E. Mackay, Ph.D.

Objective: Voxel-based morphometry is a method for detecting group differences in the density or volume of brain matter. The authors reviewed the literature on use of voxel-based morphometry in schizophrenia imaging research to examine the capabilities of this method for clearly identifying specific structural differences in patients with schizophrenia, compared with healthy subjects. The authors looked for consistently reported results of relative deficits in gray and white matter in schizophrenia and evaluated voxel-based morphometry methods in order to propose a future strategy for using voxel-based morphometry in schizophrenia research.

Method: The authors reviewed all voxel-based morphometry studies of schizophrenia that were published to May 2004 (15 studies). The studies included a total of 390 patients with a diagnosis of schizophrenia and 364 healthy volunteers.

Results: Gray and white matter deficits in patients with schizophrenia, relative to healthy comparison subjects, were reported in a total of 50 brain regions. Deficits were reported in two of the 50 regions in more than 50% of the studies and in nine of the 50 regions in one study only. The most consistent findings were of relative deficits in the left superior temporal gyrus and the left medial temporal lobe. Use of a smaller smoothing kernel (4–8 mm) led to detection of a greater number of regions implicated in schizophrenia.

Conclusions: This review implicates the left superior temporal gyrus and the left medial temporal lobe as key regions of structural difference in patients with schizophrenia, compared to healthy subjects. The diversity of regions reported in voxel-based morphometry studies is in part related to the choice of variables in the automated process, such as smoothing kernel size and linear versus affine transformation, as well as to differences in patient groups. Voxel-based morphometry can be used as an exploratory whole-brain approach to identify abnormal brain regions in schizophrenia, which should then be validated by using region-of-interest analyses.

Given the early demonstration of smaller ventricle size in schizophrenia (1) and the known genetic predisposition to the disorder (2), researchers for the past 25 years have looked for more specific structural abnormalities in schizophrenia. Many brain regions have been implicated in schizophrenia by means of magnetic resonance imaging (MRI) studies; findings have included ventricular enlargement, total brain volume deficits, and deficits in brain volume within the frontal, temporal, and parietal regions (3–6). To date, no single region has been consistently reported to be abnormal.

Several meta-analyses and reviews of brain abnormalities in schizophrenia have been conducted (7–13). Their findings include larger ventricles, smaller mean cerebral volume, reversed asymmetry in the superior temporal gyrus, and smaller volume of the medial temporal lobes in patients with schizophrenia. In a thorough meta-analysis of studies of regional brain volumes in schizophrenia, Wright et al. (7) found that in 58 studies involving a total of 1,588 independent patients with schizophrenia, the mean cerebral volume of the patients was 2% smaller than the mean for the healthy comparison subjects. Another meta-analysis with more strict inclusion criteria showed that leftward asymmetry of the planum temporale (superior temporal gyrus) and the sylvian fissure was significantly less pronounced in schizophrenia patients than in healthy comparison subjects (9). Others have focused on more localized morphometric abnormalities. For instance, Zakzanis et al. (8) evaluated the strength and consistency of study findings published between 1980 and 2000 that reported temporal lobe abnormalities in patients with schizophrenia. They found heterogeneity in the findings, with some studies showing abnormal structure and function and others reporting no difference between patients and healthy comparison subjects. Other meta-analyses focused on medial temporal and diencephalic structures have reported a moderate effect size for findings of relative deficits in gray matter (10–13).

Studies reporting structural abnormalities in schizophrenia have included a variety of patient groups that differ in age at onset, medications received, length of treatment with medication, symptoms, presence of comorbidity, and severity of illness. Because of a possible...
VOXEL-BASED MORPHOMETRY

progression of neuropathology during the illness, age and severity effects could cause variations in brain morphometry over time (14–16). In the following review, we therefore distinguish first-episode and chronic patient groups. In the imaging literature, the most prominent method used to investigate structural abnormalities is region of interest analysis. Such analyses rely on a priori defined regions of interest and manual outlining or stereological procedures to obtain volumetric measurements. These user-dependent methods could contribute a bias to the results (17, 18). In addition, region of interest analysis tends to focus on specific hypothesized regions, and findings in unexpected or unspecified regions can be overlooked (19). Finally, manual region of interest techniques are laborious, which can hinder efficient processing of large cohorts.

Thus, an automated, efficient whole-brain analysis to detect structural differences would provide a means of identifying areas of cerebral abnormality in a user-independent way. One such method is voxel-based morphometry, which was first reported in a study that used statistical parametric mapping (20). Voxel-based morphometry involves a voxel-wise comparison of the probability of the presence of gray or white brain matter, sometimes described as the density or concentration of gray or white

### TABLE 1. Technical Specifications of Voxel-Based Morphometry Studies of Schizophrenia Published to May 2004

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study</th>
<th>Subjects</th>
<th>Magnetic Resonance Imaging (MRI) Scanner</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Marcelis et al., 2003 (22)</td>
<td>31 schizophrenia patients; 32 nonpsychotic first-degree relatives; 27 healthy comparison subjects</td>
<td>1.5-T scanner</td>
<td>Brain Activation and Morphological Mapping (BAMM) (27); Analysis of Functional NeuroImages (AFNI) piecewise linear rescaling</td>
</tr>
<tr>
<td>2</td>
<td>Spalletta et al., 2003 (25)</td>
<td>28 (chronic) schizophrenia patients; 28 age-, sex-, and education-matched healthy comparison subjects</td>
<td>1.5-T scanner</td>
<td>Linear transformation to Montreal Neurological Institute template; INSECT algorithm for white matter class</td>
</tr>
<tr>
<td>3</td>
<td>Sigmundsson et al., 2001 (24)</td>
<td>27 right-handed, chronic schizophrenia patients with enduring negative symptoms; 27 healthy comparison subjects</td>
<td>1.5-T General Electric scanner</td>
<td>Nonstatistical parametric mapping registration and segmentation (27); tissue class program (30)</td>
</tr>
<tr>
<td>4</td>
<td>Kubicki et al., 2002 (18)</td>
<td>16 first-episode schizophrenia patients; 18 healthy comparison subjects; 16 first-episode affective psychosis patients</td>
<td>1.5-T General Electric scanner</td>
<td>Statistical Parametric Mapping (SPM) 99</td>
</tr>
<tr>
<td>5</td>
<td>Ananth et al., 2002 (31)</td>
<td>20 schizophrenia patients; 20 healthy comparison subjects</td>
<td>2-T Siemens scanner</td>
<td>MATLAB 5.3, SPM 99</td>
</tr>
<tr>
<td>6</td>
<td>Job et al., 2002 (32)</td>
<td>34 first-episode schizophrenia patients; 36 healthy comparison subjects</td>
<td>1.0-T scanner</td>
<td>Analyze 7.5.5 program (correction for field inhomogeneity); SPM 99</td>
</tr>
<tr>
<td>7</td>
<td>Shapleske et al., 2002 (23)</td>
<td>41 hallucinating schizophrenia patients; 31 nonhallucinating schizophrenia patients; 32 healthy comparison subjects</td>
<td>1.5-T General Electric scanner</td>
<td>SPM 99</td>
</tr>
<tr>
<td>8</td>
<td>Hulshoff Pol et al., 2001 (21)</td>
<td>159 patients with schizophrenia or schizophreniform disorder; 158 healthy comparison subjects</td>
<td>1.5-T Philips scanner</td>
<td>ANIMAL algorithm, personal software</td>
</tr>
<tr>
<td>9</td>
<td>Salgado-Pineda et al., 2003 (33)</td>
<td>13 right-handed, male, first-episode, paranoid, neuroleptic-naive schizophrenia patients; 13 healthy comparison subjects</td>
<td>1.5-T General Electric Signa scanner</td>
<td>SPM 99</td>
</tr>
<tr>
<td>10</td>
<td>Pallere-Martinot et al., 2001 (34)</td>
<td>20 male chronic schizophrenia patients; 20 healthy male comparison subjects</td>
<td>1.5-T General Electric Signa scanner</td>
<td>SPM 96</td>
</tr>
<tr>
<td>11</td>
<td>Wright et al., 1999 (35)</td>
<td>42 schizophrenia patients; 52 healthy comparison subjects</td>
<td>1.0-T Picker HPQ Vista System and 1.5-T ANALYZE scanner</td>
<td>SPM</td>
</tr>
<tr>
<td>12</td>
<td>Wilke et al., 2001 (36)</td>
<td>48 schizophrenia patients; 48 age- and sex-matched healthy comparison subjects</td>
<td>1.5-T GE Signa scanner</td>
<td>SPM 99</td>
</tr>
<tr>
<td>13</td>
<td>Moorhead et al., 2004 (37)</td>
<td>25 (14 male, 11 female) schizophrenia patients; 29 (13 male, 16 female) healthy comparison subjects</td>
<td>1.0-T Siemens scanner</td>
<td>SPM 99, MATLAB</td>
</tr>
<tr>
<td>14</td>
<td>Salgado-Pineda et al., 2004 (38)</td>
<td>14 (7 male, 7 female) right-handed schizophrenia patients; 14 healthy comparison subjects</td>
<td>1.5-T GE Signa scanner</td>
<td>SPM 2</td>
</tr>
<tr>
<td>15</td>
<td>Hulshoff Pol et al., 2004 (39)</td>
<td>159 patients with schizophrenia or schizophreniform disorder; 158 healthy comparison subjects</td>
<td>1.5-T Philips scanner</td>
<td>ANIMAL algorithm, personal software</td>
</tr>
</tbody>
</table>

a Data for subjects with learning disability and subjects with comorbid schizophrenia and learning disability were not included in review.
tients with schizophrenia. Several groups have adapted the concepts and steps of these methods, which are based on statistical parametric mapping, for voxel-based morphometry with other software (21–25). We reviewed the voxel-based morphometry studies in schizophrenia in terms of methodological differences and patient populations to identify the brain regions that are most consistently implicated in the disorder.

**Method**

Studies were considered for the review by using the following inclusion criteria: 1) they were published through May 2004 as an article (rather than a letter or an abstract), 2) they compared a group of subjects with schizophrenia (either first-episode patients or chronic patients) and a healthy comparison group, and 3) they utilized voxel-based morphometry analysis of MRI datasets to investigate differences in whole-brain matter by using an automated procedure, and they followed the general steps of normalization, segmentation, smoothing, and statistical analysis described in the early literature on the method (20, 26). Table 1 lists the articles included in the review and provides information about the subject groups, software used, voxel-based morphometry process (template used, affine registration, and varying methods), and smoothing kernel in each study (18, 21–25, 31–39).

For each study, we identified the regions that were reported to be reduced in patients relative to comparison subjects, in order to establish which relative deficits are most consistently reported and whether specific methodological factors contribute to inconsistencies in results. We excluded gender effects and comparisons between additional groups, such as first-degree relatives or subjects with other psychiatric disorders. Therefore, the results included in our review represent the statistical comparison of brain matter volume in patients with that in healthy comparison subjects, regardless of any other comparisons each study might have reported. Hulshoff Pol et al. reported gray and white matter analyses for the same cohort in separate manuscripts (21, 39), and these results were combined for presentation in Figure 1. Voxel-based morphometry data are variously referred to as density, concentration, or volume. For ease of description, we will use the term "volume" to describe results, although we acknowledge that in most cases, these are not true structural volumes.

Regions with significant differences were identified according to the labeling reported by the authors in each study. Because studies used different smoothing kernels and defined the medial temporal regions in different ways, we did not distinguish sub-areas within the medial temporal region; in particular, the amygdala, entorhinal cortex, and hippocampus were all considered as part of the medial temporal lobe in this review.

**Results**

Fifty regions were reported to be reduced in patients with schizophrenia relative to healthy comparison subjects. Figure 2 shows the percentages of studies that found reductions in each of the 50 regions. Two brain regions were found to be reduced in patients, relative to comparison subjects, in more than 50% of the studies: the left medial temporal lobe (nine of 15 studies) and the left superior temporal gyrus (eight of 15 studies). In about 50% of the studies we reviewed, relative deficits in patients were reported in the following regions: left inferior frontal gyrus, left medial frontal gyrus, right superior temporal gyrus, and left para-
hippocampal gyrus. In contrast, significant relative deficits were reported in the following regions in only one of the 15 studies: left inferior parietal lobe, left and right middle temporal gyrus, right fusiform gyrus, right and left posterior cingulate gyrus, right precentral gyrus, and left inferior temporal gyrus.

Figure 1 and Figure 2 summarize the findings of all 15 studies. Only three of the 15 studies in this review used a first-episode patient group; the findings of these studies are summarized in the last three columns in Figure 1. Of the regions that were reported to be significantly smaller in schizophrenia patients in at least six of the 15 studies,
two regions showed a disparity between chronic and first-episode patients. Reduced left medial frontal gyrus volume in schizophrenia patients was reported in seven (64%) of 11 studies of chronic patients and in none of the three studies of first-episode patients. Reduced right anterior cingulate volume was reported in all three studies of first-episode patients that included analysis of this region and in only three (27%) of 11 studies of chronic patients.

In Figure 3 the 50 regions implicated were highlighted according to the Talairach demon (40) and displayed using mri3dX (http://www.aston.ac.uk/lhs/research/groups/nrg/mri3dx). The images are colored according to the percentage of studies that reported significant reductions in patients with schizophrenia (Figure 1 and Figure 2). The lateral views in Figure 3 show the relative deficits in the superior temporal and inferior frontal cortex, and the coronal and axial views show the relative deficits in the medial frontal and anterior cingulate cortex and in the medial temporal lobe.

We focused on the regions that were reported to be significantly reduced in patients with schizophrenia in more than 50% of the studies: the superior temporal gyrus (Figure 4, Table 2) and the medial temporal lobe (Figure 5, Table 3). In an analysis of the consistency of the results for each structure, the coordinate of the significant difference reported in each study was plotted as a sphere. The size of the smoothing kernel was used to determine the diameter of the sphere. For instance, if a study used an 8-mm smoothing kernel and reported a coordinate in the superior temporal gyrus, we created a sphere 8 mm in diameter and plotted it by using the reported coordinate. In Figure 4 and Figure 5, the color scale of the spheres is graded by the stringency of the statistics used. Dark red signifies papers with analyses by trend or cluster with no overlap between coordinates; lighter red signifies papers with analyses by voxel peak height. The overlap between the results for the different papers is additive, and areas with greater overlap are nearer to yellow. The coordinates of each reported significant deficit are given for the superior temporal gyrus in Table 2 and for the medial temporal lobe in Table 3.

The smoothing kernel is an important variable in the voxel-based morphometry method, and a range of kernel widths, from 4 mm to 12 mm, was used in these studies. Figure 6 shows the percentage of studies reporting reduced volume in schizophrenia patients in each brain region, according to whether a small (4–8 mm) or large (10–12 mm) smoothing kernel was used in the study. This analysis showed that smaller smoothing kernels were used in a greater number of studies that reported volume reductions in small structures such as the medial temporal lobes, parahippocampal gyrus, thalamus, superior temporal gyrus, medial frontal gyrus, and anterior cingulate.

Discussion

Summary of Results

We reviewed 15 studies that employed voxel-based morphometry in 14 different cohorts of schizophrenia patients.
and comparison subjects and found that of the 50 regions for which deficits in volume or density were reported, two regions were identified in more than 50% of studies: the left superior temporal gyrus (in 57% of studies) and the left medial temporal lobe (in 69% of studies). Deficits in nine regions were reported in only one of the 15 studies. Use of a smaller smoothing kernel (4–8 mm) tended to yield a greater number of areas with significant deficits, particularly in the smaller structures.

**Methodological Considerations**

Methodological differences in implementation of voxel-based morphometry may explain why some regions were reported to be reduced in schizophrenia patients in only one of the studies we reviewed. For example, nine studies used cluster-level significance in the statistical analyses, and this method has been reported to cause false positive results (20). To compare statistical procedures and determine an effect size, data on peak location, cluster size, and statistical threshold would be required for each region, and these data were not reported in every study we reviewed. In Figure 4 and Figure 5 we attempted to qualify these parameters for the superior temporal gyrus and medial temporal lobe in terms of the size (smoothing kernel), weight (statistical threshold), and location of the spheres.
According to this qualitative representation, no systematic differences between regions identified with cluster-level and voxel-level statistics were detected.

In addition, the segmentation method in voxel-based morphometry requires a clear contrast between the different tissue types to avoid indistinguishability in regions where central gray matter structures have image intensities that most resemble those of white matter. These regions are particularly likely to show partial volume effects, which occur when a single voxel contains a mixture of tissue types instead of the assumed single tissue type. These effects can also arise in white matter regions near the ventricles, where voxels at the interface appear as gray matter. Patients with schizophrenia are particularly vulnerable to such partial volume effects because they have a higher ventricle-to-brain ratio than healthy volunteers (3, 41).

Improvements in voxel-based morphometry have been reported. Optimized voxel-based morphometry, described by Good et al. (26), includes use of a study-specific template and nonlinear normalization and takes into account Jacobian determinants. The optimized method involves the following steps: 1) segmenting the MR images in native space; 2) “cleaning” the segmented images by removing nonbrain voxels; 3) estimating normalization parameters to a normalized, averaged brain matter template; 4) reapplying these parameters to the whole-brain MR images; 5) segmenting the normalized image; and 6) cleaning the normalized brain matter image again (36). These steps are followed by the optional reintegration of the volume changes that occurred during nonlinear normalization by modulation with the Jacobian determinant of the deformation matrix, allowing for the evaluation of gray matter volume (in contrast to density) changes. Four studies in this review used this optimized method (22, 23, 31, 42).

The level of sensitivity of voxel-based morphometry, as well as its validation with region of interest analysis, has yet to be fully determined (18, 43). However, in a growing number of studies, voxel-based morphometry was vali-
those that used small smoothing kernels. Therefore, as the
often in studies that used large smoothing kernels and in
studies that used a smaller smoothing kernel. Superior
temporal lobe abnormalities were more often reported in
the most consistently reported abnormalities, medial
size of the smoothing kernel used by each study. Among
Table 1. Figure 6 displays the study findings in terms of the
detected (51). The studies we reviewed used a wide range
scale at which morphological changes are most sensitively
determined by region of interest analysis (18, 44–48). Keller et al.
(47) demonstrated methodological consistency between
automated voxel-based morphometry and manual analy-
sis of hippocampal volume in patients with temporal lobe
epilepsy. In addition, Kubicki et al. (18) replicated a previ-
ous region of interest finding (in the same cohort) of
smaller gray matter volume in the left posterior superior
temporal gyrus in schizophrenia patients, relative to
healthy comparison subjects and affective psychosis pa-
tients, as well as less density in the left medial temporal
lobe in patients with schizophrenia (49). In some cases,
voxel-based morphometry is less sensitive than region of interest methods (48), but sensitivity can be increased by
using a small smoothing kernel and a small-volume cor-
correction, as well as a study-specific template (32).

Wilke et al. (50) sought to substantiate voxel-based
morphometry by systematically determining the best
combination of processing parameters for examination of
malformations in cortical development, based on recom-
endations related to the optimized processing method
(26). Wilke et al. used 99 permutations of normalization
procedures and smoothing kernels, as well as density ver-
sus volume measurements, to compare 20 patients with
known cortical malformation to 53 healthy age- and gen-
der-matched comparison subjects. The results that most
accurately matched the known cortical malformations
were obtained with an affine-only approach and a 6-mm
smoothing kernel. According to the matched filter theo-
rem, the width of the smoothing kernel determines the scale at which morphological changes are most sensitively
detected (51). The studies we reviewed used a wide range
of smoothing kernels, as reported in the last column of
Table 1. Figure 6 displays the study findings in terms of the
size of the smoothing kernel used by each study. Among
the most consistently reported abnormalities, medial
temporal lobe abnormalities were more often reported in
studies that used a smaller smoothing kernel. Superior
temporal gyrus abnormalities were reported equally as
often in studies that used large smoothing kernels and in
those that used small smoothing kernels. Therefore, as the
findings of Wilke et al. suggested, this variable should be
chosen by using a hypothesis-driven approach for a spe-
cific region of interest, and it should be noted that use of a
small smoothing kernel can possibly lead to false positive
results in studies of schizophrenia patients.

Medial Temporal Lobe Findings

Left medial temporal lobe deficits in patients with
schizophrenia, relative to comparison subjects, were
found in nine of 13 studies, and right medial temporal
lobe deficits in patients were reported in only three of 13
studies. (The remaining studies examined white matter
only.) Our definition of the medial temporal lobe did not
include the parahippocampal gyrus, which was reported
to be relatively smaller in the left hemisphere in schizo-
phrenia patients in seven of the 15 studies and relatively
smaller in the right hemisphere in three of the 15 studies.
This finding of a lateralized deficit in brain matter sup-
ports previous suggestions of less hemispheric domi-
nance in brains of patients with schizophrenia (52).

Changes in brain density or volume in the limbic region
in patients with schizophrenia have been reported for sev-
eral years (7, 14, 19, 53–55). In a meta-analysis of 58 region
of interest studies, in which a volume of 100% in the com-
parison group was assumed, patients with schizophrenia
were found to have an overall volume of 94% in the left and
right amygdala, 94% in the left hippocampus/amygdala,
95% in the right hippocampus/amygdala, and 93% in the
left and 95% in the right parahippocampus (7). The major-
ity of voxel-based morphometry studies reviewed here re-
ported volume deficits in similar areas, including the left
amygdala, left hippocampus, and left parahippocampal
gyrus. Some postmortem studies agree with these imaging
findings (55–58), although others found no volume
changes in patients with schizophrenia (12, 59, 60). Several
postmortem studies suggested that researchers who have
reported volume deficits in the amygdala and hippocam-
pus may actually have misinterpreted volume changes in
the parahippocampal gyrus, because the boundaries of the
hippocampus and amygdala are difficult to determine (61–

---

### Table 2. Summary of Significant Findings of Reduced Volume of the Superior Temporal Gyrus in Schizophrenia in Voxel-Based Morphometry Studies Published to May 2004

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study</th>
<th>Peak Coordinates</th>
<th>Smoothing Kernel Size (mm)</th>
<th>Statistical Analysis Used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>x     y     z</td>
<td>x     y     z</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Marcelis et al., 2003 (22)</td>
<td>-4.8</td>
<td>18.6</td>
<td>6.6</td>
</tr>
<tr>
<td>2</td>
<td>Sigmundsson et al., 2001 (24)</td>
<td>-29</td>
<td>-6</td>
<td>-26</td>
</tr>
<tr>
<td>3</td>
<td>Kubicki et al., 2002 (18)</td>
<td>-18</td>
<td>-39</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Job et al., 2002 (32a)</td>
<td>-13.86</td>
<td>-0.84</td>
<td>-16.78</td>
</tr>
<tr>
<td>5</td>
<td>Kubicki et al., 2002 (18)</td>
<td>-18</td>
<td>-39</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Shapleske et al., 2002 (23)</td>
<td>-20</td>
<td>-8</td>
<td>-156</td>
</tr>
<tr>
<td>7</td>
<td>Hulshoff Pol et al., 2001 (21)</td>
<td>-15</td>
<td>-9</td>
<td>-15</td>
</tr>
<tr>
<td>8</td>
<td>Salgado-Pineda et al., 2003 (33)</td>
<td>-32</td>
<td>-20</td>
<td>-16</td>
</tr>
<tr>
<td>9</td>
<td>Wright et al., 1999 (31)</td>
<td>-18</td>
<td>-6</td>
<td>-12</td>
</tr>
<tr>
<td>10</td>
<td>Moorhead et al., 2004 (33)</td>
<td>-27</td>
<td>2</td>
<td>-17</td>
</tr>
<tr>
<td>11</td>
<td>Moorhead et al., 2004 (33)</td>
<td>-21</td>
<td>-24</td>
<td>-9</td>
</tr>
</tbody>
</table>

* a The authors used a small-volume correction on the data for the temporal lobe.
In imaging research, the limbic regions could be susceptible to partial volume effects in the adjacent temporal or inferior horns of the lateral ventricles, as discussed previously. Nevertheless, the overall consistency of the MRI findings, including the findings of the automated voxel-based morphometry studies, implicates the limbic regions in schizophrenia. Future voxel-based morphometry research should use a small smoothing kernel, small-volume

63). In imaging research, the limbic regions could be susceptible to partial volume effects in the adjacent temporal or inferior horns of the lateral ventricles, as discussed previously. Nevertheless, the overall consistency of the MRI

findings, including the findings of the automated voxel-based morphometry studies, implicates the limbic regions in schizophrenia. Future voxel-based morphometry research should use a small smoothing kernel, small-volume

TABLE 3. Summary of Significant Findings of Reduced Volume of the Medial Temporal Lobe in Schizophrenia in Voxel-Based Morphometry Studies Published to May 2004

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study</th>
<th>Left</th>
<th>Right</th>
<th>Smoothing Kernel Size</th>
<th>Statistical Analysis Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Shapleske et al., 2002 [23]</td>
<td>–42 –3 8</td>
<td>37 19 –19</td>
<td>4.2</td>
<td>Cluster</td>
</tr>
<tr>
<td>13</td>
<td>Moorhead et al., 2004 [37]a</td>
<td>–62 –3 0</td>
<td>60 –4 5</td>
<td>12</td>
<td>Voxel peak height</td>
</tr>
</tbody>
</table>

a The authors used a small-volume correction on the data for the amygdala and hippocampus.
Specific Brain Regions in Patients With Schizophrenia

Reduced superior temporal gyrus volume in schizophrenia patients was the most frequent bilateral finding across the 15 studies; 57% of studies found this deficit in the left hemisphere, and 50% found it in the right hemisphere. This finding is supported by reports of reduced volumes of temporal lobe structures in recent MRI and postmortem investigations in the anterior (64–66), posterior (49, 67, 68), and total area of the superior temporal gyrus (69, 70). Other groups have not found morphological abnormalities in the superior temporal gyrus in schizophrenia (71, 72). In some studies, the deficit is reported to be lateralized to the left (64, 66, 68, 69, 73), but in other studies, the difference is bilateral (49, 65, 70, 71). In one study, the volume deficit in the planum temporale was bilateral, and there was loss of asymmetry of the surface area (67). In addition, two reviews (4, 6) reported superior temporal gyrus abnormalities in 67%–81% of studies (when the analysis was not restricted to gray matter), compared with the 50%–57% reported here, suggesting that the sensitivity of voxel-based morphometry may not be equivalent to that of region of interest studies.

The range of peak superior temporal gyrus coordinates reported by the studies included in this review is shown in Table 2. The locations of these coordinates include areas of the left arcuate fasciculus (24), Heschl's gyrus (18), uncinate fasciculus (23, 36), Brodmann's area 21 (37), and Brodmann's area 22 (22, 23, 35). In the right hemisphere, the coordinates are located in the right arcuate fasciculus (24), primary auditory cortex (18), Brodmann's area 38 (35, 37), Brodmann's area 22 (36, 37), and Brodmann's area 41 (31). The variety of volume deficits across the region should be considered in light of the known morphology of the superior temporal gyrus: not only is it subdivided both structurally and functionally, but it also has potentially varying and abundant connections with the parietal, prefrontal, and superior temporal regions. These connections have been defined by tracer projections in rhesus monkeys (73).

Morphometric abnormalities in this region have been linked to symptom type and severity. For instance, Barta et al. (65) found a significant correlation between the severity of auditory hallucinations and volume deficits in the left superior temporal gyrus. Kim et al. (74) found that abnormalities of the anterior and posterior regional matter in the superior temporal gyrus were associated with psychotic and negative symptoms, depending on the side of the brain in which the abnormalities occurred. In addition to comparing voxel-based morphometry results for patients and healthy subjects, Shapleske et al. (23) compared results for patients with and without a history of verbal auditory hallucinations. They found that patients with a history of hallucinations had volume deficits in the left insula and adjacent temporal lobe. These studies demonstrated that more specific morphological abnormalities can be illuminated in schizophrenia when patients are grouped according to specific prominent symptoms. Future voxel-

**Figure 6. Percentage of Voxel-Based Morphometry Studies (N=15) With Significant Findings of Volume Deficits in Specific Brain Regions in Patients With Schizophrenia, by Size of the Smoothing Kernel Used in the Study**

<table>
<thead>
<tr>
<th>Region of Interest</th>
<th>Size of Smoothing Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right middle temporal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left middle occipital gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right postcentral gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right middle occipital gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right precentral gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right posterior cingulate gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left posterior cingulate gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left inferior parietal lobe</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right fusiform gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right cerebellum</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left middle temporal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left inferior temporal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left precentral gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right medial orbital gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left medial orbital gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right inferior parietal lobe</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right rectus gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left rectus gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right inferior temporal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left anterior cingulate gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right uncinate fasciculus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left anterior internal capsule</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right arcuate fasciculus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left arcuate fasciculus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right anterior internal capsule</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right cuneus (precuneus)</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left cuneus (precuneus)</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left middle frontal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right middle frontal lobe</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right postcentral gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left postcentral gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right fusiform gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right inferior frontal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left uncinate fasciculus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right middle frontal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right parahippocampal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right caudate nucleus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left caudate nucleus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left thalamus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right insula</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left insula</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left inferior frontal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right medial frontal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Right anterior cingulate gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right superior temporal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left medial frontal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left inferior temporal gyrus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left superior temporal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Right thalamus</td>
<td>4–8 mm (N=7)</td>
</tr>
<tr>
<td>Left parahippocampal gyrus</td>
<td>10–12 mm (N=7)</td>
</tr>
<tr>
<td>Left medial temporal lobe</td>
<td>4–8 mm (N=7)</td>
</tr>
</tbody>
</table>

* The total Ns for the percentages do not include studies in which the specific region was not investigated. Two studies by Hulshoff Pol et al. (21, 39) (study 8 and study 15 in Table 1) were considered as one study in the calculation of percentages.

corrections, careful segmentation, and validation with region of interest analysis to further clarify these limbic region findings.
based morphometry studies should take this approach to reduce patient heterogeneity and further delineate the functional deficits associated with morphological abnormalities in the superior temporal gyri.

Conclusions

We reviewed 15 studies that used voxel-based morphometry to identify structural abnormalities in the brains of patients with schizophrenia. Of the 50 regions investigated, the areas reported to be significantly reduced in schizophrenia patients in more than 50% of the studies were the left superior temporal gyrus and the left medial temporal lobe. Fifty percent of the studies reported volumetric decreases in the left parahippocampal gyrus, right superior temporal gyrus, left inferior frontal gyrus, and left medial frontal gyrus.

It is interesting to note that at least 31% of the studies did not find volume deficits in the left medial temporal lobe. Several explanations could account for this result. First, the methods varied considerably between studies. The most critical factor may be the size of the smoothing kernel. We found that the discriminatory power for detecting changes in small areas was lower with smoothing kernels of greater width. A second factor is heterogeneity of the disease process. Voxel-based morphometry produces clearer results in studies where the patient group has been carefully chosen to be homogeneous. For example, Krams et al. (75) studied patients with mirror movements due to Kallmann's syndrome, Abell et al. (76) studied a group of subjects with Asperger's syndrome who were selected on the basis of their performance on tests of theory of mind, and Watkins et al. (77) studied members of a four-generation family with an inherited speech disorder whose results on tests of speech had no overlap with the results for nonaffected members of the family. Thus, it is likely that more consistent abnormalities in schizophrenia would be reported if voxel-based morphometry were applied to clearly defined patient groups characterized by symptom profile, age, and disease duration.

References


Am J Psychiatry 162:12, December 2005

http://ajp.psychiatryonline.org
VOXEL-BASED MORPHOMETRY


42. Bagary MS, Symms MR, Barker GJ, Mutsatsa SH, Joyce EM, Ron MA: Gray and white matter brain abnormalities in first-episode schizophrenia inferred from magnetization transfer imaging. Arch Gen Psychiatry 2003; 60:779–788


Messages From “Fran”

I got a phone message the other night from a person I do not know. The man represented himself as the partner of a patient I had never met. I shall call her “Fran.”

Fran had telephoned me over the past several years: sometimes weekly, sometimes only once every few months. On occasion, her messages praised my telephone voice and my outgoing message, which she told me were far, far better than those of my colleagues in my on-call group. At times, she left messages pleading with me to see her. In other messages, she would excoriate me for being cold and indifferent to her. Once she said she was leaving her home, strongly hinting that she planned to kill herself. In that message, she simply wished to say goodbye and to have me know what was to happen to her. On some occasions, her message would be limited to her sincerely promising that she would never bother me again with more messages. Moreover, I suspect, many of the mysterious hang-ups I have gotten, where no messages were left, originated with her.

Once she called at a time when she likely believed I would not be in my office. Because I was still there, I lifted the phone and introduced myself. I sensed that she felt both surprise and embarrassment.

I am one of six members of a psychiatric on-call group. We cover each other’s practices on weekends and vacations. One member, Charles, had seen Fran at his office over a period of several years, before she moved many hundreds of miles away. Although she occasionally returned to our area on visits and would see him then, he had referred her to a psychiatrist in her new community. My colleague told me that the new psychiatrist seemed to be largely pharmacologically oriented.

It was after her move that she began the habit of calling Charles, me, and several of the other on-call group members. My impression is that, for whatever reason, she telephoned me more frequently and persistently than the others. I would like to believe that this was because of the high quality of my telephone voice. I cannot confidently assert the superiority of my outgoing message because, over the years, the outgoing messages that each of us leaves on our answering machines have come largely to resemble one another.

When the calls to me first began, they were presented as being efforts to reach Charles, for whom I was covering. On those early occasions, I would return Fran’s calls, listen to her complaints—about him, about her medications, about the on-call group, about both her old and new psychiatrist, about me—and then I would speak to her about my not contacting Charles and about her need to work with her new psychiatrist. Eventually, I chose not to return her calls.

The message from the man I do not know but who represented himself as Fran’s partner, the patient I had never met, started by saying that Fran would want him to call. She was now in an intensive care unit and terminally ill. Barring a miracle, she would soon be dead. After a pause, he added, “So no more calls…. She got a great deal out of keeping in touch with all of you up there, so she would have wanted you to know this.”

There is a possibility that the message he left was not factual, but the emotionality with which it was delivered satisfied me that, at least, the closeness to death was likely true. In any case, his call to me and the message it contained stirred up many feelings in me: loss, sadness, guilt for having decided—for the most part—not to respond to her messages with return calls, irritation upon recalling the times I had been perplexed by

“We should never make the error of underestimating the power of connection.”
and made anxious by her messages, and annoyance because of needing to sort out if I should respond to her at all or in some different way.

Even when I know my patients well and see them on a regular basis, I can be puzzled by what leads them to feel “better.” If they, in my or their perception, appear to benefit from our work together, I often cannot be clear as to what has allowed their improvement to occur. Obviously, in a relationship as tenuous as the one with Fran, it is harder still to imagine the forces that drive it.

However, what comes through clearly in this situation is the powerful regard with which patients so often hold us. There is such a strong relational pull that it causes me to marvel at the power of the interpersonal.

Given that Fran was never my patient, given her seductiveness and her tirades, the sporadic nature of her calls, and the aggravation often associated with them, it would have been easy to overlook what her partner’s message meant. Whether it was clearly helpful or not to Fran—or to me—we had a relationship, and it was significant to both of us. We may not know or understand exactly what uses one might make of the relationship, but we should never make the error of underestimating the power of connection.

Four days after the first message from Fran’s partner, I received a second: Fran had died. I was asked to relay the message to Charles.

EDMUND C. LEVIN, M.D.

Address correspondence and reprint requests to Dr. Levin, 2424 Dwight Way, #2, Berkeley, CA 94704; eclevin@earthlink.net (e-mail).
Chekhov's stories and plays show deep psychological insight. He realized that scientific rationalism could not give us the answers to some important questions, such as, "What is the meaning of life?" Thus, he presented man as a victim in an absurd world.

Chekhov graduated from medical school in 1884 and started working as a physician in a suburban Moscow hospital. Eight years later, he purchased the rundown Melikhovo estate 70 km south of Moscow. He lived there until his declining health due to lung tuberculosis forced him to move to the milder climate of Yalta.

In Melikhovo, Chekhov saw hundreds of patients, made more than a thousand house calls, and fought against cholera and illiteracy (1). Although he practiced medicine by day, at night he wrote masterpieces—stories and two magnificent plays: The Seagull and Uncle Vanya. Although Chekhov was not a wealthy man, he did not charge his patients. In 1889, he made a journey across Siberia to Sakhalin Island, where Russian convicts were kept. The hardships of that journey and his 3-month sojourn on the island were frightful. His report of the trip includes a moving description of brutal beatings he witnessed that made such an impression on the Russian public that corporal punishment was abolished for women in 1897 and for men in 1904 (2).

Chekhov created hundreds of characters who show weakness, passivity, and ineffectiveness. His stories (Tolstoy compared them to impressionist paintings) and his best plays (including Three Sisters and The Cherry Orchard) still strongly appeal to readers and theatergoers the world over. Chekhov's numerous characters include an array of more than 30 medical doctors who are hindered by various problems and poor working conditions. In his story titled Ward No. 6, set in the psychiatric division of a provincial hospital, Dr. Ragin encounters a sparkling paranoid man named Gromov who has been confined for proclaiming that truth and justice must one day triumph. Ironically, Dr. Ragin is trapped by his superiors in his own ward, and after a beating by a nurse, he dies of a stroke. When the novelist Nikolay Leskov read this masterpiece, he said, "Ward No. 6 is Russia" (3).

Chekhov clearly showed that the symbiosis of the muse and Æsculapius may increase the writer's capabilities. He wrote to a friend, "Medicine is my lawful wife, literature my mistress. When I tire of the one, I spend the night with the other. As long as it does not become a regular habit, it is not humdrum and neither of them suffers from my infidelity" (4). Thanks to his literary talent and his psychological approach, Chekhov improved both the modern story and the modern play.

References


RAJKO IGIĆ, M.D., Ph.D.

Address correspondence and reprint requests to Dr. Igić, John H. Stroger, Jr., Hospital of Cook County, 637 S. Wood, Rm. 427 DX, Chicago, IL 60612; r.igic@excite.com (e-mail). Photograph published with permission of the Granger Collection, New York.
Identifying Dieters Who Will Develop an Eating Disorder: A Prospective, Population-Based Study

Helen A. Doll, M.Sc.
Beverley A. Davies, B.Sc.

**Objective:** The aims of the study were to identify the characteristics of the dieters most at risk of subsequently developing an eating disorder and to evaluate the feasibility of using a brief questionnaire to identify such dieters in advance.

**Method:** A general population cohort of 2,992 young women who were dieting was identified. On four occasions over the subsequent 2 years, this cohort was sent a questionnaire concerning eating habits and attitudes. Participants whose responses suggested that they had developed an eating disorder were interviewed to establish their true case status. The baseline questionnaires of those who did and did not subsequently develop an eating disorder were compared to identify features that predicted future case status.

**Results:** One hundred four of the dieters developed an eating disorder of clinical severity during the 2 years of follow-up. Their baseline questionnaire scores differed in many respects from those who had not developed an eating disorder. Items associated with developing an eating disorder were selected by using three different statistical methods. A simple case-predicting instrument based on one of five items scoring above an optimal cut point had a sensitivity of 71% and a specificity of 72% (overall efficiency of 72%).

**Conclusions:** Dieters who will develop an eating disorder within the next 2 years have distinctive features. It is feasible to identify them in advance with reasonable efficiency with a brief questionnaire. This questionnaire could be incorporated into routine health assessments, thereby identifying those at high risk.

Eating disorders are a major source of physical and psychosocial morbidity among young women (1–3). It would be valuable to be able to detect those at most risk of developing an eating disorder, either to prevent the disorder from developing or to be able to start treatment early. Because clinical experience and research evidence have indicated that eating disorders commonly begin with behavior that resembles normal dieting (1, 4, 5), young women who are dieting constitute an important high-risk group, although only a small minority will develop an eating disorder (6, 7).

The overall aim of the present study was to develop a means of identifying young female dieters at most risk of subsequently developing an eating disorder. There were two specific aims: to determine whether there are features that predict the development of an eating disorder and to evaluate the feasibility of using a brief questionnaire to identify those at most risk.

**Method**

**Design**

A cohort of young women who were dieting but did not have an eating disorder were asked to complete a validated measure of eating disorders features. These women were then recontacted on four occasions 6 months apart and asked to complete additional copies of the same measure. At each occasion, the women whose responses suggested that they had developed an eating disorder were interviewed to confirm their diagnostic status. The initial questionnaires of the participants who had and had not developed an eating disorder were compared to identify whether there were features that predicted the subsequent development of an eating disorder and, if so, whether they could be used to construct an efficient case-predicting instrument.

**Participants and Procedures**

Practically all of the British population is registered with a local family physician (8). The names and addresses of the great majority of the local population are therefore to be found on family physicians’ lists. A good way of obtaining a general population study group is to contact the people on these lists. This procedure was used in the present study, the sampling frame being the patient registers of 44 family practices located in urban and rural parts of central England. The names and addresses of 17,144 women, ages 16 to 23 years inclusive, were obtained in this manner. A letter was sent to each, inviting her to take part in the study and enclosing a baseline self-report questionnaire with a stamped, self-addressed envelope. Ten thousand seventy-seven of these questionnaires (58.8%) were returned.

It is difficult to identify the true response rate in a study using this recruitment method because family physicians’ registers are invariably overinclusive, particularly with regard to this age group. This is because the registers include the names of people who have subsequently moved away but have not yet informed their physicians. Therefore, we could not conclude that nonresponse was indicative of the participants, having received the questionnaire and chosen not to take part. For this reason, we undertook a detailed pilot study using two other family practice case...
Table 1. Baseline Demographic and Clinical Characteristics of Young Female Dieters Who Did and Did Not Develop an Eating Disorder

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total Group (N=2,992)</th>
<th>Subjects Without an Eating Disorder (N=2,888)</th>
<th>Subjects With an Eating Disorder (N=104)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Age (years)</td>
<td>19.83</td>
<td>2.36</td>
<td>19.84</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>165.54</td>
<td>6.69</td>
<td>165.49</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>63.64</td>
<td>11.05</td>
<td>63.66</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>23.22</td>
<td>3.74</td>
<td>23.22</td>
</tr>
<tr>
<td>Dietary restraintb</td>
<td>2.74</td>
<td>1.01</td>
<td>2.74</td>
</tr>
<tr>
<td>Severity of eating disorder featuresd</td>
<td>2.63</td>
<td>1.04</td>
<td>2.63</td>
</tr>
</tbody>
</table>

b Restraint subscale score from the Eating Disorder Examination Questionnaire (16).
cSignificantly different from the subjects without an eating disorder (t=3.40, df=2835, p<0.01).
dGlobal subscale score from the Eating Disorder Examination Questionnaire (16).
eSignificantly different from the subjects without an eating disorder (t=6.23, df=2843, p<0.001).

The questionnaire of the 10,077 respondents were coded to identify those who were currently dieting. This was done by selecting those whose scores were in the top tertile on a measure of dietary restraint embedded within the questionnaire and then excluding those who reported either a current or past eating disorder or a medical condition or treatment known to affect eating habits or weight (e.g., thyroid disease, pregnancy, steroid medication). In this way, we identified 2,992 young women who were currently dieting but did not have an eating disorder or any of the other exclusion criteria.

These women formed the cohort that was followed up on four occasions over 2 years. This involved sending them more copies of the questionnaire at 6-month intervals and, if their responses suggested that they might have developed an eating disorder, asking if we might interview them. Those who agreed were then assessed (usually in their homes) with the Eating Disorder Examination interview. At each stage in this process, the participants were informed that they could withdraw from the study if they wished, and written informed consent was obtained from all of those who were interviewed. The study was approved by the relevant human subjects committee for each locality.

Measures

Eating Disorder Examination. The Eating Disorder Examination interview (10) was used to make diagnoses of eating disorders. The Eating Disorder Examination is widely regarded as the optimal instrument for this purpose (11–13). As part of the assessment, the participants were weighed by using calibrated portable scales, and their height was measured (thereby allowing their body mass index to be calculated: kg/m²). With this information, it was possible to apply operational definitions of the DSM-IV diagnoses of anorexia nervosa and bulimia nervosa based on the Eating Disorder Examination interview (10). Diagnosis of the other DSM-IV eating disorder category, eating disorder not otherwise specified, was made by two experienced clinicians (C.G.F. and Z.C.) after they were briefed in detail about the participant’s clinical state. The participants were given this diagnosis if they did not meet the diagnostic criteria for anorexia nervosa or bulimia nervosa yet clearly had an eating disorder comparable in severity to those seen among patients receiving treatment at eating disorder clinics. These judgments were made independently and blind to the participants’ identities and follow-up points. There were few disagreements between the two clinicians’ judgments, and these invariably concerned cases of threshold severity. Each was resolved upon discussion, a rule being not to make a diagnosis of eating disorder not otherwise specified if either clinician remained uncertain about the participant’s case status. The same procedure has been used in previous studies of eating disorders and their course (14, 15).

Eating Disorder Examination Questionnaire. The Eating Disorder Examination Questionnaire (16) is a self-report version of the Eating Disorder Examination interview. It has a 4-week time period, and it assesses the features of eating disorders. Four subscales may be derived from the instrument, together with a global score. The dietary restraint subscale was used in the present study as the measure of dieting. This subscale has five items comprising the following questions:

1. Have you been deliberately trying to limit the amount of food that you eat to influence your shape or weight (whether or not you have succeeded)?
2. Have you gone for long periods of time (8 waking hours or more) without eating anything at all in order to influence your shape or weight?
3. Have you tried to exclude from your diet any foods that you like in order to influence your shape or weight (whether or not you have succeeded)?
4. Have you tried to follow definite rules regarding your eating (for example, a calorie limit) in order to influence your shape or weight (whether or not you have succeeded)?
5. Have you had a definite desire to have an empty stomach with the aim of influencing your shape or weight?

Each item refers to the past 28 days and is rated from 0 to 6 where 0=none days, 1=1–5 days, 2=6–12 days, 3=13–15 days, 4=16–22 days, 5=23–27 days, and 6=every day. In common with the other subscales (and the global score), the dietary restraint subscale score is the mean of the items rated.

Performance on the Eating Disorder Examination Questionnaire has been extensively studied, both in isolation and in comparison with the Eating Disorder Examination interview (16–24). The findings indicate that in many respects the Eating Disorder Examination Questionnaire is a reasonable substitute for the Eating Disorder Examination interview. This is particularly true of the dietary restraint subscale. Questions were added to the Eating Disorder Examination Questionnaire concerning the exclusion criteria just noted.

http://ajp.psychiatryonline.org
dictor and optimal cut point chosen in terms of their sensitivity. A form of recursive partitioning that considers at each step all possible algorithm to identify future cases. Signal detection analysis is performed to identify, through the creation of a decision tree, the most sensitive and specific function, which features best discriminated future cases from future noncases. Unstandardized and standardized discriminant function results were similar for both undichotomized and dichotomized predictor variables (cut off at the optimal cut point); the latter are presented for simplicity and for consistency with the results from the signal detection analysis. Cox proportional hazards regression analysis. This analysis was performed to determine which features were independently predictive of time to the development of an eating disorder. The results are presented as hazard ratios with 95% confidence intervals (CIs). Both graphical methods and the use of Schoenfeld residual plots (25) indicated that the proportional hazards assumption was reasonable. Linear discriminant function analysis. This analysis was performed to identify, through the creation of a discriminant function, which features best discriminated future cases from future noncases. Unstandardized and standardized discriminant function coefficients are presented. The most efficient threshold on the function score in terms of sensitivity and specificity was identified with the use of values assigned to the costs of false positives versus false negatives and with the use of a variety of prior probabilities, including a default of 50:50 (and with a 70:30 case versus noncase ratio selected a priori as likely to generate the most useful instrument). Signal detection analysis. This analysis was used to determine, through creation of a decision tree, the most sensitive and specific algorithm to identify future cases. Signal detection analysis is a form of recursive partitioning that considers at each step all possible predictors (at every possible cut point), with the optimum predictor and optimal cut point chosen in terms of their sensitivity and specificity. Thus, the subjects with missing data were excluded on a node-by-variable basis. The cut point was set in advance because no optimal predictor was associated with an outcome at p<0.01. The merits of signal detection analysis were summarized by Agras and colleagues (26). The analyses were performed with varying emphases on sensitivity versus specificity when 50:50 was the default. Because the avoidance of false negatives (i.e., increased sensitivity) was viewed as the priority, 70:30 was selected a priori as likely to generate the most useful instrument.

Results
Development of Eating Disorders Over Follow-Up
The baseline characteristics of the 2,992 young female dieters are shown in Table 1. The response rates at each follow-up point were 76%, 68%, 63%, and 60%, respectively. The amount of missing data was modest, with the great majority of items from the Eating Disorder Examination Questionnaire completed by over 97% of the participants. Those who complied with follow-up had a lower baseline global score on the Eating Disorder Examination Questionnaire than those who did not (mean=2.50, SD=1.00, versus mean=2.74, SD=1.07, respectively) (t=6.24, df=2430.5, p<0.001) (mean difference=0.24, 95% confidence interval [CI]=0.17–0.32, effect size=0.23).

Over the follow-up, the Eating Disorder Examination Questionnaire responses of 457 participants suggested that they might have developed an eating disorder. These participants were therefore asked if they would be willing to be interviewed. Three hundred ninety-two (85.8%) agreed, of whom 104 (26.5%, 3.5% of the total study group) (95% CI=2.8%–4.2%) were found to have a DSM-IV eating disorder on the Eating Disorder Examination. Nearly forty percent (39.5%) of the cases developed within the first 6 months of follow-up, with the remainder developing at a slowly decreasing rate across the remaining 18 months (23%, 20%, and 17%, respectively). The eating disorder diagnoses of the 104 subjects were as follows: 10 with anorexia nervosa (9.6%), 19 with bulimia nervosa (18.3%), and 75 with eating disorder not otherwise specified (72.1%). Table 1 shows the baseline characteristics of those who did and did not subsequently develop an eating disorder, and Table 2 shows the characteristics of the 104 subjects with eating disorders.

An item-by-item comparison of the baseline responses on the Eating Disorder Examination Questionnaire of the participants who developed an eating disorder with those

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Subjects With Anorexia Nervosa (N=10)</th>
<th>Subjects With Bulimia Nervosa (N=19)</th>
<th>Subjects With Eating Disorder Not Otherwise Specified (N=75)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity of eating disorder features*</td>
<td>3.31 1.14</td>
<td>4.48 0.81</td>
<td>3.59 0.99</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>44.10 4.36</td>
<td>63.21 8.01</td>
<td>65.49 13.19</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>16.22 1.02</td>
<td>22.10 2.39</td>
<td>23.49 4.68</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>19.40 2.67</td>
<td>20.53 2.55</td>
<td>20.42 2.46</td>
</tr>
</tbody>
</table>

* Global subscale score from the Eating Disorder Examination Questionnaire (16).
who did not reveal that the former group had significantly \( p < 0.05 \) higher scores on most of the items on the Eating Disorder Examination Questionnaire (31 of 44, 70.4%), as well as a higher overall score. A large number of items on the Eating Disorder Examination Questionnaire were therefore available for inclusion in a case-predicting questionnaire.

**Development of a Case-Predicting Questionnaire**

The results across all three statistical methods were consistent, with the same five items from the Eating Disorder Examination Questionnaire plus body mass index selected for inclusion in the potential case-finding questionnaire (Table 3). Two additional items from the Eating Disorder Examination Questionnaire were selected with the signal detection method. Increasing numbers of scores above each item’s optimal cut point were associated with an increased likelihood of developing an eating disorder. A simple screening instrument based on the presence of one single item scoring above its cut point had a sensitivity of 79% and a specificity of 52%, whereas one based on two items had a sensitivity of 62% and specificity of 76% (overall efficiency=75%).

With Cox regression analysis, high scores on five items from the Eating Disorder Examination Questionnaire plus low body mass index were identified as predictive of the time for developing an eating disorder (Table 4). Jackknife validity analysis (27) indicated little bias in the model. With linear discriminant function analysis, high scores on four items from the Eating Disorder Examination Questionnaire plus low body mass index discriminated between future case and noncase status. A simple screening instrument based on one of these five items having a score above the optimal cut point had a sensitivity of 71% and a specificity of 72% (overall efficiency=72%). Similarly, decision rules and specification of prior probabilities (70:30 as optimum for sensitivity versus specificity) selected a test \((\text{threshold}=0)\) that would correctly identify 71% of future cases \((\text{specificity}=72.8%, 95\% \text{ CI}=71.0\%–74.6\%)) with a 28% rate of false positives \((\text{specificity}=72.8%, 95\% \text{ CI}=70.4\%–73.8\%)) and its likelihood ratio was 2.54. Jackknife “leave one out” cross-validation (27) indicated that the performance of the test was excellent, with the values of the sensitivity and specificity being unchanged.

On signal detection analysis, with a 70:30 emphasis on sensitivity versus specificity, seven items from the Eating Disorder Examination Questionnaire plus body mass index were selected. The item providing the best cut point was the frequency of self-reported binge eating, followed

---

**TABLE 3. Items Selected From the Eating Disorder Examination Questionnaire for a Case-Predicting Questionnaire and Their Association With the Development of an Eating Disorder**

<table>
<thead>
<tr>
<th>Item and Optimal Cut Point for Signal Detection Analysis</th>
<th>N</th>
<th>Subjects With an Eating Disorder</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>( \chi^2 )</td>
</tr>
<tr>
<td>Objective bulimic episodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3 episodes/28 days</td>
<td>2,325</td>
<td>63</td>
<td>2.7</td>
</tr>
<tr>
<td>≥3 episodes/28 days</td>
<td>333</td>
<td>32</td>
<td>9.6</td>
</tr>
<tr>
<td>Laxative misuse/self-induced vomiting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 episodes/28 days</td>
<td>2,827</td>
<td>95</td>
<td>3.4</td>
</tr>
<tr>
<td>≥2 episodes/28 days</td>
<td>36</td>
<td>8</td>
<td>22.2</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥19</td>
<td>2,628</td>
<td>81</td>
<td>3.1</td>
</tr>
<tr>
<td>&lt;19</td>
<td>180</td>
<td>19</td>
<td>10.6</td>
</tr>
<tr>
<td>Secret eating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;13 instances/28 days</td>
<td>2,742</td>
<td>86</td>
<td>3.1</td>
</tr>
<tr>
<td>≥13 instances/28 days</td>
<td>128</td>
<td>17</td>
<td>13.3</td>
</tr>
<tr>
<td>Fear of losing control over eating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;23 instances/28 days</td>
<td>2,454</td>
<td>70</td>
<td>2.9</td>
</tr>
<tr>
<td>≥23 instances/28 days</td>
<td>419</td>
<td>33</td>
<td>7.9</td>
</tr>
<tr>
<td>Desire to have an empty stomach</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;16 instances/28 days</td>
<td>2,254</td>
<td>62</td>
<td>2.8</td>
</tr>
<tr>
<td>≥16 instances/28 days</td>
<td>613</td>
<td>40</td>
<td>6.5</td>
</tr>
<tr>
<td>Preoccupation with food</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 instances/28 days</td>
<td>2,256</td>
<td>61</td>
<td>2.7</td>
</tr>
<tr>
<td>≥6 instances/28 days</td>
<td>623</td>
<td>42</td>
<td>6.7</td>
</tr>
<tr>
<td>Preoccupation with shape or weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;13 instances/28 days</td>
<td>2,367</td>
<td>67</td>
<td>2.8</td>
</tr>
<tr>
<td>≥13 instances/28 days</td>
<td>503</td>
<td>36</td>
<td>7.1</td>
</tr>
<tr>
<td>Total of previous items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1,479</td>
<td>22</td>
<td>1.5</td>
</tr>
<tr>
<td>1</td>
<td>1,686</td>
<td>18</td>
<td>2.6</td>
</tr>
<tr>
<td>2</td>
<td>357</td>
<td>19</td>
<td>5.3</td>
</tr>
<tr>
<td>3</td>
<td>217</td>
<td>22</td>
<td>10.1</td>
</tr>
<tr>
<td>4</td>
<td>123</td>
<td>15</td>
<td>12.1</td>
</tr>
<tr>
<td>≥5</td>
<td>55</td>
<td>8</td>
<td>14.5</td>
</tr>
</tbody>
</table>

\(^a\) Selected for signal detection analysis only.

\(^b\) Linear trend.
by eating in secret, low body mass index (<19), preoccupation with food and eating, desire to have an empty stomach, frequency of purging, fear of losing control over eating, and preoccupation with shape or weight. Use of the resulting decision tree (available upon request from the first author) correctly identified 70% of future cases (sensitivity=69.5%, 95% CI=60.2%–78.8%), with a 25% rate of false positives (specificity=75.1%, 95% CI=73.4%–76.8%). Its overall efficiency was 74.9% (95% CI=73.3%–76.6%), and its likelihood ratio was 2.79. Bootstrap validation (27) confirmed the robustness of the decision tree because, although none of the trees was exactly the same (as would be expected from the way the trees are constructed), they incorporated broadly similar items with similar cut points. Moreover, the trees had comparable levels of sensitivity and specificity.

Finally, interactions between the predictor variables and age were investigated. No factor was found to have a significant interaction with age in the Cox regression model, nor did age feature in the signal detection analysis.

**Discussion**

The two aims of the present study were to identify the characteristics of the dieters most at risk of subsequently developing an eating disorder and to evaluate the feasibility of using a brief questionnaire to identify such dieters in advance. This necessitated recruiting a large community-based group of young women, identifying a subgroup who were currently dieting, and then following up on them at repeated intervals to see who had developed an eating disorder and who had not. These steps were accomplished, and both aims were achieved. It should be noted that the dieters were identified on the basis of a high score on a measure of dietary restraint—that is, a measure of attempting to restrict food intake—rather than a measure of actual dietary restriction. There is limited evidence that the two are separable (28). Both dietary restraint and dietary restriction are thought likely to increase the risk of developing an eating disorder through cognitive and physiological mechanisms, respectively.

As expected, only a small proportion of the dieters developed an eating disorder, and most were cases of eating disorder not otherwise specified. This high proportion of subjects with eating disorder not otherwise specified relative to subjects with anorexia nervosa and bulimia nervosa parallels the distribution of the three diagnoses in most clinical settings (29). Not surprisingly, the dieters who developed an eating disorder had more disturbed eating habits and attitudes at recruitment than those who did not, as reflected in their higher scores on the global subscale of the Eating Disorder Examination Questionnaire. Also unsurprising is the fact that several of the features that best discriminated future cases from future noncases were features that are seen in people with eating disorders, albeit at a more severe level. Other ominous features were less predictable: namely, eating in secret; preoccupation with food, eating, shape, or weight; fear of losing control over eating; and wanting to have a completely empty stomach.

There was substantial overlap in the items on the Eating Disorder Examination Questionnaire selected by the three different statistical methods. The efficiencies of the instruments derived from the discriminant function and the decision tree analyses were similar and the same as those from a simple case-predicting instrument derived from one of five items scoring above the optimal cut point (sensitivity=71%, specificity=72%). Both methods would involve completing a brief questionnaire that would identify about 70% of future cases.

The strengths of the present study include the size of the cohort, which resulted in a sufficiently large number of dieters being studied for 104 cases of eating disorder cases to develop; the method of recruitment, which circumvented certain of the selection biases that would have resulted had we advertised for dieters; and the 2-year follow-up, which provided sufficient time for many cases to develop. Other strengths include the use of clinical methods and thresholds to define case status and the fact that the core measure, the Eating Disorder Examination Questionnaire, has been well validated and is known to be acceptable to the relevant population. The relatively high rates of response are also of note.
A limitation of the study is its reliance on a measure of eating habits and attitudes to predict future case status instead of also testing the performance of other variables. We decided to focus on eating habits and attitudes for three largely pragmatic reasons. First, given current knowledge about risk factors for eating disorders (5), we thought that eating habits and attitudes were likely to be better predictors of developing an eating disorder than other variables. Second, we thought that young women who are dieting would be more willing to answer questions about their eating habits than questions about other aspects of their lives. Third, we were concerned about overburdening our participants with questions, given that they had little reason to participate. The addition of other variables might have enhanced our ability to predict future case status.

Another limitation is the age and gender of the group because it did not include participants under 16 years old or men. Although an attempt has already been made to develop an instrument for use with younger teenagers (30), the relative rarity of eating disorders among men precludes them from a study of this type. A third limitation is that the participants were followed up for just 2 years, so some later-onset cases will have been missed. Fourth, the fact that those who did not comply with follow-up had higher scores on the Eating Disorder Examination than those who did may have influenced the findings. Finally, because all of our analyses were necessarily exploratory, the desirability of replication must be stressed, although in this instance, opportunities may be limited given the scale required of the research project.

In conclusion, this study has shown that young female dieters who will develop an eating disorder within the next 2 years have distinctive features and that it is potentially feasible to identify them in advance. The case-predicting questionnaire required is brief and easy to complete and score, and its content is acceptable to young women. It could therefore be incorporated into routine health assessments. Women scoring positively could be flagged, with this information made available to inform subsequent consultations. In addition, these women could be the focus of preventive interventions (31, 32). The findings may also be of relevance to young women considering embarking on weight-loss programs because the combination of dietary restraint and the features identified in the present study would confer a higher risk of developing an eating disorder. It could be argued that such women should be informed of this risk and perhaps advised against embarking on such programs. If they chose to go ahead, it would seem wise for them to do so cautiously and perhaps with some external monitoring.

References
Association Between the Insulin-Like Growth Factor 2 Gene (IGF2) and Scores on the Eating Attitudes Test in Nonclinical Subjects: A Family-Based Study

Rachel Bachner-Melman, M.A.  
Ada H. Zohar, Ph.D.  
Lubov Nemanov, M.D.  
Uriel Heresco-Levy, M.D.  
Inga Gritsenko, M.S.  
Richard P. Ebstein, Ph.D.

Objective: An interesting candidate gene for eating disorders is the gene for insulin-like growth factor 2 (IGF2). Located on chromosome 11p15.5, IGF2 is a member of the insulin family of polypeptide growth factors that is involved in development and growth. Consistent with its profile of metabolic actions, an association has been reported between a single nucleotide polymorphism (SNP) in the 3′ untranslated region of the IGF2 gene (ApaI) and body mass index. This investigation extended these studies and investigated the psychological and behavioral implications of this hormone’s impact on metabolism and body composition.

Method: The authors tested nonclinical subjects from 376 families for three IGF2 SNPs and for eating disorders, as reflected in scores on the 26-item Eating Attitudes Test, a self-report questionnaire widely used as a screening instrument.

Results: A highly significant association was observed between the IGF2 ApaI G allele and scores on the Eating Attitudes Test overall and each of its subscales (bulimia, dieting, and oral control). Additionally, a significant association was observed between this polymorphism and body mass index.

Conclusions: The current finding that the IGF2 ApaI G polymorphism, which predisposes to weight gain, may also contribute to the pathology of eating disorders is intriguing. Neurotransmitter modulation of appetitive behavior is the focus of most hypotheses regarding the etiology of severe eating disorders. The current results to some measure challenge this view, and inborn metabolic tendencies to weight gain in some women may trigger constant dieting, which in predisposed individuals eventually leads to severe eating disorders.

The etiology of severe eating disorders—anorexia nervosa, bulimia nervosa, and binge eating disorder—is complex, and multiple influences confer risk for this behavior (1). As noted by Kaye and Strober (2), the stereotyped clinical presentation, sex distribution, and age at onset support the likelihood that there is some biological vulnerability to this disorder. Large twin studies show that the co-twin of a twin affected with anorexia nervosa is 26 times as likely to have a lifetime diagnosis of bulimia nervosa as is the co-twin of an unaffected twin (3), strengthening the notion that genes are important in the etiology of anorexia nervosa and bulimia nervosa.

A familiar hypothesis regarding the genetic basis of complex diseases, such as anorexia nervosa and bulimia nervosa, is that numerous polymorphisms contribute somewhat to risk. If this notion is true, then a cost-effective, complementary strategy for discovering “eating disorder” genes is to study easily accessible and large nonpatient groups rather than solely examining less available and necessarily smaller clinical groups. Common polymorphisms in the nonclinical group are predicted to partially contribute to an eating disorder phenotype that can be readily ascertained by self-report questionnaires relevant to eating behavior and estimations of body mass index. Once such genes are identified in the nonpatients, their role in pathology can be tested in the clinical group exhibiting an extreme phenotype. Studying phenotypes, or endophenotypes, is emerging as a robust strategy in unraveling the genetic architecture of complex diseases (4). An interesting candidate gene that makes “biological sense” for a role in eating disorders and has not yet been examined, to our knowledge, is the gene for insulin-like growth factor 2. This gene, IGF2, is located on chromosome 11p15.5 (5). Insulin-like growth factor 2, also known as somatomedin A, is a single-chain polypeptide that shares an amino acid sequence homology of about 47% with insulin and about 31% with relaxin and with them comprise the insulin family of polypeptide growth factors. Their functions include mediation of growth hormone action, stimulation of growth of cultured cells, stimulation of the action of insulin, and involvement in development and growth. They appear to be autocrine regulators of cell proliferation. Consistent with this profile of metabolic actions, an association has been reported between a single...
nucleotide polymorphism (SNP) in the 3' untranslated region of the IGF2 gene (ApaI) and body mass index (6) that accounts for a small percentage of the population variance. A number of studies appear to confirm this initial finding, including suggestive evidence from a genome scan for obesity (7–10). IGF2 is also expressed in fetal and adult brains (11, 12).

The transmission disequilibrium test (13) and its extension to quantitative traits (14, 15) provide an efficacious procedure for detecting linkage and association, especially in the presence of population admixture. Even though this procedure has been widely used, it is not suitable for groups composed of families with multiple siblings, which decrease the efficiency of the analysis performed. More recently, a new unified approach, the family-based association test (16), to test the association between three family configuration and any type of phenotype (either qualitative or quantitative) has been developed. As our subject group was composed of families with multiple siblings, we applied this new procedure, which allowed us to use all the information contained in our study group despite the varied number of siblings in these families. In this large, nonclinical group of 376 nuclear families we used the family-based association test (16) to test the association between three IGF2 SNPs and eating behavior. Eating behavior was assessed by using the 26-item Eating Attitudes Test (17, 18), which is probably the most widely used standardized measure of symptoms and concerns characteristic of eating disorders. Numerous studies have used the Eating Attitudes Test as a screening tool. Moreover, overall scores show heritability of approximately 40% in normal female twins (19).

Method

Subjects

The respondents were primarily college students at Israeli colleges and their families, recruited by word of mouth and advertisements on campus notice boards. The group analyzed in the current study included 376 families including 1,549 persons: 12% of the families had one offspring apiece, 71% had two siblings, 11% had three siblings, and 6% had four or more siblings. The siblings ages were between 14 and 34 years (average=22.30).

Each contact person received a number of questionnaires (equal to the number of participating siblings) and two sterile test tubes per family member for DNA sampling; each test tube contained 10 cm³ of Aquafresh mouthwash. Questionnaires were completed by siblings but not parents. After complete description of the study to the subjects, written informed consent was obtained. The completed questionnaires and DNA samples were returned by mail or hand-delivered to an office. The contact person received a modest monetary incentive, and the study was approved by the local institutional review board and by the Israeli Ministry of Health Genetics Committee.

Instruments

Eating Attitudes Test. The Eating Attitudes Test is a 26-item self-report factor-analytically derived scale, originally validated on 160 women with eating disorders and 140 female nonclinical comparison subjects (18). It is reliable and valid, correlates highly with the original 40-item scale (r=0.98) (18), and screens for cases of eating disorders in both clinical and nonclinical subjects. Each item is scored on a 6-point Likert scale with answers ranging from “never” to “always.” For clinical purposes, the three least frequent categories (“never,” “rarely,” and “sometimes”) are given a score of 0, “often” is scored as 1, “usually” is scored as 2, and “always” is given a score of 3. In order to maximize variance, however, we assigned scores for each item ranging from 0 (“never”) to 5 (“always”). Scores above 20 are generally considered indicative of risk for developing eating disorders. In the group we examined, 10.9% of the subjects scored above this threshold, 14.4% of the females and only 3.0% of the males. Similar results have been observed in college groups in the United States (20).

The three Eating Attitudes Test subscales are dieting, bulimia, and oral control. The dieting subscale contains 13 items, which tap restrictive behaviors such as intake reduction for weight loss purposes, feeling guilt after eating, and preoccupation with thinness and weight loss. The bulimia subscale contains seven items, which assess binge eating, vomiting, and food preoccupation. It distinguishes not only eating disordered from comparison subjects but also women with a restricting type of eating disorder from women with a bulimic type. The oral control subscale contains six items, which assess the degree of self-control over eating and perceived external pressure to eat.

Eating Disorder Inventory subscales. The Eating Disorder Inventory-2 is a self-report measure of symptoms generally related to eating disorders (21). It contains 11 standardized subscales, each independently derived and representing a unique trait. In the present study, the subscales for body dissatisfaction and drive for thinness were used. They have excellent internal consistency, content validity, and criterion-based and construct validity and good test-retest reliability in eating disordered and healthy comparison subjects (21).

The drive for thinness subscale (seven items) assesses preoccupation with body weight, fear of gaining weight, desire to be thin, and food intake restriction. The body dissatisfaction subscale (nine items) measures overall satisfaction with the shape and size of various parts of the body. Respondents are asked to state how often, on a scale of 0 (never) to 5 (always), they think, for example, that their hips or thighs are too large.

TABLE 1. Primers for the First and Second Polymerase Chain Reactions in Assays of Single Nucleotide Polymorphisms (SNPs) of the IGF2 Gene in 1,549 Nonclinical Subjects (805 Siblings) From 376 Families

<table>
<thead>
<tr>
<th>SNP</th>
<th>Polymorphism</th>
<th>Gene Bank</th>
<th>First Polymerase Chain Reaction Primer</th>
<th>Taqman Probea</th>
</tr>
</thead>
<tbody>
<tr>
<td>2207</td>
<td>C/T</td>
<td>X07868 nucleotide 2207 (NT28310; 912,702 bp)</td>
<td>Forward: ACAGAGATGCCCTGTTGAGG; reverse: AGGGATGTACCTGCTGTTG</td>
<td>TTTTTTTTTTTTTTTTTTTTTTTTCTACACCAGGACTGAAAAGG</td>
</tr>
<tr>
<td>1156</td>
<td>T/C</td>
<td>Y13633 nucleotide 1156 (NT28310; 929,533 bp)</td>
<td>Forward: CACACTTCTCCTCCTCTA; reverse: CCTGCGGAAAACAAAGG</td>
<td>TTTTTTTTTTTTTTTTTTTTTTTTTCGAGGAGGAGGAGG</td>
</tr>
</tbody>
</table>

a Kit from Applied Biosystems (Foster City, Calif.).
DNA Extraction and Genotyping

DNA was obtained from all family members and extracted by a Master Pure kit (Epicentre Biotechnologies, Madison, Wis.). SNPs were assayed by using an ABI SNaPshot kit (Applied Biosystems, Foster City, Calif.), and the products were analyzed in an ABI 310 DNA analyzer (Applied Biosystems). The primers for the first and second polymerase chain reactions are shown in Table 1. A ReddyMix master mix was used (Abgene, Surrey, U.K.) at a magnesium concentration of 1.5–2.5 mM MgCl2. The first polymerase chain reaction was carried out as follows. The sample was initially heated at 95°C for 5 minutes followed by 35 cycles of 95°C (30 seconds), 55°C (30 seconds), 72°C (90 seconds), and a final extension step of 72°C for 10 minutes. These SNPs have been described by Gaunt et al. (9).

The quality of genotyping was determined as follows. First, all families were initially screened for Mendelian consistency by using seven highly polymorphic microsatellite markers. Problematic families that were not consistent with Mendelian inheritance (<1%) were excluded from the study. Second, since all subjects in the current study were family members, genotype errors with the three IGF2 SNPs appearing as Mendelian inconsistency, which were automatically flagged by the family-based association test, were reexamined either by searching for data entry errors or by regenotyping such families for the IGF2 SNPs. Third, in all cases when the output resulted in borderline classifications, the polymerase chain reaction procedure was repeated. Fourth, quality control and estimation of the error rate (percentage of miscalled genotypes) were evaluated by reanalysis of 5% of the families. The observed error rate was estimated to be less than 0.5%. Fifth, since deviation from Hardy-Weinberg equilibrium in random samples may be indicative of problematic assays (22), the frequencies of the three IGF2 SNPs were examined for the Hardy-Weinberg equilibrium. No significant deviation from the Hardy-Weinberg equilibrium was observed (IGF2 1156: χ²=1.87, df=1, n.s.; 2207: χ²=0.71, df=1, n.s.; ApaI: χ²=0.59, df=1, n.s.). Sixth, the allele and genotype frequencies reported in our study were similar to those reported in a British population, e.g., the frequency of the ApaI AA genotype was 31% in our group and 28% in the British study (9).

Statistical Analysis

We tested for the presence of associations between the three SNPs and scores on the Eating Attitudes Test and Eating Disorder Inventory-2 by using the family-based association test (http://www.biostat.harvard.edu/~fbat/fbat.htm), which allows for inclusion of both triads and extended families in the analysis and is also correlated with Eating Attitudes Test scores, but weakly (r=0.13, p<0.001). As shown in Table 2, analysis with the family-based association test showed a highly significant association between the IGF2 ApaI G allele and Eating Attitudes Test scores overall as well as for each of its subscales: bulimia, dieting, and oral control. Significant association was also observed between the IGF2 ApaI G allele and the scores on the body dissatisfaction and drive for thinness subscales of the Eating Disorder Inventory-2. Subjects with the G allele had higher scores. Haplotype analysis as provided for in the program for the family-based association test showed that the association with the CAT haplotype (2207-ApaI-1156) was significant for scores on the total Eating Attitudes Test (z=-3.04, p=0.003) and all three subscales: bulimia (z=3.05, p=0.002), dieting (z=3.05, p=0.002), and oral control (z=-2.91, p=0.05). This haplotype was associated with lower Eating Attitudes Test scores. Note, however, that the genetic information is primarily generated by the ApaI SNP, whereas little additional information is provided by the other two SNPs genotyped in this study, 1156 and 2207.

The data were also analyzed with a qualitative approach by setting a cutoff point on the Eating Attitudes Test of 20. Individuals scoring higher than 20 on this scale are generally considered at risk for eating disorders (20), and we categorized subjects with scores above 20 as "affected" in the family-based association test. This also indicated preferential transmission of the IGF2 ApaI G allele to the group with high scores on the Eating Attitudes Test (z=3.94, p=0.00008). Hence, the quantitative and qualitative approaches to data analysis yielded similar results.

We first estimated the effect size of the IGF2 ApaI G allele by comparing mean scores of all female subjects with the AA and GG genotypes on the Eating Disorder Inventory-2 drive for thinness subscale (AA: mean=12.71, SD=9.99; GG: mean=16.13, SD=9.09). This difference was significant (t=2.06, df=310, p=0.04, N=312). The effect size of GG homozygosity is small (0.36) according to Cohen’s widely used interpretation (24). Similar results in regard to effect size were obtained for the other eating-related measures. Small effect sizes have generally been observed for genes contributing to behavioral quantitative trait loci representing a variety of phenotypes. For example, in a meta-analysis of the DRD4 7-repeat allele and attention deficit hyperactivity disorder, an overall odds ratio of 1.9 (corresponding to a small effect size) was observed (25).

We also tested for association between the three IGF2 SNPs and body mass index (Table 3). A highly significant association was observed between the IGF2 ApaI G allele and body mass index.

All three SNPs are in linkage disequilibrium (Table 4).

Discussion

One particular SNP, ApaI, in the IGF2 gene has been shown to have a modest influence on body mass index and an inherited predisposition to weight gain (9). The current finding that it may also contribute to the pathology of eating disorders, as evidenced by its association with scores
on the Eating Attitudes Test, is intriguing. Previous studies of this gene in humans have focused on its role in metabolism and body composition. The current findings extend our knowledge of this gene’s action by exploring the impact of the common $IGF2$ polymorphisms on psychological processes related to eating behavior. Neurotransmitter modulation of appetitive behavior is the focus of most hypotheses regarding the etiology of severe eating disorders (26–28). However, the current results to some measure challenge this view, and we hypothesize that inborn meta-

### Table 2. Association of Scores on the Eating Attitudes Test and Eating Disorder Inventory-2 With Three Single Nucleotide Polymorphisms (SNPs) of the $IGF2$ Gene in 845 Nonclinical Subjects in 376 Families

<table>
<thead>
<tr>
<th>Test and $IGF2$ SNP</th>
<th>Allele</th>
<th>Allele Frequency</th>
<th>Number of Informative Families</th>
<th>Number of Observed Number Under Null Hypothesis</th>
<th>Variance in Observed Number</th>
<th>Association Between Eating Measure and SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$E[S]$</td>
<td>$Var[S]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$z$</td>
<td>$p$</td>
<td></td>
</tr>
<tr>
<td>Eating Attitudes Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total scale 2207</td>
<td>C</td>
<td>0.910</td>
<td>14</td>
<td>332,000</td>
<td>373,750</td>
<td>1993,312</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>106</td>
<td>2268,000</td>
<td>2279,250</td>
<td>23870,062</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>195</td>
<td>4510,000</td>
<td>4479,750</td>
<td>45193,438</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>94</td>
<td>2142,570</td>
<td>2319,535</td>
<td>27866,119</td>
</tr>
<tr>
<td>Bulimia subscale</td>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>12</td>
<td>229,000</td>
<td>231,500</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>102</td>
<td>1498,000</td>
<td>1582,500</td>
<td>13668,750</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>191</td>
<td>3170,000</td>
<td>3146,000</td>
<td>24130,375</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>94</td>
<td>1371,000</td>
<td>1558,500</td>
<td>14590,625</td>
</tr>
<tr>
<td>Dieting subscale</td>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>12</td>
<td>229,000</td>
<td>231,500</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>102</td>
<td>1498,000</td>
<td>1582,500</td>
<td>13668,750</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>191</td>
<td>3170,000</td>
<td>3146,000</td>
<td>24130,375</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>94</td>
<td>1371,000</td>
<td>1558,500</td>
<td>14590,625</td>
</tr>
<tr>
<td>Oral control subscale</td>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>12</td>
<td>416,000</td>
<td>426,750</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>102</td>
<td>33255,000</td>
<td>36335,750</td>
<td>14712,938</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>191</td>
<td>5282,000</td>
<td>5448,750</td>
<td>35267,188</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>96</td>
<td>3061,000</td>
<td>2940,250</td>
<td>28343,938</td>
</tr>
<tr>
<td>Eating Disorder Inventory-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body dissatisfaction subscale</td>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>12</td>
<td>416,000</td>
<td>426,750</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>102</td>
<td>2555,000</td>
<td>2635,750</td>
<td>24712,938</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>191</td>
<td>5282,000</td>
<td>5448,750</td>
<td>35267,188</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>96</td>
<td>3061,000</td>
<td>2940,250</td>
<td>28343,938</td>
</tr>
<tr>
<td>Drive for thinness subscale</td>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>12</td>
<td>407,920</td>
<td>422,870</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>102</td>
<td>2581,340</td>
<td>2625,185</td>
<td>23922,417</td>
</tr>
<tr>
<td>Apal 1156</td>
<td>C</td>
<td>0.305</td>
<td>196</td>
<td>5084,310</td>
<td>5337,792</td>
<td>51082,621</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>96</td>
<td>3010,030</td>
<td>2905,852</td>
<td>28178,522</td>
</tr>
</tbody>
</table>

---

a The family-based association test allows the use of additive, recessive, or dominant models in its analysis. The dominant model was used for the results shown here. The analysis was biallelic. The empirical variance option was used; it computes the variance around the summation statistic, $Var[S]$, empirically without assumptions about the recombination parameter or degree of correlation between multiple siblings in a family. The family-based association test uses the generic form $S=\sum T\times X$ as a test statistic, where summation ($S$) is over all offspring in all families in the data set. Body mass index and sex were entered as covariates in the family-based association test since both of these variables affect Eating Attitudes Test scores.

b With at least one heterozygote parent.

c Null hypothesis: no association in the presence of linkage (empirical variance option).

d The women scored significantly higher than men on the Eating Attitudes Test (women: mean=38.4, SD=18.04; men: mean=24.33, SD=14.22, t=13.78, df=853, p<0.0001), and the Eating Attitudes Test scores were significantly, albeit weakly, correlated with body mass index in female subjects (r=0.17, N=590, p<0.001).
IGF2 GENE AND EATING DISORDERS

TABLE 3. Association Between Body Mass Index and Three Single Nucleotide Polymorphisms (SNPs) of the IGF2 Gene in 1,549 Nonclinical Subjects (805 Siblings) in 376 Families

<table>
<thead>
<tr>
<th>IGF2 SNP</th>
<th>Allele</th>
<th>Allele Frequency</th>
<th>Number of Informative Familiesa</th>
<th>Number of Transmitted Alleles</th>
<th>Association Between Body Mass Index and SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Observed (S)</td>
<td>Expected Under Null Hypothesis (E[S])b</td>
</tr>
<tr>
<td>2207</td>
<td>C</td>
<td>0.910</td>
<td>14</td>
<td>407.920</td>
<td>422.870</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.090</td>
<td>105</td>
<td>2581.340</td>
<td>2625.185</td>
</tr>
<tr>
<td>Apal</td>
<td>A</td>
<td>0.314</td>
<td>203</td>
<td>5336.600</td>
<td>5597.953</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.684</td>
<td>107</td>
<td>3749.100</td>
<td>3270.980</td>
</tr>
<tr>
<td>1156</td>
<td>C</td>
<td>0.305</td>
<td>196</td>
<td>5084.310</td>
<td>5314.482</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.695</td>
<td>96</td>
<td>3010.030</td>
<td>2905.852</td>
</tr>
</tbody>
</table>

a In the test of association with body mass index, sex was used as a covariate.
b With at least one heterozygote parent.
c Null hypothesis: no association in the presence of linkage (empirical variance option).

TABLE 4. Linkage Disequilibrium Between Three Single Nucleotide Polymorphisms (SNPs) of the IGF2 Gene in Nonclinical Subjects in 376 Families

<table>
<thead>
<tr>
<th>Conventional Measure of Association</th>
<th>Chi-Square Analysis</th>
<th>Log of Probability</th>
<th>Cramer’s V</th>
<th>Uncertainty Coefficient (U)</th>
<th>Conventional Measure of Disequilibrium</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP Pair</td>
<td>N</td>
<td>χ²</td>
<td>df</td>
<td>p</td>
<td>D</td>
</tr>
<tr>
<td>2207 and Apal</td>
<td>1,407</td>
<td>50.33</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>0.189</td>
</tr>
<tr>
<td>2207 and 1156</td>
<td>1,407</td>
<td>52.35</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>0.192</td>
</tr>
<tr>
<td>Apal and 1156</td>
<td>1,410</td>
<td>32.77</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>0.152</td>
</tr>
</tbody>
</table>

In response to caloric surplus, fasting plasma levels of insulin increased more among the subjects with the IGF2 Apal GG genotype than among those with AA or AG. The changes were independent of changes in total fatness. The Apal G allele is also associated with significantly higher levels of IGF2 mRNA than is the A allele, showing a role for this polymorphism in the transcription of this gene (30). As expected for a hormone that in humans increases growth, the IGF2 knockout mouse presents a lighter, smaller phenotype (31) and a transgenic IGF2 mouse overexpressing insulin-like growth factor 2 is heavier than controls (32).

In the current study, the IGF2 Apal G allele was associated with a number of phenotypes (various measures of eating behavior, body mass index). The association across phenotypes reflecting direct measurements of eating behavior is not surprising since these variables (Eating Attitudes Test, its subscales, Eating Disorder Inventory-2 body dissatisfaction and drive for thinness subscales, body mass index) are correlated. The current results therefore support the notion that the association between the IGF2 Apal G allele and various measures of eating behavior and body mass index are “driven” by the G allele’s effect on body mass and composition (6–10). It is worth noting that the metabolic profile associated with the IGF2 Apal G allele precedes the onset of eating disorders and sets the stage, by its impact on adiposity and body mass index, for the onset of the clinical syndromes associated with eating disorders in some women. Although we have emphasized the risk conferred by the IGF2 Apal G allele, it should be underscored that the rare AA homozygote subjects (approximately 9% in this population) are seemingly protected by their low body mass index against the development of eating disorders.
We are aware of at least one other study, of binge eating, that showed a significant association between an eating disorder and a "metabolism" gene, the MC4R coding region (the lep-the binding domain) of the proopiomelanocortin gene (POMC) (33). All carriers of MC4R mutations were given a diagnosis of binge eating disorder, suggesting to the authors that MC4R is a candidate gene in the control of eating behavior. The current report, based on an investigation of a nonclinical population, suggests that a second "metabolism" gene, IGF2, also contributes to the control of eating behavior in humans.

Anorexia nervosa, bulimia nervosa, and binge eating disorder present a complex phenotype associated with aberrant eating behaviors, body image distortions, impulse and mood disturbances, hormonal and mineral imbalances, and characteristic temperament and personality traits. Employing a strategy of decomposing the phenotype of severe eating disorders into quantitative trait loci, we show in the present study that the IGF2 ApaI SNP is associated with scores on the Eating Attitudes Test, an instrument sensitive to pathological eating behavior, in a nonclinical population. Our results suggest the hypothesis that some individuals predisposed to modest increases in body mass index due to common polymorphisms in the IGF2 gene, and perhaps unduly influenced by the media message that thin is better (34, 35), embark on a course of constant dieting that slowly evolves into a pattern of abnormal eating behavior and, in a small percentage of such individuals, the full-blown clinical syndromes.

Received April 20, 2004; revision received July 28, 2004; accepted Sept. 24, 2004. From the Department of Psychology and the Scheinfeld Center for Human Genetics in the Social Sciences, Hebrew University, the Department of Psychology, Division of Behavioral Sciences, Ruppin Academic Center, Emek Hefer, Israel; and the Department of Research, Sarah Herzog Memorial Hospital, Jerusalem. Address correspondence and reprint requests to Dr. Ebstein, Scheinfeld Center for Human Genetics in the Social Sciences, Department of Psychology, Hebrew University, Jerusalem 91905, Israel; ebstein@mscc.huji.ac.il (e-mail). Partially supported by the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities (Dr. Ebstein).

References
Objective: The authors sought to evaluate patterns and predictors of relapse among women with eating disorders.

Method: Interviews were conducted bi-annually to annually to assess symptoms of eating disorders, axis I disorders, treatment, and psychosocial function on a weekly basis for women diagnosed with anorexia nervosa (N=136) or bulimia nervosa (N=110) and prospectively followed for 9 years. At the last follow-up, 229 (93%) of the subjects had been retained in the study group.

Results: Relapse occurred in 36% of the women with anorexia nervosa and 35% of the women with bulimia nervosa. Women with intake diagnoses of anorexia nervosa, restricting subtype, tended to develop bulimic symptoms during relapse, whereas women with intake diagnoses of anorexia nervosa, binge-purge subtype, or bulimia nervosa tended to return to bulimic patterns during relapse. Greater body image disturbance contributed to a risk of relapse in both eating disorders, and worse psychosocial function increased the risk of relapse in bulimia nervosa.

Conclusions: These results may explain the long-term efficacy of interpersonal therapy for bulimia nervosa and suggest that focused body image work during relapse prevention may enhance long-term recovery from eating disorders.

Relapse is a significant problem for individuals with eating disorders, with rates of relapse ranging from 22% to 51% across outcome studies of anorexia nervosa and bulimia nervosa (1–5). The identification of predictors of postremission relapse could reveal key targets for the prevention of future episodes of illness (6). Studies of bulimia nervosa have identified a number of posttreatment predictors of relapse (7–10). Many of these, such as higher frequencies of vomiting (7), higher dietary restraint (8), greater dissatisfaction with body image (9), and the over-importance of weight and shape (10), could be characterized as residual symptoms of an eating disorder resulting from a failure to fully respond to treatment. In contrast, studies of anorexia nervosa have provided few posttreatment predictors of relapse (5), with most studies reporting on the predictive significance of intake variables (4, 11).

A limitation of previous work on the relapse of eating disorders is the time course used in most treatment outcome studies. These investigations typically have conducted assessments at intake, during treatment, at the end of treatment, and then at a single follow-up. Thus, for several studies, relapse was captured only for individuals who recovered by the end of treatment and who were ill at follow-up. This timeline does not capture events that may occur in the interim, and, for captured events, it provides little to no information concerning how individuals fared between recovery and relapse. Thus, many of these studies do not identify targets for relapse prevention that are distinct from the original targets of treatment.

The Present Study

The purpose of this study was to examine the rates and predictors of relapse in patients with anorexia nervosa and bulimia nervosa. This study extends results from a previous investigation of recovery and relapse in this cohort (12) by including a longer duration of follow-up (from 7.5 years to 9 years) and by focusing specifically on postremission predictors of relapse. The previous investigation examined intake variables only as predictors of both recovery and relapse in the full study group and found no significant relapse predictors (12). In contrast to treatment outcome studies, the present study was designed to follow the naturalistic course of anorexia nervosa and bulimia nervosa by using a prospective longitudinal design. This study advances the existing methods of other outcome studies of eating disorders in several ways. We followed a large number of women with anorexia nervosa or bulimia nervosa at study entry, and interviews were conducted every 6–12 months with the participants since the initiation of the study in 1987. This allowed a sensitive assessment of the course of illness and the time until clinical improvements and relapse. The duration of follow-up was long, and attrition was low. Among the variables measured prospectively after the remission of eating disorders, we examined four sets of predictors: body image disturbance, other axis I disorders, treatment, and psychosocial function.
Table 1. Demographic and Clinical Characteristics of Women With Remitted Eating Disorders

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients With Remitted Anorexia Nervosa (N=42)</th>
<th>Patients With Remitted Bulimia Nervosa (N=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First to Third Quartile</td>
<td>First to Third Quartile</td>
</tr>
<tr>
<td>Age at intake</td>
<td>Median</td>
<td>Median</td>
</tr>
<tr>
<td>(years)</td>
<td>20.7</td>
<td>24.3</td>
</tr>
<tr>
<td>Duration of intake</td>
<td>18.7–22.5</td>
<td>20.2–27.9</td>
</tr>
<tr>
<td>episode (years)</td>
<td>3.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Duration of follow-up</td>
<td>16.6–6.5</td>
<td>5.5–8.3</td>
</tr>
<tr>
<td>(years)</td>
<td>9.0</td>
<td>9.5</td>
</tr>
<tr>
<td>Postremission</td>
<td>8.0–9.5</td>
<td>8.5–10.5</td>
</tr>
<tr>
<td>follow-up (years)</td>
<td>5.6</td>
<td>6.7–9.4</td>
</tr>
<tr>
<td>Comorbid lifetime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axis I disorder</td>
<td>38</td>
<td>71</td>
</tr>
<tr>
<td>hospitalization for</td>
<td>90</td>
<td>86</td>
</tr>
<tr>
<td>any disordera</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>hospitalization for</td>
<td>57</td>
<td>33</td>
</tr>
<tr>
<td>eating disordera</td>
<td>48</td>
<td>12</td>
</tr>
</tbody>
</table>

a Significant difference between groups (p<0.05).

Method

Definitions

In line with previous work (6, 13, 14), we used the MacArthur guidelines to define remission and relapse. The Psychiatric Status Rating Scale (15) was used to denote the level of symptoms according to DSM-III-R criteria. Scores of 5 or 6 indicate a person who meets the full criteria. All participants had scores of 5 or 6 for either anorexia nervosa or bulimia nervosa at study entry. Remission was defined as having a period of 8 consecutive weeks in which no or minimal symptoms of the syndrome were present (Psychiatric Status Rating Scale score of 1 or 2), consistent with the MacArthur guidelines for depression. Relapse represented a return to full-syndrome criteria after a period of remission for either anorexia nervosa or bulimia nervosa (Psychiatric Status Rating Scale score of 5 or 6).

Participants

Most participants were women seeking treatment for an eating disorder at the Massachusetts General Hospital Eating Disorders Unit and at other Boston-area eating disorder programs between October 1987 and June 1990. An additional cohort of 21 women who met DSM-III-R criteria for anorexia nervosa was recruited in 1991. The inclusion criteria for participation in the study were 1) a DSM-III-R diagnosis of anorexia nervosa or bulimia nervosa, 2) being female, 3) a minimum age of 12 years, 4) residence within 200 miles of Boston, 5) English speaking, and 6) no evidence of organic brain syndrome or terminal illness. Of the 294 women who met participation criteria, 250 (85%) agreed to participate in the longitudinal study. Four subjects dropped out of the study after their intake interview and before the first follow-up interview. Thus, the total size of the study group was 246. After complete description of the study to the subjects, written informed consent was obtained.

When we retrospectively applied DSM-IV criteria to intake data, 136 (55%) of these women met DSM-IV criteria for anorexia nervosa at intake, and 110 (45%) met DSM-IV criteria for bulimia nervosa at intake. Most participants (N=235, 96%) received some form of treatment during follow-up, and a portion (N=90, 37%) received inpatient treatment. At the last follow-up, 229 (93%) subjects had been retained in the study group. Detailed demographic data for the full group have been presented elsewhere (16).

As reported previously (12), the rates of remission were significantly higher in the patients with bulimia nervosa than in those with anorexia nervosa. Over the course of follow-up, 83 women with intake diagnoses of bulimia nervosa (75%) achieved remission, and 42 women with intake diagnoses of anorexia nervosa (31%) achieved remission. Table 1 provides data regarding age, duration of follow-up, and clinical characteristics for the remitted groups. The patients who recovered from anorexia nervosa had higher weights and shorter durations of illness at intake than those who did not recover (12). No differences were found between the patients who recovered from bulimia nervosa and those who did not (12).

Procedure

The individuals who appeared to meet study inclusion criteria were scheduled for an interview that was conducted in person by a trained research assistant. Interviewers confirmed the presence of full-syndrome eating disorders and assessed current and lifetime psychiatric disorders. Objective measures of height and weight were obtained during this intake interview. Follow-up interviews were conducted every 6 months during the first 5 years of the study and then were conducted annually during the remaining years of the study and thus required retrospective recall of the previous 6 months to 1 year. Follow-up interviews were conducted in person whenever possible. Previous research indicates that interview mode (telephone versus in-person) has little effect on reports of eating disorder symptoms in follow-up studies of eating disorders (17, 18). Mean, median, and modal duration of follow-up for the full study group were 8.6, 9.0, and 9.5 years, respectively. The subjects were paid for the initial and each follow-up interview.

Measures

At baseline, each participant’s lifetime axis I psychiatric history was assessed by using the Schedule for Affective Disorders and Schizophrenia—Lifetime Version (19) that was modified to include DSM-III-R diagnostic criteria for anorexia and bulimia nervosa. The 1983 Metropolitan Insurance Company’s height and weight norms were used to calculate percent expected body weight (20).

During follow-up assessments, the Longitudinal Interval Follow-Up Evaluation (21), adapted for eating disorders by inclusion of a section to probe for symptoms of eating disorders, was used to assess eating pathology, comorbid axis I disorders, psychosocial function, and treatment. In addition to the behavioral eating disorder symptoms indicated in DSM-III-R, the following DSM-III-R cognitive symptoms were probed: fear of becoming fat, misperception of body weight or shape, and overconcern with weight or shape. Undue influence of weight or shape on self-evaluation was not measured because this was introduced in DSM-IV in 1994 after the initiation of the longitudinal study. Research has indicated that both overconcern with weight and shape and undue influence of weight and shape were equally effective in differentiating patients with bulimia nervosa from comparison subjects with no eating disorders (22). Relevant to the present study, the Longitudinal Interval Follow-Up Evaluation adapted for eating disorders probed for information to determine the presence of comorbid mood disorders, alcohol use disorders, and drug use disorders according to the Research Diagnostic Criteria (RDC), coded in weekly intervals.

Use of the following forms of treatment was also assessed on a weekly basis: individual outpatient psychotherapy, clinician-led group psychotherapy (either inpatient or outpatient), treatment with fluoxetine, and inpatient treatment. Fluoxetine was analyzed specifically because it has been associated with improved out-
Table 2. Predictors of Relapse in Women With Anorexia Nervosa

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coding Range</th>
<th>Coefficient</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>Likelihood Ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate model predictor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misperception of body</td>
<td>0, 1</td>
<td>2.36</td>
<td>10.60</td>
<td>2.36–47.75</td>
<td>14.72</td>
<td>&lt;0.0002</td>
</tr>
<tr>
<td>Fear of fat</td>
<td>0, 1</td>
<td>1.46</td>
<td>4.32</td>
<td>1.46–12.78</td>
<td>7.56</td>
<td>0.006</td>
</tr>
<tr>
<td>Psychosocial function</td>
<td>0–5</td>
<td>0.81</td>
<td>2.24</td>
<td>1.24–4.07</td>
<td>6.67</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Concern about weight or shape</td>
<td>0, 1</td>
<td>1.58</td>
<td>4.85</td>
<td>1.07–21.95</td>
<td>5.78</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td>Global Assessment of Functioning Scale score</td>
<td>0–100</td>
<td>–0.05</td>
<td>0.95</td>
<td>0.91–0.99</td>
<td>5.76</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td>0.07</td>
<td>1.08</td>
<td>1.02–1.14</td>
<td>4.63</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>Individual psychotherapy</td>
<td>0, 1</td>
<td>1.14</td>
<td>3.13</td>
<td>1.04–9.35</td>
<td>4.47</td>
<td>&lt;0.04</td>
</tr>
<tr>
<td>Fluoxetine (mg/day)</td>
<td></td>
<td>–0.03</td>
<td>0.97</td>
<td>0.90–1.06</td>
<td>0.56</td>
<td>0.45</td>
</tr>
<tr>
<td>Group psychotherapy</td>
<td>0, 1</td>
<td>0.53</td>
<td>1.70</td>
<td>0.21–13.47</td>
<td>0.22</td>
<td>0.64</td>
</tr>
<tr>
<td>Psychiatric Status Rating Scale score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol addiction/abuse</td>
<td>0–3</td>
<td>0.33</td>
<td>1.40</td>
<td>0.68–2.86</td>
<td>0.66</td>
<td>0.41</td>
</tr>
<tr>
<td>Depression</td>
<td>0–5</td>
<td>0.07</td>
<td>1.07</td>
<td>0.79–1.46</td>
<td>0.21</td>
<td>0.65</td>
</tr>
<tr>
<td>Drug addiction/abuse</td>
<td>0–3</td>
<td>–0.02</td>
<td>0.98</td>
<td>0.52–1.85</td>
<td>0.47</td>
<td>0.65</td>
</tr>
<tr>
<td>Mania</td>
<td>0–5</td>
<td>–0.16</td>
<td>0.85</td>
<td>0.31–2.32</td>
<td>0.12</td>
<td>0.73</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>0, 1</td>
<td>–14.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td><strong>Multivariate model predictor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosocial function</td>
<td></td>
<td>0.43</td>
<td>1.54</td>
<td>0.67–3.53</td>
<td>1.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Concern about weight or shape</td>
<td></td>
<td>0.96</td>
<td>2.62</td>
<td>0.54–12.80</td>
<td>1.66</td>
<td>0.24</td>
</tr>
<tr>
<td>Misperception of body</td>
<td></td>
<td>2.58</td>
<td>13.20</td>
<td>2.67–65.20</td>
<td>15.94</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td>0.06</td>
<td>1.06</td>
<td>0.99–1.14</td>
<td>2.52</td>
<td>0.11</td>
</tr>
<tr>
<td>Individual psychotherapy</td>
<td></td>
<td>1.14</td>
<td>3.14</td>
<td>1.00–9.85</td>
<td>4.11</td>
<td>0.05</td>
</tr>
</tbody>
</table>

a No women with intake diagnoses of anorexia nervosa who relapsed had drug use disorders or received hospitalization. This prevented inclusion of these variables in Cox regression analyses.

Training and Reliability of Interview Assessments

Because of the duration of follow-up within the longitudinal study, significant effort went into training interviewers and ensuring adequate interrater reliability. A five-step training program modeled after the National Institute of Mental Health’s Collaborative Psychobiology of Depression Study was implemented for training interviewers. First, interviewers learned the RDC nosological categories. Second, they practiced interviews on nonpatient volunteers and trained interviewers who engaged in role-playing as patients. Third, they observed and scored training tapes of expert interviewers and resolved any deviations from expert ratings. Fourth, they observed actual interviews. Fifth, they were observed by a senior interviewer while conducting study interviews. Ongoing supervision by a senior interviewer was available throughout the study.

To ensure high interrater reliability, a semiannual monitoring and recertification program was conducted. Audiotaped interviews of the participants with anorexia nervosa and those with bulimia nervosa were randomly selected for review and coding by all active raters working on the study at that time, excluding the original interviewer. Three measures of rater agreement were performed: kappa, percentage, and interclass correlation coefficients (ICC). The interviewers were required to achieve a kappa of at least 0.6; otherwise, retraining was instituted. For the most recent 6-month period for which we collected reliability data, across all categories and three raters, agreement ranged from 88% to 90% and ICCs (3,3 two-way mixed-effect, with Spearman-Brown correction) ranged from 0.93 to 0.94.

Statistical Methods

We used time-varying proportional hazards (Cox) regression (24) to determine the influence of course variables after remission on time to relapse. Thus, week 1 represented the first week of the 8-week period that defined remission. Predictors were examined prospectively from week 1 until 12 weeks before the onset of relapse. By time-lagging predictor variables, we ensured that the predictors preceded the event they were intended to predict rather than potentially being a consequence of relapse. We focused on variables measured prospectively after remission to take advantage of the prospective study design and because these variables have the greatest relevance as targets for relapse prevention.

Nested models were compared by using the likelihood ratio test statistic that follows a chi-square distribution. A Cox model produces a model coefficient ($\beta$), a hazard ratio (exp[$\beta$]), a confidence interval (CI) for the hazard ratio, a likelihood ratio for the coefficient, and a p value for the coefficient. Hazard ratios greater than 1 increase hazard (shorten time to relapse), and ratios less than 1 decrease hazard (lengthen time to relapse). A ratio that does not differ significantly from 1 has no statistically significant effect on time to relapse. Because of the large number of univariate analyses, a p value of 0.01 was set for statistical significance for interpretation of these analyses.

Finally, we developed a multivariate regression model to predict time to relapse using methods suggested by Hosmer and Lemeshow (25). Briefly, this approach involves inclusion of variables demonstrating a univariate association with an outcome at the p<0.20 level as well as any variables of clinical importance, regardless of their univariate association. After fitting this initial multivariate model, each covariate of the model is removed. The remaining covariates are evaluated to determine whether the new model has lost an important effect (a 20% or greater change in the coefficient of any of the remaining covariates is evidence of an important interaction or confounding effect for the removed covariate). This process continues until no variables can be removed from the model. A p value of 0.05 was set for statistical significance of variables in the final multivariate model. Statistical analyses were performed using the R statistical package (26, 27).

Results

Forty-two women with intake diagnoses of anorexia nervosa (31%) achieved remission during the course of follow-up. Among these women, 15 relapsed (representing 36% of those whose illness remitted, or 11% of the total group). Although there was no significant difference in the rates of relapse between the women with intake diagnoses...
of restricting anorexia nervosa and the women with intake diagnoses of binge-purging anorexia nervosa, the majority of the women who relapsed reported binge eating (10 of 15, 67%) or purging (11 of 15, 73%) during their relapse episode. One-third (5 of 15, 33%) of the women with intake diagnoses of anorexia nervosa developed bulimia nervosa in their relapse episode. Only two women (13%) reported no bulimic symptoms during their relapse episode. One of these women had an intake diagnosis of anorexia nervosa, restricting subtype, and the other woman had an intake diagnosis of anorexia nervosa, binge-purge subtype.

Eighty-three women with intake diagnoses of bulimia nervosa (75%) achieved remission during the course of follow-up. Among these, 29 relapsed (representing 35% of those whose illness remitted, or 26% of the total group). In contrast to symptomatic patterns in the relapse of anorexia nervosa, most patients with intake diagnoses of bulimia nervosa who relapsed returned to having symptoms of bulimia nervosa (27 of 29, 93%). Only two women (7%) developed anorexia nervosa in their relapse episode. Despite differences in remission rates and patterns of symptoms after relapse, the women with intake diagnoses of anorexia nervosa and bulimia nervosa did not demonstrate significantly different relapse rates (log rank \( \chi^2 = 0.0007, df=1, p=0.98 \)). Thus, just over one-third (35%) of both groups with remitted illnesses relapsed.

Table 2 presents predictors of relapse in women with anorexia nervosa. From univariate models of variables prospectively measured after remission, predictors of relapse included worse psychosocial function, overconcern with weight or shape, and worse scores on the Global Assessment of Functioning Scale (GAF) (DSM-IV). In the final multivariate model for anorexia nervosa, psychosocial function, overconcern with weight or shape, Global Assessment of Functioning Scale all demonstrated a significant effect. Among these, worse psychosocial function and overconcern with weight or shape demonstrated statistical significance in increasing the risk of relapse.

Table 3 presents predictors of relapse in women with bulimia nervosa. From univariate models of variables prospectively measured after remission, predictors of relapse included worse psychosocial function, overconcern with weight or shape, and worse scores on the Global Assessment of Functioning Scale (GAF) (DSM-IV). In the final multivariate model for bulimia nervosa, psychosocial function, overconcern with weight or shape, GAF Scale scores, and mania score on the Psychiatric Status Rating Scale all demonstrated a significant effect. Among these, worse psychosocial function and overconcern with weight or shape demonstrated statistical significance in increasing the risk of relapse.

**Discussion**

Overall, just over one-third of the women who recovered from either eating disorder relapsed. These rates are consistent with previous studies (1–5). One interesting distinction between relapse in women with anorexia nervosa and women with bulimia nervosa is the extent to which crossover occurred predominantly in one direction. That is, the women who recovered from anorexia nervosa frequently relapsed into a bulimic syndrome. However, the women who recovered from bulimia nervosa did not relapse into an anorexic syndrome. This may reflect factors that contribute to achieving remission, a prerequisite for relapse. A minority of women with intake diagnoses of anorexia nervosa achieved remission over the period of observation (N=42, 31%). Thus, we examined relapse predictors in women with anorexia nervosa who were in some ways less representative of their diagnostic group. Conversely, most women with bulimia nervosa did achieve remission (N=83, 75%). Potentially, features that contribute to an increased probability of remission may have made women from both...
diagnostic groups more similar, contributing to similar rates of relapse and relapse symptom patterns, even though these features are fairly common in women with bulimia nervosa and not as common in women with anorexia nervosa. Such features may relate to the difficulty of sustaining low weight, either because of metabolic pressure to maintain weight or behavioral pressure to develop binge episodes in response to dietary restriction.

Like findings from previous investigations (9, 10), predictors of relapse included cognitive features of anorexia nervosa and bulimia nervosa. However, to the extent that these symptoms are part of the definition of both disorders, their continued presence would logically be related to the return of the full disorder. Thus, the association between body image disturbance and relapse may reflect a limitation in how we defined recovery from eating disorders (28). Within the current study, remission was defined as a score on the Psychiatric Status Rating Scale of 1 or 2 for 8 consecutive weeks. Although both scores require abstinence from eating-disordered behaviors, a score of 2 on the Psychiatric Status Rating Scale allows lingering concerns about weight and shape. This was intended to reflect normative discontent that is present in women who have never suffered from an eating disorder, and it is consistent with how many eating disorder studies define remission (29). Because these concerns appear to increase the risk for relapse, focused work on body image (30, 31) may help patients achieve lasting recovery from both behavioral and cognitive symptoms of eating disorders. Unfortunately, data collected on ongoing therapy did not reveal the content of that treatment before, during, or after remission from eating disorders. After remission of an eating disorder, it is quite likely that ongoing treatment would be related to other axis I disorders. Thus, it is not possible to evaluate the possible efficacy of focused work on body image on reducing the risk of relapse in the present study group. In addition, naturalistic follow-up studies are not designed to evaluate treatment efficacy because the same factors that might predict increased treatment-seeking likely contribute to relapse (32). This may explain why relapse in anorexia nervosa was predicted by increased individual psychotherapy.

Worse psychosocial function was retained in the final multivariate model for predicting relapse in both anorexia nervosa and bulimia nervosa and was a significant predictor in the multivariate model for relapse in patients with bulimia nervosa. This result may explain the association between psychosocial stressors and relapse previously described for bulimia nervosa (33). Women with poor psychosocial function may be less equipped to cope with life stressors. Thus, stress coupled with poor coping may lead to a return of symptomatic behaviors. Troop and Treasure (34) reported significant differences in coping responses to life stressors in the year before the onset of an eating disorder among patients with anorexia nervosa, bulimia nervosa, and comparison subjects with no history of eating disorders. In addition, poor psychosocial function among recovered patients may contribute to the emergence of such stressors. Such a pattern may explain why interpersonal psychotherapy has demonstrated efficacy in the treatment of bulimia nervosa at follow-up (35). Interpersonal therapy may reduce the risk of relapse by helping women recognize and cope with psychosocial stressors that contribute to disordered eating. Alternatively, problems in psychosocial function may not directly affect the risk of relapse but may serve as a marker of a third underlying variable—such as severity of axis II pathology—that increases the risk for relapse. Although axis II disorders were assessed at baseline, they were not assessed throughout the follow-up period and could not be included in analyses of postremission predictors of relapse.

The strengths of this study include the careful assessment of the course of eating disorders in a large group of women with intake diagnoses of either anorexia nervosa or bulimia nervosa. Both the long duration of follow-up and the low rate of attrition increased the probability that events of remission and relapse were captured. The weaknesses of this study include the low rates of full remission of anorexia nervosa. Although this is representative of results from other anorexia nervosa outcome studies (36), the low rate of remission reduced the number of women who had the potential to relapse and limited our power for the prediction of relapse. Thus, analyses for anorexia nervosa could detect only large effect sizes and may have captured some false positives. Finally, our results may not generalize to individuals who never seek treatment or who differ from our group demographically.

Future studies should examine interventions targeted at helping individuals improve body image and cope with psychosocial stressors to prevent relapse. Perhaps individuals with poor psychosocial function funnel distress into body dissatisfaction as a retreat to something they feel they can control. However, attempts to alter shape and weight do not improve psychosocial problems. Thus, the combination of poor psychosocial function and increased concern about weight and shape may trigger the return of full eating disorders. Teaching patients to cope effectively with psychosocial stressors and to accept their bodies may help prevent relapse into eating disorders.
RELAPSE OF EATING DISORDERS

References

Is Impaired Set-Shifting an Endophenotype of Anorexia Nervosa?

Joanna Holliday, Ph.D.
Kate Tchanturia, Ph.D.
Sabine Landau, Ph.D.
David Collier, Ph.D.
Janet Treasure, M.D., Ph.D., F.R.C.P., F.R.C.Psych.

Objective: Set-shifting difficulties have been reported in subjects with anorexia nervosa and appear to persist after recovery; therefore, they may be endophenotypic traits. The goals of this study were to investigate whether set-shifting difficulties are familial by examining discordant sister-pairs in comparison with healthy unrelated women and to replicate, with a broader battery, the lack of influence of an acute illness state on neuropsychological performance.

Method: Forty-seven pairs of sisters discordant for anorexia nervosa and 47 healthy unrelated women who were comparable in age and IQ completed neuropsychological tasks selected to assess set-shifting ability. Analyses of variance with standard errors that are robust against correlations within family clusters were used to compare the groups. Results were adjusted for obsessive-compulsive, anxiety, and depression symptoms. Subjects with acute (N=24) and fully remitted (N=23) anorexia nervosa were compared to assess state versus trait effects.

Results: Sisters with and without anorexia nervosa took significantly longer than unrelated healthy women to shift their cognitive set (CatBat task) and demonstrated greater perceptual rigidity (Haptic Illusion task) but did not differ significantly from each other. Women with anorexia nervosa were slower than other groups on Trail Making tasks. Women who had fully recovered from anorexia nervosa made significantly fewer errors than those with acute anorexia nervosa on the Trail Making alphabet task, but these subgroups did not differ on other measures.

Conclusions: Both affected and unaffected sisters had more set-shifting difficulties than unrelated healthy women. This finding, together with the replicated finding that set-shifting difficulties persist after recovery, suggests that set-shifting difficulties are trait characteristics and may inform the search for the endophenotype in anorexia nervosa.

The classification of psychiatric disorders on the basis of overt clinical phenotypes might not be optimal in the search for vulnerability genes and other etiological factors because the genotype-phenotype relationship in complex disorders is indirect. For this reason, there has been renewed interest in the search for “endophenotypes,” measurable disease-associated traits that have a simpler relationship with underlying genes than clinical measures (1). One possible endophenotype of complex psychiatric disorders is neuropsychological function.

Executive functioning includes the processes that supervise the operation of other cognitive processes, such as inhibiting actions, restraining and delaying responses, attending selectively, setting goals, planning, and organizing, as well as set-maintaining and set-shifting, located primarily in the prefrontal cortex. Studies by our group (2–4) examined executive functioning in eating disorders, focusing on set-shifting tasks. Set-shifting involves the ability to move back and forth between tasks, operations, or sets (5). Using selected perceptual and cognitive tasks, our group (3) found that individuals with anorexia nervosa took significantly longer to “shift set” than subjects with similar IQ who did not have anorexia nervosa. Some difficulty in set-shifting persisted in women who had recovered from anorexia nervosa (4), suggesting that set-shifting difficulties are not purely a function of the acute illness state. The hypothesized association between set-shifting difficulties and anorexia nervosa have face validity in that individuals with anorexia nervosa are often described as persistent, rigid, conforming, or obsessive personalities (6, 7).

An endophenotype linked to heritable etiological factors must fulfill several criteria: it must be state-independent; it must associate with illness and co-segregate within families; and it must be found at a higher rate in nonaffected family members than in the general population (1). If a characteristic fulfills these criteria but is not proven to be heritable, it is termed a “biological marker.” Our previous work indicated that the set-shifting abnormalities in anorexia nervosa were independent of the stage of illness (4). Thus, the primary aim of this study was to examine whether unaffected first-degree relatives (sisters) shared this pattern of neuropsychological impairment. We used the same test battery as in the previously reported study, and our subjects were women with ano-
reexia nervosa, their healthy sisters, and healthy comparision women of similar age and IQ. We tested whether healthy sisters demonstrated set-shifting characteristics at a higher rate than the healthy unrelated women, and we tried to replicate, with a broader battery, the finding that these abnormalities are independent of acute illness state.

**Method**

**Participants**

Forty-seven women with anorexia nervosa and one healthy sister of each were invited to participate in this study as part of a collaborative, multicenter study conducted across Europe to investigate risk factors for eating disorders. Individuals with anorexia nervosa were recruited from a clinical population of individuals receiving treatment for eating disorders within the South London and Maudsley National Health Service Trust (N=18 [38%]) and from a register of 500 individuals with past or current eating disorder (N= 29 [62%]). Individuals on the register were ascertained from clinical populations and service user groups within the United Kingdom. Forty-four of the participants with anorexia nervosa (94%) were newly recruited for this study and have not been reported on previously. Three of the individuals with anorexia nervosa completed the neuropsychological battery for a previous paper published by our group (3). Twenty-eight (60%) of the individuals with anorexia nervosa reported a lifetime history of anorexia nervosa, binge-purge subtype, and the remaining 19 individuals (40%) were diagnosed as having anorexia nervosa, restrictive subtype.

Forty-seven comparison women of normal weight with no history of eating disorders were recruited by public advertisement in the local community. The comparison group consisted of the 35 women previously reported on by our group (3) and 12 additional participants recruited for this study.

All participants were native English speakers, and 135 (96%) of the women were of white Caucasian origin. Ethical approval for the study was obtained from the South London and Maudsley Trust Research Ethics Committee. After complete description of the study to the subjects, written informed consent was obtained from all participants.

Individuals were included in the sister-pair group if they had a DSM-IV lifetime diagnosis of anorexia nervosa and a healthy sister who did not have any form of eating disorder (anorexia nervosa, bulimia nervosa, or eating disorder not otherwise specified). Women with anorexia nervosa approached about this study were asked if they had a sister who might be willing to participate. With their permission, the sister was contacted and invited to take part. Sisters had to be less than 10 years apart in age and to have lived in the same family as the patient for a minimum of 8 years. If the patient had more than one sister, the sister closest in age was approached. Each sister was interviewed independently.

Of the 52 sister-pairs originally recruited for this study, three were excluded because the sister met criteria for anorexia nervosa, bulimia nervosa, or eating disorder not otherwise specified. Individuals with neurological illness, head injury, comorbid psychotic disorder, or learning difficulties were also excluded (N=2). Of the women with a lifetime diagnosis of anorexia nervosa, 23 (49%) met the criteria for full recovery (normal weight for at least a year and regular menses), 21 (45%) had acute anorexia nervosa, and three (6%) had recently recovered normal weight (body mass index >17.5). Women at different stages of recovery were included in this study because our primary aim was to examine neuropsychological performance in individuals with anorexia nervosa, their unaffected sisters, and healthy unrelated women. In the subsidiary analysis comparing acutely ill and recovered individuals, the three women who had recently recovered normal weight were included in the acutely ill group because they were currently inpatients receiving treatment for anorexia nervosa and continued to fulfill all the DSM-IV criteria with the exception of body mass index. A previous study by our group (4) examining the effect of short-term weight gain on neuropsychological performance showed no significant improvement across any of the tasks.

None of the healthy unrelated women were taking psychotropic medication. Within the sister-pair group, one of the healthy sisters reported taking antidepressant medication (a selective serotonin reuptake inhibitor [SSRI]) at the time of the study. Seventeen (36%) of the women with anorexia nervosa were currently taking psychotropic medication: one reported taking a tricyclic antidepressant (imipramine), one reported taking mirtazapine, and the rest were taking SSRIs. The proportion of individuals taking psychotropic medication was similar among those who were acutely ill and those with remitted anorexia (nine [38%] of 24 and eight [35%] of 23, respectively).

**Measures**

Eating disorder diagnoses were made according to DSM-IV criteria on the basis of a clinical interview (a European adaptation of the Longitudinal Interval Follow-Up Evaluation [8]) and the Eating Disorder Examination (9). The Eating Disorder Examination has demonstrated good interrater reliability in terms of diagnoses (kappa=0.82–1.0) and illness history (kappa=0.80–0.99) variables. Diagnostic validity (compared with clinical notes) yielded kappas between 0.77 and 1.0 for sequential diagnoses (Anderluh et al., personal communication).

The neuropsychological battery consisted of several paradigms assessing both perceptual and cognitive set-shifting ability: the Haptic Illusion task (10, 11), Brixton Test (12), Trail Making task (13), CatBat task (4), and neurological test for dyssyndiachokinesis (14). A full description of the tasks is provided in a previous publication (3) and is also available at http://www.eatingresearch.com.

Participants also completed the National Adult Reading Test, 2nd ed. (15), to provide an estimate of premorbid intellectual ability; the verbal fluency task (13) as a general measure of executive function; and the Hospital Anxiety and Depression Scale (16) and Maudsley Obsessive Compulsive Inventory (17) to assess current anxiety, depression, and obsessive-compulsive symptoms.

**Data Analyses**

The nominal significance level was chosen to be 1% to adjust for multiple neuropsychological tests. Women with anorexia nervosa, their healthy sisters, and healthy unrelated women were compared on demographic, clinical, and neuropsychological variables by using an analysis of variance (ANOVA) model. However, analysis of data from sibling pairs needs to account for the fact that measures from siblings are not independent. Standard methods for group comparison such as ANOVA assume that observations are independent. In contrast, because members of the same family share a number of characteristics, observations from the same family pair or cluster might be positively correlated. We dealt with this by using ANOVA with standard errors that are robust against correlations within family clusters (18). When a significant group difference was detected, post hoc t tests, again based on robust standard errors, were used to carry out three groupwise comparisons at a Bonferroni-adjusted significance level of 0.33%.

Box plots were used to check that the remaining assumptions of the model were fulfilled. The distribution of the clinical and neuropsychological variables appeared normal apart from depression scores on the Hospital Anxiety and Depression Scale, which were positively skewed in the healthy unrelated group and the unaffected sister group. Transformation of this variable (by square root) produced a normal distribution; therefore, the transformed values were used in the analysis. The Mann-Whitney U
test and the Wilcoxon signed rank test were used to carry out pairwise comparisons of the low error frequencies on the neuropsychological tasks. In a subsidiary analysis, to assess the possible influence of acute illness state, independent t tests and Mann-Whitney U tests were used as appropriate to compare women who were fully recovered from anorexia nervosa and those with acute anorexia nervosa.

Since anxiety and depression scores on the Hospital Anxiety and Depression Scale and scores on the Maudsley Obsessive Compulsive Inventory differed between the groups (see Results section) and might potentially influence neuropsychological performance, Spearman’s correlation coefficient was used to determine any relationship between these variables and neuropsychological test scores. Although these features may be considered part of the clinical phenotype of anorexia nervosa, when a relationship was observed, anxiety, depression, and/or Maudsley Obsessive Compulsive Inventory scores were included as covariates in the analysis to control group comparisons for these potential confounders. Both unadjusted and adjusted results are presented. Given that antidepressant (particularly SSRIs) medication may also affect performance on the neuropsychological battery, medicated (N=17) and nonmedicated (N=30) subgroups within the anorexia nervosa group were compared to test this possibility. Two-tailed tests were used throughout. All analyses were carried out using Stata (19).

### Results

The women with anorexia nervosa, their healthy sisters, and the healthy unrelated women were well balanced in age and IQ (Table 1). The women with anorexia nervosa had a wide range of illness duration (median=6.0 years, interquartile range=3.0–9.0). The median age at onset for anorexia nervosa was 16.0 years (interquartile range=13.0–19.0). The median lowest reported body mass index was 13.5 (interquartile range=11.9–14.9), indicating that the majority of the women had experienced a severe form of anorexia nervosa. The women with anorexia nervosa obtained significantly higher depression, anxiety, and obsessive-compulsive symptom scores than their healthy sisters and the healthy unrelated women (Table 1).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Women With Anorexia Nervosa (N=47)</th>
<th>Healthy Sisters (N=47)</th>
<th>Healthy Unrelated Women (N=47)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>F (df=2, 92)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>26.3 10.2</td>
<td>27.6 9.6</td>
<td>26.5 6.1</td>
<td>1.74</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>17.9 2.7</td>
<td>22.8 4.4</td>
<td>22.1 2.3</td>
<td>39.21</td>
</tr>
<tr>
<td>National Adult Reading Test score</td>
<td>15.6 6.1</td>
<td>16.9 6.2</td>
<td>15.6 4.3</td>
<td>1.57</td>
</tr>
<tr>
<td>Full-scale IQ (WAIS-R)</td>
<td>111.3 7.6</td>
<td>109.8 7.7</td>
<td>111.4 5.5</td>
<td>1.41</td>
</tr>
<tr>
<td>Hospital Anxiety and Depression Scale scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>11.1 4.9</td>
<td>5.8 3.4</td>
<td>5.5 3.1</td>
<td>25.28</td>
</tr>
<tr>
<td>Depression Maudsley Obsessive Compulsive Inventory total score</td>
<td>7.0 3.0–10.0</td>
<td>2.0 0.0–4.0</td>
<td>1.0 0.0–3.0</td>
<td>10.97</td>
</tr>
<tr>
<td>Maudsley Obsessive Compulsive Inventory total score</td>
<td>8.0 5.0–13.0</td>
<td>4.0 2.0–5.0</td>
<td>2.0 1.0–5.0</td>
<td>19.38</td>
</tr>
</tbody>
</table>

* p values were derived using standard errors that are robust against correlations within family clusters.

Anxiety and depression scores on the Hospital Anxiety and Depression Scale and scores on the Maudsley Obsessive Compulsive Inventory did not correlate with scores on the CatBat task, the Brixton Test, the neurological test for dysdiadochokinesis, or the Trail Making shifting task (all Pearson correlation coefficients below 0.20 in absolute value); therefore, unadjusted results are presented for these variables (Table 2). The number of perceptual illusions was positively correlated with depression score on the Hospital Anxiety and Depression Scale (r=0.27, N=141, p=0.001) and total score on the Maudsley Obsessive Compulsive Inventory (r=0.27, N=141, p=0.002). Total score on the Maudsley Obsessive Compulsive Inventory was also correlated with response time on the Trail Making alphabet task (r=0.26, N=141, p=0.004). Maudsley Obsessive Compulsive Inventory scores and Hospital Anxiety and Depression Scale depression scores, therefore, were considered covariates, as appropriate, for the number of perceptual illusions and response time on the Trail Making alphabet task.

The three groups were equivalent in verbal fluency and did not differ on the nonshift component of the CatBat task. Compared with their healthy sisters, the women with anorexia nervosa took significantly longer to complete the alphabet component of the Trail Making task (estimated difference=3.8 seconds, 99% CI=1.0 to 6.6) (t=3.55, df=92, p=0.001 [unadjusted]; t=3.17, df=92, p=0.002 [adjusted]) and showed a similar but statistically nonsignificant delay on the motor component (estimated difference=2.0 seconds, 99% CI=–0.1 to 4.1) (t=2.5, df=92, p<0.02). The same directional effect was observed in a comparison of the women with anorexia nervosa and the healthy unrelated women, but these differences did not reach significance when the Bonferroni-adjusted significance level of 0.003 was applied (motor component: 2.4 seconds, 99% CI=–0.7 to 5.5, t=2.01, df=92, p<0.05; alphabet component: 4.0 seconds, 99% CI=–0.9 to 8.1, t=2.57, df=92, p<0.02 [unadjusted]; t=1.95, df=92, p<0.06 [adjusted]). Healthy sisters...
did not differ from healthy unrelated women on any of the general cognitive tasks.

On the set-shifting tasks, sisters with no eating disorders took significantly longer than healthy unrelated women on the shift component of the CatBat task, equating to a mean difference of 9.2 seconds (99% CI = 1.5 to 17.0) (t = 3.14, df = 92, p = 0.002), and had a higher ratio of Cat time to Bat time of 0.40 (99% CI = 0.06 to 0.74) (t = 3.12, df = 92, p = 0.002). A similar but nonsignificant effect was observed when women with anorexia nervosa were compared with healthy unrelated women for Cat time (estimated difference = 8.4 seconds, 99% CI = –0.9 to 17.8) (t = 2.37, df = 92, p = 0.02) and the CatBat ratio (increase of 0.36, 99% CI = 0.04 to 0.68) (t = 2.96, df = 92, p = 0.004). The women with anorexia nervosa and their unaffected sisters took significantly longer than healthy unrelated women to adjust to the shift component of the CatBat task, equating to a mean difference of 9.2 seconds (99% CI = 1.5 to 17.0) (t = 3.13, df = 92, p = 0.002). A secondary analysis excluding the three women who recently achieved normal weight from the acute group produced similar, nonsignificant findings (all p values > 0.1). Furthermore, within the anorexia nervosa group, body mass index did not correlate significantly with performance on any neuropsychological task (all Pearson correlation coefficients were below 0.20 in absolute value).

The performance of affected and unaffected sisters did not differ significantly on any set-shifting task. There were no differences in error rate between groups on the Trail Making tasks (data not shown but available on request). A comparison of individuals with anorexia nervosa who were or were not receiving current psychoactive medication revealed no significant group differences across all neuropsychology tasks (p = 0.2).

In a subsidiary analysis to address the influence of illness state on performance in the women with anorexia nervosa, fully recovered women (N = 23) were compared with those who were acutely ill and those who had recently recovered normal weight (N = 24). The recovered group made significantly fewer errors than the acutely ill on the alphabet component of the Trail Making task (p = 0.003, Mann-Whitney U test). Four (17%) of the fully recovered women made one or more errors on the alphabet, compared with 15 (63%) of those with acute anorexia nervosa. Error rates on these tasks were not correlated with anxiety or depression scores on the Hospital Anxiety and Depression Scale or Maudsley Obsessive Compulsive Inventory; therefore, there was no need for adjustment. Recovered women did not differ significantly from those with acute anorexia nervosa on any of the other neuropsychological tasks (all p > 0.1).

A secondary analysis excluding the three women who had recently achieved normal weight from the acute group produced similar, nonsignificant findings (all p values > 0.1). This study set out to examine the possibility that set-shifting difficulties, thought to be trait markers for anorexia nervosa, could be classified as endophenotypes on the basis of findings in first-degree relatives. We found that the set-shifting difficulties evident in women with anorexia nervosa were shared by their healthy sisters. The most striking results were apparent from the CatBat cogni-

![Table 2. Neuropsychological Performance on the Non-Set-Shifting and Set-Shifting Tasks of Women With Anorexia Nervosa, Their Sisters, and Healthy Unrelated Women](http://ajp.psychiatryonline.org)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Women With Anorexia Nervosa (N=47)</th>
<th>Healthy Sisters (N=47)</th>
<th>Healthy Unrelated Women (N=47)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-set-shifting tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal fluency task (total words)</td>
<td>45.3 ± 12.6</td>
<td>41.0 ± 9.9</td>
<td>43.4 ± 10.1</td>
<td>1.53</td>
</tr>
<tr>
<td>Cat time (before shift)</td>
<td>28.0 ± 12.7</td>
<td>27.6 ± 14.7</td>
<td>28.9 ± 14.8</td>
<td>0.11</td>
</tr>
<tr>
<td>Trail Making time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor task</td>
<td>20.9 ± 5.6</td>
<td>19.0 ± 2.5</td>
<td>18.5 ± 5.7</td>
<td>3.25</td>
</tr>
<tr>
<td>Alphabet task</td>
<td>25.9 ± 7.5</td>
<td>22.1 ± 5.2</td>
<td>21.9 ± 7.3</td>
<td>6.58</td>
</tr>
<tr>
<td>Set-shifting tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bat time (after shift)</td>
<td>30.7 ± 20.1</td>
<td>31.5 ± 14.7</td>
<td>22.2 ± 13.7</td>
<td>5.28</td>
</tr>
<tr>
<td>CatBat shift ratio</td>
<td>1.24 ± 0.60</td>
<td>1.28 ± 0.67</td>
<td>0.88 ± 0.57</td>
<td>6.14</td>
</tr>
<tr>
<td>Perceptual illusions (Haptic Illusion task)</td>
<td>15.2 ± 8.0</td>
<td>14.6 ± 10.1</td>
<td>8.1 ± 7.4</td>
<td>11.46</td>
</tr>
<tr>
<td>Brixton Test errors</td>
<td>10.5 ± 4.4</td>
<td>9.9 ± 3.9</td>
<td>11.7 ± 4.0</td>
<td>2.46</td>
</tr>
<tr>
<td>Dysdiadochokinesis</td>
<td>16.5 ± 4.5</td>
<td>16.4 ± 4.1</td>
<td>18.4 ± 3.6</td>
<td>3.42</td>
</tr>
<tr>
<td>Trail Making shifting task (time)</td>
<td>31.9 ± 10.7</td>
<td>31.2 ± 8.1</td>
<td>31.0 ± 15.1</td>
<td>0.13</td>
</tr>
</tbody>
</table>

a Significant pairwise difference between the patients with anorexia nervosa and their healthy sisters and between the patients with anorexia nervosa and the healthy unrelated control subjects (p < 0.004). After adjusting for Maudsley Obsessive Compulsive Inventory total scores, F = 4.81, df = 2, 92, p = 0.01.

b Significant pairwise difference between the patients with anorexia nervosa and the healthy unrelated control subjects (p < 0.003).

c Significant pairwise differences between the patients with anorexia nervosa and their healthy sisters and between the patients with anorexia nervosa and the healthy unrelated control subjects (p < 0.003). After adjusting for Hospital Anxiety and Depression Scale and Maudsley Obsessive Compulsive Inventory scores, F = 6.35, df = 2, 92, p = 0.003.

Discussion

This study set out to examine the possibility that set-shifting difficulties, thought to be trait markers for anorexia nervosa, could be classified as endophenotypes on the basis of findings in first-degree relatives. We found that the set-shifting difficulties evident in women with anorexia nervosa were shared by their healthy sisters. The most striking results were apparent from the CatBat cogni-
tive set-shifting task and the Haptic Illusion task, which assess perceptual rigidity. Both affected and unaffected sisters took significantly longer to shift set on these tasks. There was also evidence of slower alternation on the test for dysdiadochokinesis in women with anorexia nervosa and their unaffected sisters. None of these tests differed between women with acute anorexia nervosa and those who had fully recovered, suggesting that they are trait and not state effects. However, slower performance on some general cognitive tasks was specific to the anorexia nervosa group. The women with anorexia nervosa were slower than their healthy sisters and healthy unrelated subjects to complete nonshift components of the Trail Making task. The women who were fully recovered from anorexia nervosa were equally slow to complete these components, although they made significantly fewer errors on one of the tasks, suggestive of a different cognitive style.

This is the first study, to our knowledge, to investigate neuropsychological functioning in the unaffected siblings of individuals with an eating disorder. The finding that women with anorexia nervosa and their unaffected sisters exhibit similar difficulties in some set-shifting tasks has several implications. It suggests that reduced cognitive and perceptual flexibility may constitute a familial trait associated with a greater risk of developing anorexia nervosa rather than a consequence or scar of the illness. One possibility is that mental rigidity is linked to persistent abnormalities in the serotonergic system seen in anorexia nervosa, including elevated serotonin 5-HT metabolites in CSF and alteration in 5-HT$_{2A}$ receptor binding (20, 21). The serotonergic system has been strongly implicated in the regulation of impulsivity and cognitive inflexibility (22–25).

The results of this study are in agreement with the hypothesis that the pattern of neuropsychological functioning observed in previous studies may constitute a biological marker or heritable endophenotype in anorexia nervosa; unaffected sisters carry copies of illness susceptibility genes that are nonpenetrant for illness but manifest as the endophenotype. Indeed, it is now well accepted that human prefrontal function, which includes factors such as information processing speed and specific cognitive abilities, is under substantial genetic control (26).

Evidence of shared neuropsychological traits in unaffected relatives parallels the results of studies in other complex disorders, such as schizophrenia (27–29), where illness-associated neuropsychological traits are evident in unaffected first-degree relatives when compared with unrelated subjects. A perhaps surprising finding is that unaffected sisters in this study had scores very similar to those of their affected sisters on the tasks of interest. One would expect the sisters’ scores to be intermediate because, on average, 50% of genes are shared. However, other studies of cognitive functioning in discordant siblings (e.g., for schizophrenia) have also reported similar impairments in both affected and unaffected siblings relative to unrelated comparison subjects (29).

The set-shifting difficulties evident in this group of women with anorexia nervosa are largely consistent with previous reports from our group (2–4). Slowness on the dysdiadochokinesis test in anorexia nervosa, which did not reach statistical significance in our study, has also been reported by a different group (30, 31). Our previous study using the same test battery (3) found no effects of SSRI medication use on neuropsychological performance.

It is possible that the tasks included in our “set-shifting” battery are testing different aspects of information processing. The Haptic Illusion and CatBat tasks have an element of learning, i.e., they involve a period of training to establish the set (presentation of unequal-sized balls 15 times) or a narrative set where the letter c for cat is required (six times). When the set is changed, previous learning has to be extinguished and a new pattern put in place. Thus, these tasks may be measuring a failing in the process of extinction in individuals with anorexia nervosa. Strober (32) has developed a theory about the maintenance of anorexia nervosa in which he suggests that “genetically driven variations in mechanisms underlying fear conditioning are posited as the second-stage contributor to the morbid level of fear that quickly ensues, and its prolonged resistance to extinction.” Both the abnormalities detected and the evidence of trait effects are compatible with this theory, although they suggest that the failure to extinguish learned responses may not be limited to fear conditioning. This line of thought merits further investigation.

In contrast to the finding in the current study, in an earlier study (3) our group found that subjects with anorexia nervosa demonstrated difficulties on the Brixton Test and the Trail Making shift tasks. It is possible that these tasks are more sensitive to aspects of illness state than the CatBat and Haptic Illusion tasks. The women with anorexia nervosa in the current study had a higher current body mass index and a shorter illness duration than subjects with anorexia nervosa in our earlier study. Moreover, 37%, compared with 69% in the earlier study, were inpatients at the time of testing. Thus, the failure in replication could be ascribed to the differences in clinical state.

In common with our previous findings (3), the women with anorexia nervosa in the current study took significantly longer to complete the two nonshift components of the Trail Making task. Comparable response times between acute and recovered subjects suggest that longer response time on these tasks is not a function of acute state. However, the absence of this feature in the healthy sisters suggests that this is either an individual-specific trait or a residual scar effect of having had anorexia nervosa.

An intriguing finding was that recovered subjects made significantly fewer errors than all other groups on the Trail Making alphabet task. Combined with the greater response times in the anorexia nervosa group relative to comparison subjects, across several tasks, this may indi-
cate a pattern of responding in the recovered state that is consistent with a more cautious cognitive style. This finding is in keeping with the slow, accurate response style reported previously in anorexia nervosa (33). The links between neuropsychological performance and aspects of personality in anorexia nervosa such as perfectionism and obsessiveness, especially regarding concern over mistakes (34), would be an interesting avenue for future research.

The strengths of this study include the use of a discordant sibling-pair design, which is a powerful tool in investigating the extent to which illness-associated characteristics are unique or shared within families (35). Eating disorder diagnoses were based on a semistructured interview that uses a timeline approach to elicit lifetime eating disorder symptoms, circumventing some of the reliability questions raised by self-reported, questionnaire-based diagnoses. The neuropsychological battery used in this study was specifically selected to test the hypothesis that individuals with anorexia nervosa might have difficulty in set-shifting and was identical to that used in a number of previous studies, allowing comparisons to be made. We were able to consider the influence of depression, anxiety, obsessive-compulsive symptoms, and medication use in our analysis, all of which may be considered potential confounders when assessing neuropsychological performance.

This study has some limitations. Because women with anorexia nervosa were recruited at different stages of recovery and treatment for this study, we could not control for the effects of current nutritional state on neuropsychological performance. In addition, we cannot determine from these data whether the pattern of set-shifting difficulties reported in women with anorexia nervosa and their healthy sisters is specific to this disorder or is a feature of other axis I disorders. The finding that anxiety, depression, and obsessive-compulsive symptoms had minimal effects on performance, however, suggests some degree of specificity. This test battery has not been used in individuals with primary depression, anxiety, or obsessive disorders in the absence of an eating disorder; therefore, the use of this battery in these groups would represent a valuable future extension of the current research.

To conclude, this study provides further support for the possibility that information processing (set-shifting or extinction) difficulties may be part of the endophenotype in anorexia nervosa. To our knowledge, this is the first study to investigate such traits in the relatives of individuals with anorexia nervosa. The extent to which altered performance on the neuropsychological battery reflects a susceptibility to anorexia nervosa or correlates of certain personality traits is an interesting question that merits further exploration. Future studies assessing the heritability and exact form of these traits would be required to accept or refute the endophenotype theory.

References
11. Kawaguchi I: [Investigation of figural after-effects from the standpoint of set-theory.] Z Psychol Z Angew Psychol 1980; 188:377–395 (German)

Received Feb 27, 2004; revision received July 29, 2004; accepted Sept 13, 2004. From the Division of Psychological Medicine, the Social, Genetic and Developmental Psychiatry Research Centre, and the Department of Biostatistics and Computing, Institute of Psychiatry, King’s College London (KCL); and the KCL Department of Academic Psychiatry, Guy’s, King’s and St. Thomas’s Medical School, London. Address correspondence and reprint requests to Dr. Holliday, Eating Disorders Unit, Box 059, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK; joanna.holliday@hmc.ox.ac.uk (e-mail). Supported by grant 1206/87 from the Health Foundation (formerly the PPP Foundation) (Professor Treasure, Professor Collier, and Dr. Ulrike Schmidt), by grant QLK1-1999-916 from the European Commission Framework V program, and by Wellcome Trust International Fellowship number 059049 (Dr. Tchanturia).
25. Winstanley CA, Theobald DE, Dalley JW, Glennon JC, Robbins TW: 5-HT(2A) and 5-HT(2C) receptor antagonists have opposing effects on a measure of impulsivity: interactions with global 5-HT depletion. Psychopharmacology (Berl) 2004; 176:376–385
Predicting Posttraumatic Stress Symptoms From Pretraumatic Risk Factors: A 2-Year Prospective Follow-Up Study in Firefighters

Markus Heinrichs, Ph.D.
Dieter Wagner, Ph.D.
Walter Schoch, M.Sc.
Leila M. Soravia, Ph.D.
Dirk H. Hellhammer, Ph.D.
Ulrike Ehlert, Ph.D.

Objective: Most studies focusing on risk factors for posttraumatic stress disorder (PTSD) have used retrospective study designs. Only a small number of studies have prospectively examined risk factors in the immediate aftermath of trauma exposure in predicting PTSD symptoms. The purpose of this study was to identify predictive risk factors for posttraumatic stress symptoms and comorbid psychopathological symptoms present during the time before exposure to traumatic stress in a high-risk population.

Method: Forty-three professional firemen were assessed immediately after basic training (baseline) and at 6, 9, 12, and 24 months after entry into firefighter service. Subjects were screened for psychopathological symptoms, including symptoms of PTSD, depression, and anxiety. Subjects were also characterized with regard to personality traits such as self-efficacy, hostility, and alexithymia. Neuroendocrine activity was assessed by examination of awakening and diurnal salivary cortisol profiles and 24-hour urinary catecholamine excretion. Multiple linear regression analysis was used to analyze posttraumatic stress symptoms at 24-month follow-up as a function of pretraumatic characteristics.

Results: A high level of hostility and a low level of self-efficacy at baseline accounted for 42% of the variance in posttraumatic stress symptoms after 2 years. Subjects who had both risk factors at baseline showed a significant increase in measures of PTSD symptoms, depression, anxiety, general psychological morbidity, and alexithymia during the 2-year period. Biological characteristics were not predictive of the development of psychopathological symptoms.

Conclusions: These results suggest that specific personality traits may constitute markers of vulnerability to the development of psychopathological symptoms after trauma exposure. Early identification of preexisting risk factors is needed to provide effective prevention and intervention for individuals who are at risk of developing trauma-related disorders.

Posttraumatic stress disorder (PTSD) is a mental disorder that potentially follows an event in which the individual experienced, witnessed, or was confronted with either actual or threatened loss of life or serious injury invoking a response of fear, helplessness, or horror. According to DSM-IV, PTSD symptoms are subdivided into three categories: reexperiencing of the trauma, numbing of affect and avoidance of trauma-related stimuli, and symptoms of excessive arousal not present before the event. Since the introduction of PTSD to the psychiatric nomenclature, a growing amount of research has centered on diagnosis, course, and treatment of the disorder. However, only a small subset of individuals who have been exposed to a traumatic event go on to develop PTSD or other disorders (e.g., major depression, anxiety disorders) (1). The estimated lifetime prevalence of PTSD ranges from 1.3% in Germany (2) to 7.8% in the United States (3). In contrast, 89.6% of citizens in the United States are exposed to at least one traumatic event at some time in their lives (4). Thus, PTSD is a possible but not inevitable consequence of trauma exposure. Because initial evidence suggests the potential benefits of early intervention shortly after trauma (5–7), accurate identification of specific risk and protective factors is needed to provide effective treatment to those who are likely to develop long-term trauma-related psychopathology.

Aside from the most salient predictor of PTSD, which is the nature of the traumatic event itself, three other risk factors were consistently identified across studies in a meta-analysis by Brewin et al. (8): psychiatric history, family history of mental disorders, and childhood abuse. In addition, personality traits (e.g., hostility, neuroticism, self-efficacy) were also identified as predictors of PTSD symptoms (8, 9). However, it is noteworthy that the vast majority of previous studies have used retrospective designs with respect to possible risk factors, and the poten-
tial for distortion of such factors in cross-sectional research is well known from a methodological point of view. Moreover, retrospective data of trauma survivors are influenced by typical PTSD symptoms, such as avoidance and amnesic or dissociative symptoms. Thus, a putative risk factor may merely be a consequence of the disorder, not one of its causes.

Until recently, few studies prospectively examined the development of mental disorders after exposure to trauma. In most of these studies, psychological and biological data were collected in the immediate aftermath and at subsequent time points after trauma exposure (e.g., assault, motor vehicle accident) to identify mechanisms that are predictive of PTSD symptoms. It is interesting to note that the salient predictors of PTSD known from retrospective studies seem to have poor predictive value for the development of the disorder when they are assessed in a prospective design (10). For example, discriminant function analysis failed to show effects of past psychiatric history, prior trauma, or intrusive symptoms in victims of motor vehicle accidents in the immediate aftermath of a trauma (11, 12). In contrast, biological variables in the acute peritraumatic phase have been shown to more accurately predict chronic PTSD (1). For example, lower cortisol levels (11, 13) and higher resting heart rates (14, 15) shortly after motor vehicle accidents were shown in persons who developed PTSD at a follow-up time, relative to those who did not.

However, these prospective posttraumatic data do not allow the identification of predisposing factors. Is it the trauma itself or pretraumatic vulnerability that gives rise to the altered biological and psychological mechanisms immediately after trauma? And if these early posttraumatic mechanisms have been shown to have high predictive value for the development of PTSD symptoms and other mental disorders, would the identification of preexisting risk factors allow a better understanding of the development of trauma-related disorders? The analysis of pretraumatic risk factors may best be carried out with a prospective, longitudinal study design that includes data from the period before exposure to trauma, which are used to determine whether any of the factors predict subsequent PTSD symptoms. Whereas pretraumatic factors of primary victims (e.g., persons who have experienced rape or a motor vehicle accident) are very difficult to establish conclusively for methodological reasons, members of high-risk populations for trauma-related disorders who are often exposed to traumatization provide an adequate sample. Professional firefighters are regularly engaged in intense traumatic events, including exposure to gruesome injuries or death and unpredictable, dangerous situations (16–21). The estimated prevalence of PTSD is 22.2% in American firefighters and 17.3% in Canadian firefighters (22). Similarly, an estimated 18.2% of German firefighters met the diagnostic criteria for PTSD (23). In view of the fact that a community study found the highest risk of PTSD in victims of assaultive violence to be 20.9% (4), it is clear that firefighters represent a population at high risk for the development of PTSD symptoms.

Although there is still considerable need for a better understanding of the development of chronic psychopathology after a traumatic event, to date, even less attention has been given to the underlying preexisting vulnerability mechanisms. In a prospective, longitudinal study design, subjective (personality traits, psychopathological symptoms) and neuroendocrine (salivary cortisol, urinary catecholamines) characteristics were repeatedly assessed. In this study we sought to answer the question: What characteristics present at the time before exposure to traumatic stress may predict PTSD symptoms and other psychopathological symptoms in a high-risk population over the course of 2 years?

**Method**

**Participants and Procedure**

Forty-three male probationary professional firefighters (mean age=25.6 years, SD=3.5) enrolled in the study immediately after completing basic training at the fire academy. All subjects underwent a medical examination to screen out chronic diseases, mental disorders (including PTSD and past trauma history), use of medication, and drug or alcohol abuse before entering the study. Four of the original 47 subjects were excluded: three met the criteria for a mental disorder (according to assessment with the PTSD Symptom Scale [24] and the General Health Questionnaire [25]) and one met the criteria for alcohol abuse. Female firefighters were not included in the study because only one female firefighter was present during the period of recruitment. The study was approved by the institutional review board of the University of Trier. All subjects provided written informed consent before participation, and all were informed of their right to discontinue participation at any time.

After an initial screening that determined their eligibility, participants were given detailed verbal explanations of the study procedures. They then received the study pack for the first study time point. It contained full standardized written instructions for the study questionnaires and saliva and urine measures (see later description). The study pack also contained saliva sampling tubes and a container for urine collection. Participants were assessed after completing basic training (baseline) and at 6, 9, 12, and 24 months after entry into the fire departments. Immediately after basic training, probationary firefighters take up postings at operational fire stations. Particularly between 6 and 12 months after job entry, the fire departments endeavor to confront the probationary firefighters intensively with stressful on-duty events (e.g., exposure to deaths of others and to unpredictable and life-threatening situations) in order to test their eligibility for the job.

The second day of 2 consecutive days off within a defined 2-week period was used for psychological and endocrine assessments. The 2-week period was necessary in order to account for irregular changes of shift and vacation. All study materials, including completed questionnaires and saliva and urine samples, were immediately transferred to the laboratory.

**Psychological Measures**

At all five study time points, participants completed questionnaires to measure demographic items, personality characteristics, and psychopathological symptoms. The validated German versions of the following questionnaires were included: the PTSD...
PREDICTING PTSD SYMPTOMS

Symptom Scale (24), the General Health Questionnaire (25), the Zung Self-Rating Depression Scale (26), the State-Trait Anxiety Inventory (27), the SCL-90-R (28), and the Toronto Alexithymia Scale (29, 30). Self-efficacy was assessed by using the Inventory on Competence and Control Beliefs (31). All of these questionnaires are widely used and have shown satisfactory internal consistency and validity.

The PTSD Symptom Scale (24) is a self-report rating scale designed to assess the presence and severity of PTSD symptoms on a 4-point scale. The questionnaire is scored as three subscales rating reexperiencing (five items), avoidance (seven items), and arousal (five items) symptoms, according to the DSM criteria. PTSD is diagnosed if at least one reexperiencing symptom, three avoidance symptoms, and two arousal symptoms are endorsed on the scale by firefighters who were traumatized at least 1 month before the examination. According to Foa et al. (24), a symptom rates as present if the PTSD Symptom Scale item corresponding to the symptom is scored one or greater. Subsyndromal PTSD can be diagnosed if the reexperiencing symptom cluster is present plus either the avoidance or the arousal cluster (23, 32). The PTSD Symptom Scale is recommended for use as a continuous measure (24) and for the assessment of PTSD symptoms in high-risk populations (7, 33).

The General Health Questionnaire (25) is a standard screening measure based on 28 items that is used for detecting individuals with a diagnosable mental disorder and has been widely validated. The questionnaire consists of subscales for somatic symptoms, anxiety/insomnia, social dysfunction, and severe depression. In the present study, the General Health Questionnaire scoring method (0-0-1-1) was applied so that 1 point was given for each affirmative answer. The General Health Questionnaire has been recommended for screening general psychological morbidity in the high-risk population of firefighters (23).

The Zung Self-Rating Depression Scale (26) has 20 items describing depressive symptoms on a 4-point scale. Scores on the Zung Self-Rating Depression Scale indicate levels of depressive symptoms that may be of clinical significance. Several studies have established this questionnaire as a reliable and valid instrument for measuring depressive symptoms (34).

Anxiety was measured with the trait anxiety scale of the State-Trait Anxiety Inventory (27), which has 20 items that are rated on a 4-point scale. Trait anxiety denotes individual differences in anxiety proneness and refers to a general tendency to respond with anxiety to perceived threats in the environment.

The SCL-90-R (28) is a multidimensional self-report instrument designed to screen for a broad range of psychological problems and symptoms of psychopathology. The inventory contains nine primary symptom scales and three global indices with a total of 90 items, each of which is rated on a 5-point scale indicating the degree of distress associated with each symptom.

Alexithymia was assessed by using the self-report Toronto Alexithymia Scale (29, 30), which is composed of 20 items rated on a 5-point scale to measure the difficulty of recognizing and verbalizing emotions.

Self-efficacy was assessed by using the Inventory on Competence and Control Beliefs (31), which consists of 32 items rated on a 6-point scale. Self-efficacy is a predictor of the individual’s perception of competence and the capacity to act autonomously and efficiently. An example of an item of this scale is “I can pretty much determine what will happen in my life.” The reliability and validity of this questionnaire are well established (35, 36).

Endocrine Measures

Salivary cortisol. Recent studies have found the measurement of cortisol as an indicator for adrenocortical activity to be of high predictive value for the development of PTSD (for review, see references 37, 38). Numerous studies indicate that salivary cortisol is considered a reliable and valid measure of the biologically active, or unbound, fraction (39), with high correlations between salivary and plasma free cortisol (40, 41). The study packs for all five time points contained full standardized written instructions and eight prelabeled saliva sampling tubes (Salivette; Sarstedt, Rommelsdorf, Germany). Participants were instructed to collect saliva immediately upon awakening and at 30, 45, and 60 minutes thereafter. This group of samples constituted the awakening cortisol profile. Four additional samples were provided over the course of a day at 8:00 a.m., 11:00 a.m., 4:00 p.m., and 8:00 p.m. This group of samples constituted the diurnal cortisol profile. Participants were instructed to collect their saliva by chewing the cotton dental swab from the Salivette for 60 seconds. For all samples, subjects were asked to refrain from smoking, eating, or drinking anything but water for at least 30 minutes before saliva collection. Samples were kept in freezers in the subjects’ residences until delivery to the laboratory. The Salivette tubes were stored in the laboratory at –20°C until they were required for biochemical analyses. Before the samples were assayed for free cortisol, they were thawed and centrifuged at 3000 rpm for 10 minutes to obtain 0.5–1.0 ml clear saliva with low viscosity. This procedure has been shown to be valid in assessing both awakening and diurnal cortisol profiles in human research (e.g., references 36, 42). The free cortisol concentration in saliva was analyzed by using a time-resolved immunoassay with fluorescence detection, as described previously (43). The limit of detection was 0.5 nmol/liter. The inter- and intra-assay coefficients of variation were below 12% and 10%, respectively.

Urinary catecholamines. Studies of alterations of the sympathetic nervous system in PTSD have shown hyperadrenergic states with higher 24-hour urinary catecholamine excretion in PTSD patients, relative to comparison subjects (44, 45). At all five time points in the present study, written instructions and a polypropylene container for urine collection were mailed to the subjects. The subjects were instructed to urinate into the container during the 24-hour collection period and refrigerate the sample until delivery to the laboratory. After the evaluation of volume, aliquots were frozen at –80°C until assayed. Norepinephrine and epinephrine concentrations were assayed by high-performance liquid chromatography with electrochemical detection (Chromsystems, Munich, Germany). The inter- and intra-assay coefficients of variation were below 5% and 4%, respectively.

Data Analyses

Initial one-way analyses of variance (ANOVA) with repeated measures were conducted to examine the course of psychometric and biological characteristics in the entire group of subjects. The areas under the curve were calculated with the trapezoid formula, aggregating the four awakening cortisol levels and the four diurnal cortisol levels, respectively (46). To determine which variables at baseline might predict PTSD symptoms (PTSD Symptom Scale) 24 months after regular exposure to traumatic events, we first performed bivariate correlations. To avoid suppression effects and multicollinearity in the regression analysis (47, 48), only baseline variables that were significantly correlated with PTSD symptoms at follow-up and not correlated among each other were included as independent variables in the subsequent regression analysis. To cover the range of PTSD symptoms in further analyses, the PTSD Symptom Scale was used as a continuous measure (8, 24). Finally, in order to elucidate which factors most strongly predict higher levels of PTSD symptoms over the course of 2 years, we used a stepwise multiple linear regression analysis with scores on the PTSD Symptom Scale in the 2-year follow-up as the dependent variable and scores on baseline variables (identified through bivariate correlations) as the independent variables. To determine how risk factors identified by regression analysis alter the course of psychological and biological characteristics, individuals were divided into high- and low-risk
groups by median split of the baseline scores for the predictors. Differences in the course of psychological and biological characteristics between the high- and low-risk groups were then examined by using two-way ANOVAs with repeated measurement (group-by-time [repeated factor: five time points]). Where the Mauchly test of sphericity indicated heterogeneity of covariance, we verified repeated-measures results with Greenhouse-Geisser corrections. The statistical significance level was set at 5% for two-sided tests.

### Results

#### Description of Subjects

Participants were assessed after completing basic training (baseline) and at 6, 9, 12, and 24 months after entry into the fire departments. None of the subjects met the criteria for PTSD at baseline. At 24-month follow-up, seven subjects (18.3%) met the criteria for PTSD and eight subjects (18.6%) met the criteria for subsyndromal PTSD according to the PTSD Symptom Scale. Table 1 summarizes data on the subjects’ psychometric and biological characteristics over the course of the study. There was a significant increase in body weight in study participants ($F=3.08$, $df=2.7$, 74.4, $p<0.05$). All other psychometric and biological variables did not significantly change in the total group during the course of the study. There was no correlation between the number and severity of traumatic events and PTSD symptoms.

#### Predicting PTSD Symptoms From Characteristics Before Trauma Exposure

Bivariate correlations were used to assess the relationship between variables before entry into the fire departments (baseline) and PTSD symptoms 24 months later. The bivariate correlations demonstrated that the PTSD Symptom Scale score at 24-month follow-up was negatively correlated with baseline self-efficacy (assessed with the Inventory on Competence and Control Beliefs) ($r=-0.40$, $df=35$, $p=0.02$) and positively correlated with baseline hostility (assessed with the SCL-90-R) ($r=0.58$, $df=35$, $p<0.001$), baseline general psychological morbidity (assessed with the General Health Questionnaire) ($r=0.45$, $df=32$, $p=0.01$), and baseline obsessive-compulsive symptoms (assessed with the SCL-90-R) ($r=0.48$, $df=35$, $p=0.004$). These variables were included as independent variables in the regression analysis in which the score on the PTSD Symptom Scale at 24-month follow-up was the dependent variable (see Data Analyses). The results obtained in the stepwise multiple linear regression analysis revealed two significant predictors of PTSD symptoms after 2 years of service in a fire department. A high level of hostility (assessed with the SCL-90-R) and a low level of self-efficacy (assessed with the Inventory on Competence and Control Beliefs) before job entry significantly predicted higher PTSD symptom levels at 2-year follow-up. It is important to note that there was no correlation between the two variables ($r=-0.04$, $df=42$, $p=0.81$). With regard to the direction of the regression effects, self-efficacy received a negative beta weight, indicating a negative relation to PTSD symptoms. In contrast, hostility had a positive effect on PTSD symptoms. The total model explained 42% of the variance in PTSD symptoms after 2 years ($F=13.37$, $df=2$, 32, $p<0.001$). The results of the stepwise multiple linear regression analysis are depicted in Table 2.

### Course of Psychological and Biological Characteristics in Firefighters With High and Low Risk for PTSD Symptoms

To further investigate the predictive value of hostility and self-efficacy for the development of PTSD symptoms

---

### Table 1. Psychometric and Biological Characteristics of Male Firefighters at Baseline (Immediately After Basic Training) and Over the First 24 Months After Entry Into Firefighter Service

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Baseline (N=43)</th>
<th>6 Months (N=38)</th>
<th>9 Months (N=37)</th>
<th>12 Months (N=34)</th>
<th>24 Months (N=36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight (kg)</td>
<td>78.33 9.76</td>
<td>78.89 9.13</td>
<td>78.77 8.99</td>
<td>80.37 9.45</td>
<td>80.49 9.48</td>
</tr>
<tr>
<td>Psychiatric symptoms (General Health Questionnaire total score)</td>
<td>13.92 5.56</td>
<td>15.00 8.05</td>
<td>16.14 10.53</td>
<td>15.18 10.94</td>
<td>15.19 9.53</td>
</tr>
<tr>
<td>Global severity (SCL-90-R global severity index)</td>
<td>0.20 0.15</td>
<td>0.18 0.17</td>
<td>0.23 0.32</td>
<td>0.20 0.33</td>
<td>0.24 0.35</td>
</tr>
<tr>
<td>Hostility (SCL-90-R hostility subscale score)</td>
<td>0.18 0.22</td>
<td>0.12 0.25</td>
<td>0.24 0.44</td>
<td>0.27 0.61</td>
<td>0.25 0.54</td>
</tr>
<tr>
<td>Trait anxiety (State-Trait Anxiety Inventory score)</td>
<td>30.24 5.71</td>
<td>31.55 7.89</td>
<td>31.43 3.68</td>
<td>31.44 10.32</td>
<td>32.36 9.10</td>
</tr>
<tr>
<td>Depression (Zung Self-Rating Depression Scale score)</td>
<td>34.44 5.17</td>
<td>36.70 7.95</td>
<td>37.50 9.34</td>
<td>37.32 10.50</td>
<td>37.72 9.41</td>
</tr>
<tr>
<td>Alexithymia (Toronto Alexithymia Scale score)</td>
<td>41.13 6.70</td>
<td>42.27 8.30</td>
<td>42.92 6.69</td>
<td>40.49 9.63</td>
<td>42.65 11.52</td>
</tr>
<tr>
<td>Self-efficacy (Inventory on Competence and Control Beliefs score)</td>
<td>67.07 6.17</td>
<td>68.19 6.40</td>
<td>68.14 8.55</td>
<td>68.15 8.09</td>
<td>69.40 8.48</td>
</tr>
<tr>
<td>Awakening salivary cortisol profile (area under the curve, nmol/liter)</td>
<td>18.47 6.38</td>
<td>17.60 5.10</td>
<td>18.94 6.95</td>
<td>18.64 7.90</td>
<td>20.33 8.79</td>
</tr>
<tr>
<td>Diurnal salivary cortisol profile (area under the curve, nmol/liter)</td>
<td>103.41 47.21</td>
<td>101.18 43.74</td>
<td>101.10 38.66</td>
<td>119.11 50.49</td>
<td>113.24 45.37</td>
</tr>
<tr>
<td>Urinary epinephrine (µg/24 hours)</td>
<td>9.47 3.76</td>
<td>8.32 2.97</td>
<td>8.89 2.86</td>
<td>9.94 5.61</td>
<td>9.96 3.72</td>
</tr>
<tr>
<td>Urinary norepinephrine (µg/24 hours)</td>
<td>38.97 11.21</td>
<td>36.98 11.01</td>
<td>40.60 14.25</td>
<td>39.54 17.07</td>
<td>38.60 13.41</td>
</tr>
</tbody>
</table>

---

**Notes:**
- Significant increase of body weight over the course of the study ($F=3.08$, $df=2.7$, 74.4, $p<0.05$).
- The area under the curve aggregates the four salivary cortisol levels measured immediately after awakening and at 30, 45, and 60 minutes thereafter.
- The area under the curve aggregates the four salivary cortisol levels measured at 8:00 a.m., 11:00 a.m., 4:00 p.m., and 8:00 p.m.
and other comorbid symptoms, high- and low-risk groups, as determined by median split of the baseline scores of both predictors, were created. The high-risk group (N=11) had a mean hostility score of 0.33 (SD=0.19) and a mean self-efficacy score of 61.55 (SD=5.09) after completing basic training (baseline). The low-risk group (N=31) had a mean score of 0.12 (SD=0.21) for hostility and a mean score of 69.03 (SD=5.32) for self-efficacy. Subjects in the high-risk group fulfilled the criteria for both a high level of hostility and a low level of self-efficacy. All other subjects were assigned to the low-risk group. Two-way ANOVAs with repeated measurement that included the subjects were assigned to the low-risk group. Two-way ANOVAs with repeated measurement that included the

subjects as a between-subject factor were performed to assess differences in the course of psychopathological symptoms before trauma and to determine whether any measures predict subsequent PTSD symptoms (9). Professional populations at high risk for trauma-related disorders (e.g., firefighters, members of the military) are regularly engaged in traumatic events and thus provide model groups in which to explore preexisting differences that constitute risk factors for PTSD symptoms.

To our knowledge, the current study is the first to investigate prospectively the predictive power of both psychological and biological characteristics before trauma exposure in the development of subsequent posttraumatic stress symptoms. The results of the multiple regression analysis show that the combination of preexisting high levels of hostility and low levels of self-efficacy is a strong predictor of the development of PTSD symptoms in the high-risk population of firefighters. The presence of both risk factors at baseline accounted for 42% of the variance in posttraumatic stress symptoms at 2-year follow-up. Moreover, firefighters with both of these personality characteristics at baseline had a steady increase during the 2-year period in scores on measures of PTSD symptoms assessed during the 2-year period of prospective measurement.

### Discussion

Although exposure to trauma is common, PTSD and other mental disorders after trauma are relatively rare. The core methodological issue in the growing field of research on traumatic stress is to test which factors precede psychopathological symptoms that emerge after trauma in order to allow early and specific diagnosis and intervention. The best method for identifying variables that increase the risk for trauma-related disorders may be to administer a battery of measures to a large number of individuals before their exposure to trauma and to determine whether any measures predict subsequent PTSD symptoms. Professional populations at high risk for trauma-related disorders (e.g., firefighters, members of the military) are regularly engaged in traumatic events and thus provide model groups in which to explore preexisting differences that constitute risk factors for PTSD symptoms.

### Table 2. Stepwise Multiple Linear Regression Analysis of Pretraumatic Risk Factors That Are Predictive of PTSD Symptoms at 24-Month Follow-Up in Male Firefighters (N=34)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Corrected R2</th>
<th>Corrected ΔR2</th>
<th>ΔF</th>
<th>Δp</th>
<th>Standardized Beta</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostility (SCL-90-R hostility subscale score)</td>
<td>0.34</td>
<td>0.32</td>
<td>0.34</td>
<td>16.78</td>
<td>0.000</td>
<td>0.54</td>
<td>4.15</td>
<td>33</td>
</tr>
<tr>
<td>Self-efficacy (Inventory on Competence and Control Beliefs score)</td>
<td>0.46</td>
<td>0.42</td>
<td>0.12</td>
<td>6.93</td>
<td>0.013</td>
<td>–0.35</td>
<td>2.63</td>
<td>32</td>
</tr>
</tbody>
</table>

*Model: F=13.37, df=2, 32, p<0.001.

In addition, there were no significant changes in urinary catecholamines during the course of the study in the high- versus the low-risk group.

Overall, the firefighters with both risk factors—a high hostility score and a low self-efficacy score—showed a general increase in all psychopathological symptoms assessed during the 2-year period of prospective measurement.
FIGURE 1. Mean Scores on Measures of Psychopathological Symptoms and Cortisol Levels in Male Firefighters With High and Low Pretraumatic Risk for the Development of PTSD Symptoms Over 24 Months of Firefighter Servicea

a Significant interaction effect (group-by-time) on PTSD Symptom Scale score (p < 0.05), Zung Self-Rating Depression Scale score (p < 0.01), SCL-90-R score (p < 0.05), and Toronto Alexithymia Scale score (p < 0.05); significant main effect of time on PTSD Symptom Scale score (p < 0.05) and Zung Self-Rating Depression Scale score (p < 0.001); significant main effect of subject group on PTSD Symptom Scale score (p < 0.01), General Health Questionnaire score (p < 0.001), Zung Self-Rating Depression Scale score (p < 0.001), State-Trait Anxiety Inventory score (p < 0.05), SCL-90-R score (p < 0.05), and Toronto Alexithymia Scale score (p < 0.001) (two-way analysis of variance with repeated measurement).
contrast, subjects with either low levels of hostility or high levels of self-efficacy or both protective traits showed no increase in psychopathological symptoms, suggesting a significant effect of these personality factors on the development or prevention of stress-related symptoms.

The present data support and extend the clinical evidence regarding the role of personality traits in PTSD. Higher levels of hostility and anger have been associated with the development of PTSD symptoms in combat veterans; victims of violent crime, sexual assault, and accidents; and political prisoners (49–63). For example, anger accounted for more than 40% of the variance in PTSD symptoms in Vietnam veterans (60). Furthermore, lower levels of self-efficacy have previously been related to PTSD symptoms (64–66). Individuals with PTSD report lower self-efficacy levels than healthy comparison subjects, although traumatized individuals without PTSD do not differ in self-efficacy from comparison subjects (66). However, all of these studies of the predictive power of alleged risk factors began only after subjects had been exposed to trauma. The design of such studies makes it difficult to differentiate between consequences and causes.

As yet, only a small number of prospective, longitudinal studies have been conducted in this area, and they have primarily examined archival data (e.g., military records) collected pretrauma. In combat veterans, specific preexisting personality traits were shown to predict PTSD symptoms (67–69). Schnurr et al. (69) found that higher scores on the MMPI paranoia, hypochondriasis, psychopathic deviate, and masculinity-femininity scales predicted PTSD symptoms in male college graduates who later served in the Vietnam War. Bramsen et al. (67) reported that higher scores on a personality measure of negativism (akin to neuroticism) predicted subsequent PTSD symptoms among Dutch veterans who took part in the United Nations Protection Force mission in the former Yugoslavia. These initial prospective findings suggested that pretrauma personality traits are predictive of later development of PTSD symptoms after trauma exposure. Moreover, there is evidence that lower intelligence levels precede rather than follow the development of PTSD symptoms in combat veterans (70).

The identification of risk factors may provide clues regarding underlying mechanisms and could help in building new strategies to prevent the development of a disorder (71). It is surprising that little attention has been directed toward the role of protective or resilience factors against stress-related psychopathology. For example, individuals with low hostility ratings may be those who have better social coping abilities. A possible consequence of a high level of hostility, on the other hand, is social isolation or lack of social support. It should be noted within this context that recovery from PTSD is significantly influenced by the ability to preserve social support networks (8, 72–77), and in turn, social support might be an important factor for maintaining high levels of self-efficacy (73). As self-efficacy refers to an individual's feeling of confidence that they can perform a desired action (78–80), individuals with a high level of self-efficacy may be able to impose meaning on their traumatic experiences, thereby fostering recovery from them. Conversely, low self-efficacy might render life more unpredictable and uncontrollable from the perspective of the survivor of a trauma, thereby increasing the risk for long-term trauma-related psychopathology. Accordingly, the intense exposure to stressful events experienced by probationary firefighters 6–12 months after job entry led to a strong increase in all psychopathological symptoms in the firefighters who had a high level of hostility and a low level of self-efficacy, while the low-risk group showed no changes during this period (Figure 1). Future prospective studies should further explore powerful protective psychological factors (e.g., social support, self-efficacy), other personality factors, and possible underlying neurobiological mechanisms (e.g., neuropeptide oxytocin) in the development of stress-related symptoms (81–83).

In the present study, we did not find significant alterations in salivary cortisol and urinary catecholamines. Although awakening cortisol concentrations were consistently lower in the high-risk group of firefighters (those with a high level of hostility and a low level of self-efficacy) than in the low-risk group during the 2-year period, the difference did not reach statistical significance, possibly owing to the small number of subjects. Most studies of chronic PTSD have demonstrated that PTSD is associated with distinct endocrine modifications, primarily involving a highly sensitized hypothalamic-pituitary-adrenal (HPA) axis characterized by decreased basal cortisol levels and increased negative feedback regulation (e.g., references 37, 84–88). These studies raise the question of when in the course of adaptation to trauma are low basal cortisol levels first observable. For example, McFarlane et al. (11) demonstrated that subjects who had developed PTSD 6 months after a motor vehicle accident had significantly lower cortisol levels within hours after the trauma, compared to subjects who subsequently developed major depression or those who did not develop a mental disorder. Resnick et al. (89) found lower cortisol levels immediately after rape only in women with a prior history of rape or assault, although cortisol levels did not predict the subsequent development of PTSD in these women. However, the course of cortisol levels pretrauma has not yet been examined. One possible explanation for the difference in our endocrine findings, compared with those of previous studies, might be a differential extent of psychopathological symptoms. In the present study, we used a continuous measure to cover the range of PTSD symptoms (e.g., references 90–94). Future longitudinal studies should additionally use standardized diagnostic interviews at all time points to measure symptoms of
PTSD. Moreover, we did not measure PTSD symptoms related to a specific traumatic event. Despite these limitations, our results do suggest that both cortisol and catecholamine levels before trauma exposure did not predict the development of posttraumatic stress symptoms over the course of 2 years. Obviously, additional prospective, longitudinal studies that use neuroendocrinological and neuroimaging techniques and that include large study groups and start before exposure to trauma are needed to further evaluate the course of biological mechanisms in the adaptation to trauma.

An important question to be raised is to what extent the current data may be generalized to other populations. The largest group of traumatized individuals in which PTSD symptoms have been studied is male combat veterans. The present study was conducted in a population of male professional firefighters. It is important to note that the risk factors discovered in male high-risk populations cannot be applied directly to other groups, including the general population, assault and rape victims, victims of accidents, or victims of natural disasters. Because of the low predictive value of salient predictors of PTSD symptoms in prospective studies, such as past psychiatric history, prior trauma, and intrusive, avoidance, and hyperarousal symptoms in the immediate aftermath of the trauma (11, 12, 95), a new vulnerability model that includes pretrauma risk factors, type of trauma, and trauma responses is warranted.

A better understanding of pretraumatic risk factors would undoubtedly have important clinical implications with regard to the development of trauma-related disorders. This study supports the evidence for a strong predictive role of personality traits in the development of PTSD symptoms and comorbid psychopathological symptoms. Ideally, it would be desirable for future studies to further explore the interrelationships between possible preexisting personality factors and psychobiological mechanisms in the development of PTSD in order to integrate pre- and posttraumatic factors of vulnerability. For example, it might be speculated that a smaller hippocampus (96) somehow alters the ability for cognitive buffering against stress, resulting in both dysfunctional personality traits (e.g., low level of self-efficacy) and HPA axis alterations (e.g., low cortisol level). Alternatively, there could be multiple pathways for developing PTSD symptoms, and the course may or may not involve all of the mechanisms mentioned earlier. Accordingly, there is most likely no linear relationship between one possible risk factor and subsequent trauma-related symptoms.

Although the reported results need to be replicated to allow firm conclusions to be drawn, quantification of hostility and self-efficacy as risk or protective factors may eventually help the clinician and specific organizations to target individuals at high risk of developing trauma-related disorders (e.g., firefighters, police, military) at an early stage. One critical question that needs to be answered is whether specific psychological and biological characteristics could be used as exclusionary factors for some types of professions in order to protect individuals who wish to enter these professions. Finally, the results may indicate that coping skills training (e.g., anger/hostility management, self-efficacy training) could be helpful for primary and secondary prevention in high-risk populations.

References
1. Yehuda R, McFarlane AC, Shalev AY: Predicting the development of posttraumatic stress disorder from the acute response to a traumatic event. Biol Psychiatry 1998; 44:1305-1313
PREDICTING PTSD SYMPTOMS


2285
PREDICTING PTSD SYMPTOMS


Video Testimony of Long-Term Hospitalized Psychiatically Ill Holocaust Survivors

Rael D. Strous, M.D.
Mordechai Weiss, M.D.
Irit Felsen, Ph.D.
Boris Finkel, M.D.
Yuval Melamed, M.D.
Avraham Bleich, M.D.
Moshe Kotler, M.D.
Dori Laub, M.D.

Objective: Many Holocaust survivors who have both psychotic disorders and residual symptoms of posttraumatic stress disorder (PTSD) remain chronically hospitalized in psychiatric institutions. This study investigated the clinical benefits of a therapeutic process facilitating a detailed videotaped account of traumatic experience (testimony method) in elderly long-term hospitalized Holocaust survivors.

Method: Twenty-four schizophrenia patients (mean age=72.2 years) who were chronically hospitalized in Israeli state psychiatric hospitals underwent assessment by blind rating with a battery of psychiatric rating scales before and 4 months after extensive videotaped interview. The rating scales included the Positive and Negative Syndrome Scale; Clinical Global Impression (CGI); Mini-Mental State Examination (MMSE); Clinician-Administered PTSD Scale, Form 2; and Structured Interview for Disorders of Extreme Stress. Full pre- and postinterview data were available for 21 patients.

Results: Thirty-eight percent of the patients met the criteria for PTSD at the first interview, compared with only 19% at the second interview. The patients had significant reductions in functional impairment and in the severity and intensity of all posttraumatic symptom clusters (intrusion, avoidance, hyperarousal); the avoidance cluster showed the most reduction. Eleven subjects had an improvement of 30% or more in total posttraumatic severity score. No differences in Positive and Negative Syndrome Scale, MMSE, Structured Interview for Disorders of Extreme Stress, and CGI total scores were noted postinterview or between the two preinterview evaluation batteries in the comparison group. Female patients had a higher prevalence of PTSD symptoms. Total Clinician-Administered PTSD Scale, Form 2, scores and total Positive and Negative Syndrome Scale scores were inversely correlated both at baseline and at follow-up.

Conclusions: Study observations suggest clinical benefits of the testimony method in the alleviation of many posttraumatic symptoms, but not psychosis, in a cohort of psychiatrically ill Holocaust survivors, despite an interval of as many as 60 years since the traumatic events. The findings have implications for care and rehabilitation of patients many years after acute traumatic events.

(Am J Psychiatry 2005; 162:2287–2294)

Few major trauma events in history match the extent and scope of the Holocaust, which took place merely 60 years ago. The events of the Holocaust, in which persons faced the prospect of being killed or witnessing mass destruction of individuals and communities, clearly qualify as traumatic stressors that would predispose exposed persons to clusters of symptoms characteristic of a DSM diagnosis of posttraumatic stress disorder (PTSD). Although PTSD symptoms may be expected in the immediate aftermath of exposure and for some time afterward (1), it appears that many of these symptoms may become ingrained and show evidence of chronicity to the extent that acute disabling PTSD symptoms, despite the time elapsed, may still be evident more than 60 years after the traumatic events (2). In individuals with evidence of PTSD chronicity, it remains unclear, however, how other chronic psychiatric illnesses interact with the traumatic experience and posttraumatic symptoms.

Many Holocaust survivors do not voluntarily disclose their experiences. In addition, some commentators have suggested that the Holocaust, taught as history, has been ignored by many persons in the mental health community. For example, Vigoda wrote: “The patient didn’t talk, and the psychiatrist didn’t know how to ask….There was a lot of difficulty in dealing with the fear, the loss, the sadness, the loneliness” (3). Routine brief accounts of patients’ pasts, as rendered in the context of a conventional psychiatric assessment of hospitalized Holocaust survivors, surprisingly do not include extensive personal histories and often omit features of traumatic experiences that may be difficult to elicit without special clinical attention and skills. Such special attention, we suggest, is particularly important in re-evaluating patients whose initial hospitalization and diagnoses long predate more recent theoretical developments and clinical formulations regarding the association of PTSD with extreme trauma. The testimony method, pio-
neered by a group of Chilean therapists, is one such development (4). The testimony method may be defined as a form of brief psychotherapy used in the treatment of survivors of “state-sponsored violence” (5). The testimony method alleviates many chronic symptoms by transforming the painful trauma story into a cathartic experience and a document that could be useful to others. Videotaping of the testimony is framed by its purpose: the creation of an autobiographical document that has as its centerpiece the traumatic experience. It is a collaborative venture during which the interviewer recedes into the background and the patient is assisted in entering a new social context by means of the narration of personal experience (5).

No systematic study has yet been conducted to determine the potential effects and benefits of the testimonial process in Holocaust survivors who are chronically hospitalized patients with mental illness in Israel (18% of the chronic psychiatric hospital population as of 1993) (6). Some researchers have suggested that ventilation of feelings after severe persecution and the revisiting of previous disturbing experiences such as the Holocaust—rather than a focus on current coping mechanisms, realities, and positive emotions—may be counterproductive to an individual's mental well-being (7). However, it remains unknown whether recounting such experiences in the context of a controlled therapeutic environment may be beneficial for elderly patients, for whom the possibility of death is revisited because of old age and ill health (8). We hypothesized that many of these patients may experience relief if they were enabled to share their history of severe persecution more openly. Our intention was to investigate the role of video testimony as a potentially useful clinical intervention many years after the acute traumatic event and to ascertain positive or negative effects several months after the testimony process.

Method

Study Population

The study population consisted of chronically ill inpatients at the Beer Yaakov and Lev Hasharon Mental Health Centers (two large state referral institutions in Israel) recruited in the years 2002–2003 for study participation. The subjects were drawn from the approximately 100 residents housed in the hostel section for Holocaust survivors established in 2000 at these centers. This Holocaust survivor population has an age range of 59–97 years, and all have severe, chronic mental illness. To be included in the study, subjects had to meet the criteria for having been victims of Nazi persecution, as defined by the Conference on Jewish Material Claims Against Germany, Inc. (experience of being in hiding or sequestered in ghettos or concentration labor and death camps, etc.), had to have been at least 3 years old during the time of persecution, and had to be willing and capable of telling a story, even if only in fragments. Survivors were excluded if they exhibited features of major cognitive impairment or severe psychotic disorganization that would preclude participation in the video testimony process. The study was approved by the local Helsinki Committee Ethical Review Committee and the Yale Human Investigation Committee. The subjects and their legal guardians provided signed informed consent after the nature of the study and its potential risks and benefits were fully explained. Consent was also obtained from each subject's designated clinician. In addition to being informed of the right to terminate participation at any time during the study, subjects were informed that they had the right to prohibit the sharing of their video testimony and the right to withdraw it at any time from the Fortunoff Video Archive for Holocaust Testimonies or the locked collections for future medical training and research.

Study Design

After baseline clinical ratings, the subjects were randomly assigned to experimental and control groups. The experimental group subsequently underwent the videotaped testimony and its clinical follow-up. The control group continued to receive regular treatment. Four months later, both the experimental group and the control group again were administered the battery of clinical rating tests used at baseline in order to assess the potential significance of the video testimony. After this second evaluation, the control subjects provided video testimony. Four months after the video testimony intervention in the control group, the group again was administered the clinical assessment battery. Both groups continued to receive their regular pharmacological treatment throughout the study.

Clinical Assessments

Patients were rated at baseline and at 4-month intervals for the 8-month duration of the study by means of the Positive and Negative Syndrome Scale (9) and the Clinical Global Impression (CGI) severity and improvement scales (10). In addition, each subject was rated with the Clinician-Administered PTSD Scale, Form 2 (11), Structured Interview for Disorders of Extreme Stress (12), and Mini-Mental State Examination (MMSE) (13). Each subject was evaluated by the same research physician (R.D.S. or B.F.) for the duration of the study in order to preserve continuity and uniformity of assessments.

Patient Monitoring

The study was carried out in an inpatient setting where patients were closely monitored and assessed daily for any adverse events or clinical deterioration after the video testimony intervention. Guidelines for the study stipulated that any evidence of worsening in clinical state was grounds for immediate termination from the study.

Video Testimony

Before the video testimony, the interviewing staff (D.L., I.F., and three additional trained psychologists) underwent a training workshop that covered the organization and content of the video testimony process and outlined interviewer conduct during the process. The video testimony consisted of a preinterview that was intended to acquaint the subject with the interviewer, provide a preliminary impression of the subject's persecution history, and provide an opportunity for the researchers to respond to any concerns expressed by the subject. The video testimony itself lasted up to 3 hours and was carried out over one or two interview sessions, depending on the subject's needs, ability, and willingness to give testimony. During a follow-up interview, the interviewer solicited further thoughts and reflections, and the subject had the opportunity to discuss any symptoms or difficulties that occurred after the testimony. The hospital-appointed psychiatrist familiar with each patient was present on-site and available throughout all study procedures. Interviews were conducted in the subject's preferred language (Hebrew, Yiddish, German, or Polish). While obtaining the subject's personal and persecution history, the interviewer placed particular emphasis on eliciting details about the individual's Holocaust experience, "reentry" from persecu-
tion, any experiences of waiting for someone to return, the rebuilding of lost ties (with family members, friends, and community members), complicated grief processes, and the early treatment process after mental illness was recognized.

At the conclusion of each interview, the two interviewers discussed the findings and presented a clinical formulation of the subject to representatives of the subject’s treatment team. The treatment team could then decide on short- or long-term treatment interventions, which may have included changes in the patient’s living conditions, reconciliation with and return to living with family members (with any necessary social supports), a change in vocational activities (e.g., supervised leisure activities, sheltered workshops, involvement in art activities), consistent group involvement with other patients who are Holocaust survivors, and gradual development of a stable, continuous one-to-one therapeutic relationship between a member of the treatment team and the patient.

**Statistical Analysis**

Associations between variables were performed by using Pearson’s correlation coefficients. Changes in CGI and MMSE scores were analyzed by using paired t tests. Changes in Positive and Negative Syndrome Scale scores were analyzed with two-by-three multivariate analysis of variance (MANOVA), with session (before, after), and Positive and Negative Syndrome Scale cluster (positive, negative, general) as within-subject factors. A gender effect was added to the model to form a two-by-two-by-three MANOVA. Significant interactions were analyzed by using Tukey’s honestly significant difference post hoc comparisons. Similar two-by-three analysis of variance (ANOVA) models were applied to the Clinician-Administered PTSD Scale, Form 2, subscales. Scores on the intensity and functional impairment subclusters were analyzed by using paired t tests. Change of PTSD symptoms was analyzed with the sign test. The prediction of the proportional change in the intensity score was analyzed with a multiple regression model, with stepwise selection of predictors (alpha=0.05). Scores on subscales of the Structured Interview for Disorders of Extreme Stress were analyzed with two-by-six MANOVAs, with session and cluster as within-subject factors, followed by paired t tests. Associations with gender were tested with chi-square tests or grouped t tests, as appropriate, depending on the nature of the variables. In addition, a two-by-two-by-three MANOVA was performed with gender as a between-subject factor and session and cluster as within-subject factors.

**Results**

The initial study group included 24 patients (10 women, 14 men). Before the analysis, data for three patients with no postinterview assessments were omitted (one subject died of natural causes [cancer] and two subjects refused the 4-month posttestimony interview). Age ranged from 60 to 85 years (mean=71.9 years, SD=7.2). The subjects’ countries of origin were as follows: Poland (N=7), Romania (N=5), Hungary (N=3), and France, Greece, Yugoslavia, Czech Republic, Russia, and Morocco (N=1 each).

The two preinterview clinical ratings were tested for six control subjects (control subjects had two clinical ratings before the video interview in order to maintain the single-blind nature of the study). The Clinician-Administered PTSD Scale, Form 2, total severity score was higher at the clinical rating, compared to the second/first (first: mean=16.75, SD=15.41; second: mean=34.50, SD=10.38) (t=3.47, df=3, p=0.04). No differences were observed on any measure of any of the other rating scales. The second group of ratings of the control subjects was therefore used for comparison with the postinterview ratings.

**Positive and Negative Syndrome Scale, CGI, and MMSE Scores**

No differences were observed between pre- and postinterview Positive and Negative Syndrome Scale subscale or total scores, CGI scores, or MMSE scores (Table 1). Correlation analysis revealed that the Positive and Negative Syndrome Scale subscale scores before and after the intervention were associated, with coefficients ranging between 0.71 and 0.89 (all significant at p<0.001).

**Clinician-Administered PTSD Scale, Form 2, Scores**

PTSD diagnosis. Clinician-Administered PTSD Scale, Form 2, data were analyzed according to the DSM-IV criteria required for a diagnosis of PTSD (endorsement of at least one intrusion, three avoidance, and two hyperarousal items). At the first interview eight patients (38.1%) met the

---

**TABLE 1. Clinical Rating Scores Before and After Video Testimony Interview in Long-Term Hospitalized Psychiatically Ill Holocaust Survivors (N=21)**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Before Interview</th>
<th>After Interview</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Positive and Negative Syndrome Scale score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>14.4</td>
<td>4.0</td>
<td>14.3</td>
</tr>
<tr>
<td>Negative</td>
<td>22.9</td>
<td>7.6</td>
<td>23.9</td>
</tr>
<tr>
<td>Global</td>
<td>31.4</td>
<td>9.3</td>
<td>31.6</td>
</tr>
<tr>
<td>Total</td>
<td>68.6</td>
<td>17.3</td>
<td>69.8</td>
</tr>
<tr>
<td>Clinical Global Impression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improvement scale score</td>
<td>4.8</td>
<td>0.5</td>
<td>4.7</td>
</tr>
<tr>
<td>CGI score</td>
<td>4.7</td>
<td>0.7</td>
<td>3.8</td>
</tr>
<tr>
<td>MMSE score</td>
<td>21.2</td>
<td>5.5</td>
<td>21.4</td>
</tr>
<tr>
<td>Clinician-Administered PTSD Scale, Form 2, score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrusion</td>
<td>5.95</td>
<td>5.62</td>
<td>3.70</td>
</tr>
<tr>
<td>Avoidance</td>
<td>15.09</td>
<td>9.29</td>
<td>8.28</td>
</tr>
<tr>
<td>Arousal</td>
<td>6.05</td>
<td>4.65</td>
<td>4.43</td>
</tr>
<tr>
<td>Total</td>
<td>27.15</td>
<td>16.88</td>
<td>16.70</td>
</tr>
</tbody>
</table>

*a* Paired t test.
In the analysis of the intensity data (Table 2), a two-by-three MANOVA with session (before, after) and cluster (avoidance, intrusion, arousal) as within-subject factors yielded a significant session effect (F=18.64, df=1, 20, p<0.001) and a significant session-by-cluster interaction (F=9.17, df=2, 40, p<0.001). Tukey’s honestly significant difference post hoc comparisons indicated a significant reduction in avoidance symptoms (p<0.001) and in intrusion symptoms (p<0.01) but not in arousal symptoms.

Functional impairment. In addition to the presence of the three major diagnostic clusters, the DSM-IV criteria for PTSD include functional impairment. Although this requirement may seem redundant in a hospitalized population, analysis of the functional items of the Clinician-Administered PTSD Scale, Form 2 items 18 and 19 indicating social and occupational functioning, respectively, showed that a reduction in impairment was present from the preinterview (mean=3.38, SD=1.94) to the postinterview session (mean=2.62, SD=1.88) (t=2.61, df=20, p<0.02).

Associated features. Of the associated features assessed with the Clinician-Administered PTSD Scale, Form 2 items 23–30, only severity (frequency plus intensity) of memory impairment showed significant improvement postintervention (before: mean=3.67, SD=1.11; after: mean=2.67, SD=1.75) (t=2.84, df=17, p<0.02).

Change in total severity scores. The proportional change in total severity score (severity after interview/severity before interview) was calculated for each patient. Proportional changes equal to 1.00 indicate no improvement or worsening of PTSD symptoms, changes less than 1.00 indicate improvement (after < before), and changes greater than 1.00 indicate worsening (after > before). Data for one patient were omitted from the analysis because of major fluctuations in the proportional change scores, including a 570% increase in the arousal score and decrease of 4% in the intrusion score. The average proportional change was 0.68 (SD=0.34), with a range from 0.12 to 1.05. Six subjects showed no change, and one subject showed an increase of symptoms (proportional change=1.05). All other subjects showed improvement, and nine subjects had improvement of 30% or more. Of the three symptom clusters, the intrusion cluster had the greatest change (mean change=0.67, SD=0.42, median=1.0), followed by the avoidance cluster (mean change=0.72, SD=0.37, median=0.86), and the arousal cluster (mean change=0.78, SD=0.35, median=0.85). A regression model predicting the total severity of proportional change, with age, sex, severity of intrusion, severity of avoidance, and severity of arousal as predictors, was significant (F=8.20, df=2, 12, p=0.003) and explained 50% of the variance. Significant predictors were preinterview severity of avoidance (beta=−0.85, t=4.05, p=0.001) and preinterview severity of arousal (beta=−0.47, t=2.30, p=0.04). The results suggest that higher avoidance scores and lower arousal scores predict greater reduction of total severity scores.

| TABLE 2. Frequency and Intensity Scores on Symptom Subscales of the Clinician-Administered PTSD Scale, Form 2, Before and After Video Testimony Interview in Long-Term Hospitalized Psychiatrically Ill Holocaust Survivors (N=21) |
|-------------------------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Clinician-Administered PTSD Scale Subscale | Before Interview | After Interview | Analysis<sup>a</sup> |
| | Mean | SD | Mean | SD | t (df=20) | p |
| Intrusion | | | | | | |
| Frequency score | 2.25 | 2.05 | 1.75 | 2.27 | 1.81 | 0.09 |
| Intensity score | 3.90 | 3.87 | 1.95 | 2.35 | 3.35 | 0.003 |
| Avoidance | | | | | | |
| Frequency score | 6.23 | 4.07 | 3.43 | 3.38 | 3.66 | <0.002 |
| Intensity score | 8.86 | 5.84 | 4.86 | 3.94 | 4.38 | <0.001 |
| Arousal | | | | | | |
| Frequency score | 2.38 | 1.96 | 2.00 | 1.95 | 1.02 | 0.23 |
| Intensity score | 3.67 | 3.07 | 2.43 | 2.16 | 2.77 | <0.02 |

<sup>a</sup> Paired t test.
Structured Interview for Disorders of Extreme Stress Scores

Data from the Structured Interview for Disorders of Extreme Stress were analyzed according to the six subscales and a total score (Table 3). In general, the postinterview scores tended to be higher than the preinterview scores. The two-by-six ANOVA model did not reveal a significant main effect of session or a session-by-cluster interaction. However, a significant increase occurred after the video interview in scores the subscale that measures alteration and regulation of affect and impulses. No other changes were detected after the interview.

Video Interview Responders

Patients were defined as responders to the video interview if their proportional severity score showed a reduction of 30% or more and as nonresponders if the reduction was less than 30%. According to this categorization, 11 patients were responders and another 10 were nonresponders. Associations between response category and background and clinical data were calculated. Only three variables showed associations with response category: preinterview avoidance frequency and intensity scores (for both, responder > nonresponder) and postinterview Structured Interview for Disorders of Extreme Stress subscale of attention consciousness (responder < nonresponder). It should also be noted that of the seven subjects with initial PTSD who completed the study (the eighth patient refused the poststudy interview), five showed an improvement of 30% or more (range=30% to 66%), one patient showed milder improvement of 11%, and another patient showed no change.

Gender Effects

Before the video testimony interview, female subjects had a higher prevalence of PTSD symptoms (55.6%) than male subjects (16.7%), although the difference only approached significance ($\chi^2=3.50$, df=1, p=0.06). This differential effect remained after the interview process, with 33.3% of female subjects and no male subjects meeting the PTSD criteria ($\chi^2=4.67$, df=1, p<0.04). A two-by-two-by-three MANOVA analysis of the severity of different diagnostic clusters revealed a significant gender effect ($F=12.59$, df=1, 18, p=0.002) but no gender interaction. Female patients had higher severity scores both before and after the interview (before: mean=39.12, SD=17.10; after: mean=54.11, SD=16.60), compared to male patients (before: mean=19.17, SD=11.50; after: mean=10.50, SD=6.87). A significant main effect of gender was observed for total Positive and Negative Syndrome Scale scores both before and after the interview (before: mean=74.73, SD=17.73; after: mean=78.17, SD=18.33), compared to female patients (before: mean=61.00, SD=14.34; after: mean=58.67, SD=18.32). A two-by-two MANOVA performed on CGI severity scores revealed a significant gender effect ($F=4.50$, df=1, 16, p<0.05) and a significant gender-by-session interaction. The scores of male patients increased after the interview (before: mean=4.89, SD=0.60; after: mean=5.11, SD=0.60), and those of female patients decreased (before: mean=4.67, SD=0.50; after: mean=4.33, SD=0.50). Post hoc comparisons with Tukey’s honestly significant difference test revealed that male and female patients differed only on postinterview scores (p<0.001). No gender differences were observed for CGI improvement, MMSE, and Structured Interview for Disorders of Extreme Stress scores.

Associations Between Clinical Measures

Associations observed between total rating scores included an inverse correlation between total Clinician-Administered PTSD Scale, Form 2, scores and total Positive and Negative Syndrome Scale scores both at baseline ($r=–0.45$, N=20, p<0.05) and at follow-up ($r=–0.44$, N=21, p<0.05) and a positive correlation between CGI severity scale scores and total Positive and Negative Syndrome Scale scores at baseline ($r=0.62$, N=18, p=0.006) and at follow-up ($r=0.75$, N=18, p<0.001).

Discussion

The study results indicate significant improvement in posttraumatic symptoms in chronically hospitalized psychiatrically ill Holocaust survivors after video testimony, although no difference was noted in psychotic symptoms. A notable improvement of symptom severity was observed in all PTSD diagnostic clusters, particularly in severity of avoidance. Because the content of the traumatic event was being addressed and focused on, increases were noted in the current relevance of the event to the individual. Greater PTSD severity scores were noted in female subjects, similar to findings in other studies (14). It is interesting to note that a decrease in overall illness severity (CGI severity scale score) was found in female subjects, compared to a slight increase in male subjects, suggesting that women in particular may benefit from such video testimony.
In addition to the expected and intuitive correlation between illness severity and psychotic symptoms, a particularly noteworthy finding of the study is the observation of an inverse correlation between PTSD symptoms and overall psychotic state. Although this finding is speculative, it may be explained by an increased awareness of and increased capacity to focus on traumatic memories in patients with less severe psychosis who would then be exposed to more painful memories and the potential for posttraumatic symptoms. Moreover, a highly psychotic state with major mechanisms of denial and distortion of reality may serve as a protective shield against the real experience of the traumatic event. When psychosis is less pronounced, awareness of the “real world” and painful memories comes to light. From a psychoanalytic perspective, the relationship between Clinician-Administered PTSD Scale, Form 2, and Positive and Negative Syndrome Scale scores can be attributed to the gender effect noted previously. The finding that female patients scored higher on PTSD symptom scales might be related to a gender difference in ability to maintain “neurotic defenses” in the face of extreme traumatization such as the Holocaust. Female patients may have been able to maintain such defenses and, as a result, developed PTSD symptoms in the face of trauma, while male patients were less able to maintain such defenses and developed psychotic features.

The psychotherapeutic benefits of the testimonial process, observed in traumatized political refugees (5), may also be observed in Holocaust survivors with severe, chronic mental illness who were victims of persecution and human rights violations. Sharing the story through testimony psychotherapy, even in the context of a single extended interview process many years after the intensely traumatic event, can reduce chronic posttraumatic symptoms and thus may be likely to improve psychosocial functioning and quality of life. Evidence for such improvement may be noted in the following vignette. “Sara” described hiding in a house for 2 years in Belgium before being reported to the Gestapo by a neighbor:

We lived in constant fear and were always scared. We never left the house in 2 years...in the end they really did come and get us. They took us all standing in a cattle car to Auschwitz. Three days. Three nights. There they separated us from Mother. We never saw her again.

The hospital staff members had not been aware of Sara’s experience of hiding and were not aware of the intensity of her experiences during the war. Her psychiatric history supported a consistent picture of anxiety related to traumatic experience. She still had nightmares about her experiences at Auschwitz. She rarely ventured outside the hospital premises, and on occasions when she was convinced to participate in outings, she remained tense and hypervigilant. After Sara's video testimony, staff members more clearly recognized the connection between her traumatic childhood and her suspicious behavior, lack of trust, and self-neglect. After the interview and her sharing of her experience, her rapport with staff and involvement in activities, including outings, significantly improved.

We are by no means the first researchers to interview Holocaust survivors about their experiences. Many others have done so, including David Boder, who used primitive wire-recording devices in interviews in the late 1940s (15). However, we appear to be one of the first groups to use the testimony method with the goal of symptom monitoring at baseline and at follow-up. We suggest the process of creating video testimonies can be cathartic, as well as a source of material for therapeutic work and modification of the treatment approach. Moreover, the testimonial event may set in motion a process of self-reflection and a need to share thoughts about past experiences with fellow patients, family members, and health care providers. This process results in an overall improvement in posttraumatic symptoms, despite a lack of noticeable effect on the underlying schizophrenia illness. In addition, it can be expected that improvement in posttraumatic symptoms will have secondary effects on depressive ideation, which is frequently associated with PTSD in Holocaust survivors (16). The positive effects of video testimony were observed regardless of the subject’s age at the time of the Holocaust experience. Because individuals with severe mental illness such as schizophrenia may underreport trauma exposure, the results of this study suggest the importance of addressing such issues in individuals who may be expected to have developed PTSD symptoms, given their exposure to trauma, but who may not overtly report such symptoms (17).

After providing videotaped testimony, subjects may finally feel relieved to some extent of the burden of their story because it has been entrusted to safekeeping, and they may no longer fear that the knowledge will vanish. The benefits of testimony are likely to be generalizable to other patients with PTSD and psychosis, especially those who have experienced severe persecution. The study results provide the grounds for further specific studies of severe and long-standing PTSD, which may be associated with psychosis, and of the therapeutic effect and impact of testimony on individualized treatment. In addition, by giving their testimonies to the Fortunoff Video Archive for Holocaust Testimonies, individuals are afforded the opportunity to help create for themselves and others a living memorial to counteract forgetfulness, ignorance, and Holocaust denial.

Although the study observations indicating improvement in PTSD symptoms (and not psychosis) after testimony were robust, the findings may to some extent challenge the idea that patients who have been chronically hospitalized with a schizophrenia diagnosis may instead have chronic PTSD with psychosis (6). The scenario of PTSD with psychosis masquerading as schizophrenia has been suggested for populations with childhood sexual and...
physical abuse (18) and combat trauma (19), as well as other subpopulations exposed to severely traumatic events (reviewed in reference 20). In addition, classic psychoanalytic theory suggests that psychosis is a defensive response against intense internal traumatic experience (21). Further investigation is required to clarify whether some of the patients in this study may have chronic PTSD with associated psychosis rather than chronic schizophrenia. It is important to note that regardless of the diagnostic formulation (PTSD with psychosis or schizophrenia), the study observations indicate that treatment approaches such as the testimony intervention appear to be beneficial.

It is interesting to note that individuals with schizophrenia may be more vulnerable to the development of PTSD because of a higher risk of exposure to trauma in general, a lower threshold for coping with stressful life events, and potentially decreased inhibitory function of dopamine in the locus ceruleus as a result of chronic antipsychotic use, leading to noradrenergic activity enhancement and arousal as seen in PTSD (reviewed in reference 22). An alternative hypothesis within the context of a stress-diathesis model is that dehumanizing trauma and consequent PTSD provokes schizophrenia in vulnerable individuals who otherwise may not have become psychotic (22).

In this study, the mean MMSE score suggested signs of dementia in many subjects. Although Yehuda et al. (23) found accelerated cognitive decline and specific memory disturbances in Holocaust survivors with chronic PTSD symptoms, it remains unclear whether the relatively low MMSE scores in our study were influenced by the subjects’ chronic PTSD symptoms or were solely a reflection of chronic schizophrenia and the aging process. It has also been suggested that PTSD symptoms may have a delayed manifestation and may appear only with dementia onset (24).

Potential risks of the study, anecdotaly reported by other trauma survivors who underwent video testimony, include temporary anxiety before and after the testimonial process and/or sleep disturbances. In this study, however, no short-term or long-lasting adverse effects were noted, even though the subjects were strongly encouraged to share any difficult experiences with their clinicians. The limitations of the study include the inability to generalize the findings to other traumatized groups, given the relative uniqueness of the trauma and the specific ethnocultural characteristics of the subjects. Although the follow-up period in this study was limited to only 4 months, it is the intention of the investigators to follow this cohort of patients for a more extended period (more than 1 year) in order to clarify long-term effects of the study intervention.

In conclusion, the study results indicate the robust usefulness of a testimony interview in the alleviation of many posttraumatic symptoms in a cohort of psychiatrically ill Holocaust survivors, despite an interval of as many as 60 years since the traumatic event. The circumstances of the testimony process, including the use of videotaping and preservation of the interviews as part of the history of the Holocaust, may have had a special effect on these patients. In light of the videotaped testimony, aspects of the patients’ care may need to be addressed, including clinical reassessment in view of a history of trauma. For example, such a history may have resulted in symptoms that could be partly alleviated and that could be made a focus of management in many, but not necessarily in all, such patients (25). In addition, rehabilitative efforts should address the specific needs that emanate from the interaction of past trauma, long-term hospitalization, and the processes of aging. We suggest that video testimony could come to play a role in treatment planning and in the design of a specialized treatment program for such individuals, including steps to address countertransference responses in staff. Further research in subpopulations of patients with comorbid PTSD and chronic schizophrenia is needed to clarify these findings and test them in the context of larger double-blind studies.

Received July 1, 2004; revision received Oct. 12, 2004; accepted Dec. 3, 2004. From Beer Yaakov Mental Health Center; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; Lev Hasharon Mental Health Center, Netanya, Israel; and the Department of Psychiatry, Yale University School of Medicine, New Haven, Conn. Address correspondence and reprint requests to Dr. Strous, Beer Yaakov Mental Health Center, P.O. Box 1, Beer Yaakov 70350, Israel; raels@post.tau.ac.il (e-mail).

Supported by grants from the Institute for Social and Policy Studies, Yale University, and the Conference on Jewish Material Claims Against Germany, Inc.

References


15. Boder D: I Did Not Interview the Dead. Champaign, University of Illinois Press, 1949


25. Tucker WM: How to include the trauma history in the diagnosis and treatment of psychiatric inpatients. Psychiatr Q 2002; 73:135–144
Peritraumatic and Persistent Dissociation in the Presumed Etiology of PTSD

John Briere, Ph.D.
Catherine Scott, M.D.
Frank Weathers, Ph.D.

Objective: Dissociative responses that occur at the time of a trauma (peritraumatic dissociation) have been described as a major risk factor for subsequent posttraumatic stress disorder (PTSD). The current study evaluated peritraumatic dissociation and PTSD from a multivariate perspective, along with a less-investigated phenomenon: trauma-specific dissociation that begins during or after an event and continues until the time of assessment (persistent dissociation).

Method: In two studies, 52 local community participants and 386 participants from the general population with histories of exposure to at least one traumatic event were assessed for the presence of PTSD and were administered measures of dissociation and peritraumatic distress.

Results: In both studies, peritraumatic dissociation, persistent dissociation, peritraumatic distress, and generalized dissociative symptoms were associated with PTSD by univariate analyses. However, multivariate analyses in both studies indicated that PTSD status was no longer related to peritraumatic dissociation once other variables (especially persistent and generalized dissociation) were taken into account. In contrast, persistent dissociation was a strong predictor at univariate and multivariate levels.

Conclusions: Trauma-related persistent dissociation is a substantial predictor of PTSD, whereas peritraumatic dissociation ceases to predict PTSD at the multivariate level. These findings suggest that it is less what happens at the time of a trauma (e.g., disrupted encoding) that predicts PTSD than what occurs thereafter (i.e., persistent avoidance).

Since the late 1800s, researchers have noted that traumatized individuals often experience dissociative symptoms, such as depersonalization, derealization, amnesia, or fugue states (1, 2). Most theorists suggest that dissociation is a defensive process in which an individual develops the capacity to separate himself or herself from the psychic and physical pain associated with exposure to traumatic events (3). This dissociative capacity, in turn, is thought to be used by the individual in future painful circumstances (including during activated trauma memories) as a way to down-regulate his or her experience of acute psychological distress (4–6).

Researchers have reported a tendency for some individuals to dissociate during or soon after traumatic events, a phenomenon referred to as peritraumatic dissociation (7, 8). This response, typically involving depersonalization or derealization, has been linked to the subsequent development of posttraumatic stress disorder (PTSD) in a number of studies (7–12). A recent meta-analysis (13), for example, concluded that peritraumatic dissociation may be the largest known psychological risk factor for PTSD (14, 15).

Despite these findings, a smaller group of studies have failed to replicate the relationship between peritraumatic dissociation and PTSD (16–18) or have found that the relationship between peritraumatic dissociation and PTSD disappears after other variables are controlled (19). Such contrary results have led some researchers to question the relevance of peritraumatic dissociation in acute stress disorder (20), where it is a central symptomatic feature.

Among the possible reasons for conflicting findings in this area, in addition to sampling and measurement differences, are two that are relevant to the current investigation: 1) significant variability in the extent to which studies control for moderating or mediating variables in the presumed dissociation-PTSD relationship and 2) potentially incomplete specification of the temporal boundaries of peritraumatic dissociation. For example, the majority of studies have not examined the extent to which the peritraumatic dissociation they measure persists over time. As a result, it is difficult to determine whether—as some suggest—the time of onset of dissociation is the critical issue or whether the more important risk factor is the persistence of such dissociation into the long term. If peritraumatic dissociation is presumed to increase the likelihood
of PTSD by blocking initial processing, it seems likely that ongoing dissociation would have an even larger effect by disrupting such processing over a more extended period.

Inclusion of a measure of persisting dissociation in peritraumatic dissociation research would help resolve the chronology issue. If both peritraumatic and persistent dissociation were evaluated multivariately, the role of peritraumatic dissociation in PTSD could be considered with control for duration. If the peritraumatic aspect of dissociation leads to PTSD, one would expect a measure of this construct to remain a strong predictor of posttraumatic stress regardless of whether persistent dissociation was taken into account. If, however, peritraumatic dissociation is associated with PTSD only to the extent that it continues beyond the trauma, control for persistent dissociation would eliminate all or most of the relationship between peritraumatic dissociation and PTSD while preserving the relationship between persistent dissociation and PTSD.

A relationship between persistent dissociation and posttraumatic stress would not be surprising. Dissociation is a major component of the DSM-IV-TR diagnostic criteria for acute stress disorder, is listed as an associated feature of PTSD, and is described by some as an integral aspect of posttraumatic stress (21). Notably, however, DSM-IV-TR acute stress disorder criteria require that dissociative symptoms occur “while experiencing or after experiencing the distressing event” [p. 471], yet DSM-IV-TR also states elsewhere that “the disturbance lasts for a minimum of 2 days” [p. 472]. The seeming contradiction between these two criteria—one referring to peritraumatic dissociation and the other to persistent dissociation—no doubt adds to the confusion in this area.

Despite the potentially important role of trauma-specific, persistent dissociation (however defined) in the etiology of PTSD, only a few studies have examined sustained posttraumatic dissociation as a specific phenomenon (22–24). Furthermore, there are few published reports to date discriminating persistent dissociation from peritraumatic (but not necessarily persistent) dissociation in the prediction of posttraumatic stress (20). Of interest, these few studies suggest that persistent dissociation is more related to acute stress disorder or PTSD than is peritraumatic dissociation.

Sustained dissociation alone may not be sufficient to explain the relationship between peritraumatic dissociation and PTSD. Although persisting dissociation may be a potential risk factor for PTSD because it blocks the processing of trauma-specific memories, it is also possible that the presence of any dissociation may be a marker for an increased likelihood of PTSD, including dissociation that an individual experienced before the index trauma. This is referred to as “generalized dissociation” for the purposes of this study and may be especially relevant for individuals who have a history of exposure to multiple traumas.

Finally, an additional predictor of PTSD is the extent of negative emotionality that an individual experiences at the time of the trauma (25, 26). It has been hypothesized that peritraumatic dissociation serves as a way to avoid the painful feelings immediately evoked by exposure to trauma (9, 10). Therefore, the negative emotional impact of a given trauma may produce both an increased likelihood of PTSD and peritraumatic dissociation. In this instance, the causal link between peritraumatic dissociation and PTSD might be smaller or nonexistent, and as a result, any statistical relationship between theses two variables might attenuate or disappear once peritraumatic distress was taken into account. Unfortunately, beyond one early investigation (27), the multivariate relationship between peritraumatic distress and peritraumatic dissociation has rarely been studied, especially in relation to a heightened risk for PTSD.

Given these potential complexities, two studies were conducted to evaluate the relationship between PTSD and a number of potential predictors that were multivariately considered: exposure to trauma, peritraumatic dissociation, peritraumatic distress, trauma-specific persistent dissociation, and generalized dissociation.

Method

Study 1

A group of 52 trauma-exposed residents of a southern U.S. community were recruited through newspaper advertisements and flyers for a study on the effects of traumatic life events. After approval was obtained from the institutional review board at Auburn University, the participants were individually interviewed with the Clinician-Administered PTSD Scale (28) and administered a series of instruments. Among these were the Peritraumatic Dissociative Experiences Questionnaire (10), the Dissociative Experiences Scale (29), and the Detailed Assessment of Posttraumatic Stress (30).

The Clinician-Administered PTSD Scale interview has very good test-retest reliability, internal consistency, and predictive validity with respect to a Structured Clinical Interview for DSM-IV diagnosis of PTSD (31). The Peritraumatic Dissociative Experiences Questionnaire is a reliable and valid 10-item research measure that evaluates self-reported dissociative symptoms that occurred at the time of the traumatic event (9, 10). The Dissociative Experiences Scale is a widely used, 28-item measure of general dissociative symptoms with demonstrated reliability and strong evidence of various types of validity (32). The Detailed Assessment of Posttraumatic Stress is a 104-item, standardized clinical test that includes information on an individual’s history of exposure to trauma, immediate reactions to a traumatic event, and diagnostic status for PTSD and acute stress disorder (30, 33).

Relevant to the current research, the peritraumatic distress subscale of the Detailed Assessment of Posttraumatic Stress specifically asks about negative emotional responses at the time of the trauma. Another component, the trauma-specific dissociation subscale, evaluates persistent dissociation, i.e., responses such as depersonalization and derealization that reportedly occurred immediately after the trauma and have persisted to the point of the current assessment.

The average age of this convenience sample was 35.9 years (SD=1.8), and 45 (86.5%) were women. The racial breakdown was 80.8% (N=42) Caucasian, 13.5% (N=7) African American/black, 3.8% (N=2) Hispanic, and 1.9% (N=1) “other.”

Discriminant function analysis with the Statistical Package for the Behavioral Sciences (version 11.5) (SPSS Inc., Chicago) was
performed on the participants’ Clinician-Administered PTSD Scale PTSD status as a function of gender, their Peritraumatic Dissociative Experiences Questionnaire score, their Dissociative Experiences Scale score, and their peritraumatic distress and trauma-specific dissociation subscale scores from the Detailed Assessment of Posttraumatic Stress.

**Study 2**

Data for study 2 consisted of all 386 participants from the general population from the normative study of the Detailed Assessment of Posttraumatic Stress who had a history of exposure to at least one DSM-IV criterion A trauma. Analysis of this data set provided a more fine-grained replication and extension of study 1 employing a larger and potentially more representative sample of the general population and newer psychological measures of posttraumatic stress and dissociation. Approved by the institutional review board of the University of Central Florida, this study was a random mailed survey of individuals with cars or telephones in the United States and had a response rate of 11.4%. The average age of this traumatized subsample of the Detailed Assessment of Posttraumatic Stress normative study was 45.2 years (SD=16.7), 209 (52.3%) were men, and the racial breakdown was 83.2% (N=321) Caucasian, 6.5% (N=25) African American/black, 3.6% (N=14) Hispanic, 3.1% (N=12) Asian, 2.1% (N=8) Native American, and 1.6% (N=6) “other.”

Posttraumatic stress disorder in study 2 was determined with the Detailed Assessment of Posttraumatic Stress. Its diagnosis of PTSD has been shown to have good sensitivity (0.88) and specificity (0.86) relative to the Clinician-Administered PTSD Scale PTSD diagnosis (31). Peritraumatic dissociation was measured with the peritraumatic dissociation subscale of the Detailed Assessment of Posttraumatic Stress, persistent dissociation was evaluated with the Detailed Assessment of Posttraumatic Stress trauma-specific dissociation subscale, and general dissociation was measured with the six subscales of the Multiscale Dissociation Inventory (34). The Multiscale Dissociation Inventory is a psychometrically valid measure of various dissociative responses (33–35) that yields six subscale scores: depersonalization, derealization, disengagement, emotional constriction, memory disturbance, and identity dissociation.

Logistic regression analysis was used in study 2, also with use of the Statistical Package for the Behavioral Sciences, version 11.5. Variables were entered in the following order, based on the assumed chronological sequence of response to trauma. At step 1, the demographic variables of age and sex were entered. At step 2, three types of trauma exposure history, as measured by the trauma-specified section of the Detailed Assessment of Posttraumatic Stress, were added to the equation. These were sexual violence (history of childhood or adult sexual victimization), physical violence (a history of childhood or adult physically violent assaults), and accident or disaster (history of childhood or adult motor vehicle accidents, accidents occurring at home or at work, or disasters). At step 3, the participants’ scores on the peritraumatic distress subscale of the Detailed Assessment of Posttraumatic Stress were entered, followed by scores from the Detailed Assessment of Posttraumatic Stress peritraumatic dissociation subscale at step 4 and the Detailed Assessment of Posttraumatic Stress trauma-specific dissociation subscale at step 5. Finally, at step 6, all six subscales of the Multiscale Dissociation Inventory were added to the equation as a block. After all variables had been entered, the final equation was examined to determine the relative contribution of each variable to PTSD status, with control for all other variables. Of the initial 386 participants, 20 had incomplete demographic or test data, resulting in a final sample size of 366 for this analysis.

**Results**

**Study 1**

Fourteen (26.9%) of 52 trauma-exposed subjects in this group were found to have PTSD on the Clinician-Administered PTSD Scale. Discriminant analysis of participant gender and scores on the Peritraumatic Dissociative Experiences Questionnaire, the Dissociative Experiences Scale, and the trauma-specific dissociation and peritraumatic distress subscales predicting PTSD diagnosis revealed a significant multivariate relationship ($R^2=0.58$, $\chi^2=19.69$, df=5, $p<0.001$). As shown in Table 1, the standardized discriminant function coefficients and post hoc analysis of variance (ANOVA) results indicate that several variables discriminated PTSD-positive versus PTSD-negative status. The univariate ANOVA results revealed that when they were considered individually, scores on the Peritraumatic Dissociative Experiences Questionnaire, the Dissociative Experiences Scale, and the trauma-specific dissociation and peritraumatic distress subscales were no longer related to PTSD. The discriminant function coefficients indicated that although scores on the Dissociative Experiences Scale and the peritraumatic distress and trauma-specific dissociation subscales continued to be meaningful discriminators of PTSD, scores on the Peritraumatic Dissociative Experiences Questionnaire were no longer related.
TABLE 2. Logistic Regression and Univariate Correlation Analyses of PTSD in 366 Subjects With Exposure to Trauma in Study 2

<table>
<thead>
<tr>
<th>Step and Variable</th>
<th>χ²</th>
<th>df</th>
<th>p</th>
<th>Step Analysis</th>
<th>Univariate Analysis</th>
<th>Stepwise Analysis</th>
<th>Analysis of Total Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>3.64</td>
<td>2</td>
<td>n.s.</td>
<td>-0.07</td>
<td>1.06</td>
<td>0.98</td>
<td>0.05</td>
</tr>
<tr>
<td>Sex</td>
<td>0.08</td>
<td>n.s.</td>
<td>1.49</td>
<td>1.78</td>
<td>2.66</td>
<td>12.61</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>18.17</td>
<td>3</td>
<td>≤0.001</td>
<td>0.17</td>
<td>≤0.001</td>
<td>4.38</td>
<td>≤0.04</td>
</tr>
<tr>
<td>Sexual violence</td>
<td>0.20</td>
<td>≤0.001</td>
<td>9.20</td>
<td>≤0.0003</td>
<td>5.13</td>
<td>0.68</td>
<td>2.62</td>
</tr>
<tr>
<td>Physical violence</td>
<td>0.01</td>
<td>n.s.</td>
<td>0.09</td>
<td>0.85</td>
<td>0.21</td>
<td>n.s.</td>
<td>0.64</td>
</tr>
<tr>
<td>Accident or disaster</td>
<td>1.04</td>
<td></td>
<td>0.11</td>
<td>1.10</td>
<td>0.11</td>
<td>n.s.</td>
<td>1.10</td>
</tr>
<tr>
<td>Step 3: peritraumatic distress</td>
<td>23.01</td>
<td>1</td>
<td>≤0.001</td>
<td>0.33</td>
<td>≤0.001</td>
<td>18.81</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Step 4: peritraumatic dissociation</td>
<td>19.48</td>
<td>1</td>
<td>≤0.001</td>
<td>0.39</td>
<td>≤0.001</td>
<td>16.86</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Step 5: trauma-specific dissociation</td>
<td>31.11</td>
<td>1</td>
<td>≤0.001</td>
<td>0.59</td>
<td>≤0.001</td>
<td>19.88</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Step 6</td>
<td>24.29</td>
<td>1</td>
<td>≤0.001</td>
<td>0.46</td>
<td>≤0.001</td>
<td>6.28</td>
<td>≤0.02</td>
</tr>
<tr>
<td>Disengagement</td>
<td>0.46</td>
<td>≤0.001</td>
<td>2.85</td>
<td>1.64</td>
<td>6.28</td>
<td>≤0.02</td>
<td>1.64</td>
</tr>
<tr>
<td>Depersonalization</td>
<td>0.43</td>
<td>≤0.001</td>
<td>2.42</td>
<td>n.s.</td>
<td>0.55</td>
<td>2.42</td>
<td>n.s.</td>
</tr>
<tr>
<td>Derealization</td>
<td>0.44</td>
<td>≤0.001</td>
<td>2.06</td>
<td>n.s.</td>
<td>0.60</td>
<td>2.06</td>
<td>n.s.</td>
</tr>
<tr>
<td>Memory disturbance</td>
<td>0.44</td>
<td>≤0.001</td>
<td>0.11</td>
<td>n.s.</td>
<td>1.10</td>
<td>n.s.</td>
<td>1.10</td>
</tr>
<tr>
<td>Emotional constriction</td>
<td>0.57</td>
<td>≤0.001</td>
<td>6.65</td>
<td>≤0.01</td>
<td>1.55</td>
<td>6.65</td>
<td>≤0.01</td>
</tr>
<tr>
<td>Identity dissociation</td>
<td>0.43</td>
<td>≤0.001</td>
<td>0.02</td>
<td>n.s.</td>
<td>1.06</td>
<td>0.02</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

**Study 2**

A total of 22 (6.0%) of 366 trauma-exposed participants from the general population sample endorsed symptoms on the Detailed Assessment of Posttraumatic Stress that met DSM-IV criteria for PTSD. Univariate correlation analysis indicated that PTSD was associated with exposure to physical and sexual violence on both of the Detailed Assessment of Posttraumatic Stress dissociation subscales (trauma-specific dissociation and peritraumatic dissociation), peritraumatic distress, and all six subscales of the Multiscale Dissociation Inventory (Table 2). Perhaps most noteworthy of this first set of findings is the substantial univariate relationship between persistent dissociation (as tapped by the trauma-specific dissociation subscale) and PTSD, which accounted for 34.8% of the variance in PTSD diagnosis. In contrast, 15.2% of the variance in PTSD was accounted for by peritraumatic dissociation.

As shown in Table 2, stepwise logistic regression analysis revealed that PTSD was unrelated to demographic variables at step 1 but was associated with exposure to sexual and physical violence at step 2; peritraumatic distress, peritraumatic dissociation, and persistent dissociation at steps 3, 4, and 5; and disengagement and emotional constriction at step 6. In other words, the stepwise results indicated that each variable made a significant contribution to the prediction of PTSD, even when all variables entered at prior steps were taken into account. Significantly, peritraumatic dissociation remained a significant predictor at step 4 even after there was control for trauma exposure and posttraumatic distress. When all variables were considered simultaneously (i.e., after step 6), however, peritraumatic distress and peritraumatic dissociation were no longer related to PTSD, whereas persistent dissociation, disengagement, and emotional constriction continued to be significant predictors.

Because persistent dissociation was a major predictor of PTSD status in both studies, a final analysis was conducted to evaluate the use of the Detailed Assessment of Posttraumatic Stress trauma-specific dissociation subscale as a stand-alone predictor of PTSD. This analysis included six additional participants who had missing scores on other scales but whose data were complete for PTSD and the trauma-specific dissociation subscale. When the standard clinical cutoff (a T score of 65 or higher) (34) was used to index the presence or relative absence of persistent dissociation on the trauma-specific dissociation subscale, cross-tabulation with PTSD status indicated a specificity of 0.97 and a sensitivity of 0.57. Examination of Table 3 reveals that trauma-exposed individuals with nonclinical levels of persistent dissociation had a 2.9% likelihood of PTSD, whereas those with clinical levels had a 56.5% likelihood of PTSD ($\chi^2=107.10, df=1, p<0.001$).

**Discussion**

As reported in other investigations, univariate analyses in studies 1 and 2 indicated that peritraumatic dissociation is a significant predictor of PTSD status. However, this relationship ceased to be significant in both studies once persistent dissociation was taken into account by multivariate analyses. In contrast, a substantial relationship was found between persistent dissociation and PTSD at both univariate and multivariate levels. Univariate analyses in study 2 indicated that the persistence of dissociative symptoms after trauma predicted over a third of the variance in PTSD diagnosis, compared to approximately one-sixth of the PTSD variance accounted for by peritraumatic dissociation. This relationship remained in studies 1 and 2 even when the variance associated with peritraumatic and generalized dissociation had been statistically removed. Cross-tabulation of the participants’ trauma-specific dissociation scores with their PTSD status revealed that those with subclinical levels of persistent dissociation were unlikely to experience PTSD, whereas those with clinically elevated levels had more than a 50% likelihood of developing PTSD.
Of interest, the stepwise logistic results of study 2 indicated that although peritraumatic dissociation did not predict PTSD when all other variables were taken into account, it was a significant predictor at step 4, i.e., before persistent dissociation was added but after control was added for trauma exposure and peritraumatic distress. These results suggest that peritraumatic dissociation may not be simply an avoidance response to trauma-related emotional distress, at least in terms of its relationship to PTSD. Removing any variance in peritraumatic dissociation that was associated with exposure to trauma and peritraumatic distress did not eliminate the association of this variable with PTSD. Explanations for this result include the possibility that 1) although peritraumatic dissociation may arise, in part, from peritraumatic distress, other variables not monitored in this study (e.g., reduced capacity for affect regulation) also may be relevant, or 2) peritraumatic dissociation has little to do with a defensive response to trauma-related distress but instead may reflect other processes, for example, the effects of overwhelming autonomic arousal on perception and awareness.

Overall, the findings of studies 1 and 2 suggest that the primary risk for PTSD is less whether one dissociates during (or soon after) a traumatic event than whether such dissociation persists over time. Although it is possible that peritraumatic and persistent dissociation reflect the same underlying phenomenology and function, the temporal component of this response appears to be critical. At the same time, a high level of persistent dissociation is apparently not necessary to produce all cases of PTSD: nearly half of the PTSD-positive individuals in study 2 did not have clinically elevated scores for trauma-specific dissociation. Future research is indicated 1) to determine the specific causal role of persistent dissociation in PTSD, including at what level it may become contributory (for example, persistent dissociation may increase the likelihood of PTSD at lower levels than what is indexed by the established clinical cutoff for trauma-specific dissociation on the Detailed Assessment of Posttraumatic Stress) and 2) to further examine other variables that increase the likelihood of PTSD in the context of persistent dissociation. In any event, the current findings support the contention of Harvey and Bryant (20) that peritraumatic dissociation may have a less powerful relationship to PTSD than initially thought.

The current findings do not directly explain why persistent dissociation is more related to PTSD than is transient (i.e., solely peritraumatic) dissociation, although there are several possibilities. First, persistent dissociation may be an underrecognized “part” of PTSD, in which case it may constitute early evidence of an emerging posttraumatic stress syndrome in a given individual. Second, event-specific dissociation that persists over time may occur as a function of trauma severity, which, in turn, is associated with the likelihood of PTSD (36). The current study attempted to control for this possibility through multivari-ate analysis in which the relationship between persistent dissociation and PTSD was examined after control was added for trauma type (sexual versus physical violence versus accidents or disasters) and how upsetting the individual found the trauma (peritraumatic distress). At a minimum, the current findings do not contradict existing clinical theory, i.e., that persisting trauma-specific dissociation increases the likelihood of PTSD by blocking normal trauma processing, perhaps by decreasing the individual’s ongoing exposure and desensitization to traumatic memories and associated affects over time (37–39).

The current study also examined the association between generalized dissociation and PTSD. In both study 1 and study 2, overall dissociation, indexed by the Dissociative Experiences Scale and the Multiscale Dissociation Inventory, respectively, predicted variance in PTSD independent of that predicted by peritraumatic and persistent dissociation. However, study 2 suggests a smaller role for generalized dissociation once other forms of dissociation are taken into account. Although all six Multiscale Dissociation Inventory subscales were significant univariate predictors, only two (disengagement and emotional constriction) were associated with PTSD at the multivariate level. Of these two, emotional constriction would be expected to correlate with PTSD by virtue of its similarity to the emotional numbing criteria of that diagnosis.

Although generalized dissociation may represent an additional—or perhaps limited—predictor of PTSD, it is possible that the symptoms measured by the Dissociative Experiences Scale or the Multiscale Dissociation Inventory are the sum total of all forms of persisting posttraumatic dissociation that a given individual has experienced across multiple events. A participant’s score on the trauma-specific dissociation subscale reflects dissociation in response to a single specified trauma—not all traumas that the participant ever has encountered. Thus, it may be hypothesized that if trauma-specific dissociation scores were collected for all major trauma exposures per individual, scores on the Dissociative Experiences Scale or the Multiscale Dissociation Inventory might not account for additional variance in PTSD. This supposition is an empirical question, however, that deserves additional research attention.

Interpretation of the findings reported here must be tempered by the limitations of the present study. The recruitment of participants in study 1 was subject to self-selection bias because those who responded to newspaper or other media advertisements may have differed from

<table>
<thead>
<tr>
<th>PTSD</th>
<th>Trauma-Specific Dissociation (T ≥65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>339</td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
</tr>
</tbody>
</table>
those who did not volunteer. Further, the mailing method used to collect the general population sample in study 2 omitted, by definition, those who did not have mailing addresses (e.g., incarcerated, hospitalized, or homeless people), as well as those who could not read English. In addition, the response rate for this sample was below that considered optimal for mailed studies and thus may represent another example of self-selection bias. The effects of these convenience sample characteristics on the generalizability of the current results cannot be fully determined. However, the fact that studies 1 and 2 produced similar results, despite different participants, measures, and statistical approaches, suggests a significant degree of external validity.

The implications of these findings are several. First, the oft-cited notion that peritraumatic dissociation predicts PTSD may overlook the important question of whether such dissociation is sustained over time. The current results suggest that transient peritraumatic dissociative responses predict little independent variance in PTSD relative to persistent dissociation. To the extent that the persistence of dissociation is a more important variable, those studying the effects of dissociation during (or immediately after) a traumatic event may need to inquire further as to the duration of whatever dissociative symptoms have been identified. Second, the fact that peritraumatic dissociation does not make a unique contribution to the prediction of PTSD suggests that the criteria for acute stress disorder might be sharpened by discriminating between peritraumatic and persistent dissociation and by weighing the latter to a greater extent than the former.

Received April 10, 2003; revisions received Jan. 30 and July 7, 2004; accepted Sept. 13, 2004. From the Department of Psychiatry and the Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles; and the Department of Psychology, Auburn University, Auburn, Ala. Address correspondence and reprint requests to Dr. Briere, Psychological Trauma Program, LAC+USC Medical Center, 2020 Zonal Ave., Los Angeles, CA 90033; jbriere@usc.edu (e-mail).

The authors thank Psychological Assessment Resources for access to the standardization data for the Detailed Assessment of Posttraumatic Stress.

References


33. Dietrich AM: Characteristics of child maltreatment, psychological dissociation, and somatoform dissociation of Canadian inmates. J Trauma and Dissociation 2003; 4:81–100
A Prospective Study of Posttraumatic Stress and Depressive Reactions Among Treated and Untreated Adolescents 5 Years After a Catastrophic Disaster

Armen K. Goenjian, M.D.
David Walling, Ph.D.
Alan M. Steinberg, Ph.D.
Ida Karayan, Psy.D.
Louis M. Najarian, M.D.
Robert Pynoos, M.D.

Objective: This study evaluated 1) the natural course of posttraumatic stress and depressive reactions among untreated adolescents from two cities in an earthquake zone (Gumri and Spitak) and one at the periphery (Yerevan) who were differentially exposed to the 1988 Spitak earthquake in Armenia and 2) the effectiveness of brief trauma/grief-focused psychotherapy among adolescents from Gumri.

Method: One hundred twenty-five adolescents were assessed with the Child Posttraumatic Stress Disorder Reaction Index (CPTSD-RI) and the Depression Self-Rating Scale (DSRS) at 1.5 and 5 years post-earthquake. At 1.5 years, trauma/grief-focused group and individual psychotherapy was provided over 6 weeks to a group of students in Gumri.

Results: CPTSD-RI scores among untreated adolescents from Gumri and Spitak subsided significantly but mildly at follow-up, with scores from Spitak, the city at the epicenter, remaining above the cutoff for a diagnosis of PTSD. DSRS scores increased mildly in both earthquake cities but only significantly in Gumri. Among treated adolescents in Gumri, improvement in CPTSD-RI scores was three times that of the untreated Gumri comparison group. The treated group also tended to improve on their DSRS scores, whereas these scores worsened significantly among untreated subjects.

Conclusions: Untreated adolescents exposed to severe trauma are at risk for chronic PTSD and depressive symptoms. Brief trauma/grief-focused psychotherapy is effective in reducing PTSD symptoms and halting the progression of depression. This study supports the implementation of mental health intervention programs in schools after disasters to reduce trauma-related psychopathology.

The Spitak earthquake: on Dec. 7, 1988, a devastating earthquake struck northern Armenia, causing the deaths of more than 25,000 people in an area inhabited by 530,000. It caused massive destruction in four cities and numerous villages. In Spitak and Gumri, two of the most devastated cities, nearly all of the inhabitants experienced direct threats to their lives and witnessed mutilating injuries and grotesque deaths. Five years after the earthquake, only 15% of the reconstruction was completed in both cities. Living conditions were dismal, with people living in crowded, poorly insulated shacks. The omnipresence of debris and destroyed buildings served as constant reminders of the earthquake. The unemployment rate was over 70%.

Literature Review

There is a paucity of literature on the longitudinal course of posttraumatic stress disorder (PTSD) and depression among children and adolescents after disasters. La Greca et al. (1), using the Child Posttraumatic Stress Disorder Reaction Index (CPTSD-RI) (2), evaluated third- to fifth-grade students after Hurricane Andrew. Symptom severity was moderate at 3 months and mild at 10 months after the hurricane. Shaw et al. (3), also using the CPTSD-RI after Hurricane Andrew, evaluated children ages 6 to 11. Children exposed to high levels of trauma, despite some improvement, continued to experience a moderate degree of symptoms at 21 months. McFarlane (4) evaluated “posttraumatic phenomena” with the parent and teacher questionnaires of Rutter and associates among children exposed to the 1983 bush fires in Australia at 8 and 26 months postdisaster. The prevalence of these phenomena did not change over the 18-month follow-up period.

Roussos et al. (5), using the Depression Self-Rating Scale (6), reported low levels of depression 3 months after the 1999 Ano Liosia earthquake in Greece. We, using the same instrument, reported high levels of depression among children and adolescents 1.5 and 3 years after the Spitak earthquake in Armenia (7) and 6 months after Hurricane Mitch in Nicaragua (8). Both Yule and Udwin (9) and we (7) found an increase in levels of depression over time among adolescents with chronic PTSD.

There are few treatment outcome studies among adolescents after disasters. Galante and Foa (10), using a structured classroom intervention at monthly intervals over 1
academic year after the 1980 earthquake in Italy, reported a reduction in earthquake-related fears. This study did not measure posttraumatic stress or depressive reactions, and there was no comparison group. Yule (11) reported the results of cognitive-behavior therapy among girls after the sinking of the Jupiter. At 5–9 months after the disaster, the treated group had lower scores on the Revised Impact of Event Scale and the Birleson Depression Inventory. The authors did not report on the course of distress reactions among the untreated subjects.

Two years after Hurricane Iniki, Chemtob et al. (12) evaluated the outcome of four sessions of psychosocial intervention among children. Using the Kauai Reaction Inventory and a semistructured clinical interview, they found a significant, but mild, reduction of trauma-related symptoms at posttreatment. The reduction was maintained at the 10–12-month follow-up. We reported the findings among adolescents who received trauma/grief-focused group and individual psychotherapy 1.5 years after the Spitak earthquake (7). Three years after the earthquake, there was a significant decrease in posttraumatic stress symptoms and no change in the severity of depressive symptoms among treated subjects. In contrast, both posttraumatic stress and depressive symptoms worsened among untreated adolescents.

Another study of untreated school-age children evaluated 1.5 years after the Spitak earthquake revealed high rates of PTSD, depressive disorders, and separation anxiety disorder among students residing in the two heavily affected cities (13). The present study is a 5-year postearthquake follow-up of a subgroup of those adolescents. The first objective of this study was to determine the course of PTSD and depressive symptoms among untreated subjects from three cities that were differentially affected by the earthquake. The second objective was to determine the effectiveness of trauma/grief-focused psychotherapy on these variables by comparing treated and comparison groups from one of the earthquake cities.

Method

Subjects

A total of 125 subjects were evaluated from three cities in Armenia, Spitak, Gumri, and Yerevan, each city located at increasing distances from the epicenter of the earthquake. Spitak, a city a few miles from the epicenter, was almost totally destroyed. Seventeen percent of the population died. Residents experienced extreme threats to their lives and witnessed mutilating injuries and grotesque deaths. These traumatogenic experiences persisted for days after the earthquake. In Gumri, located 20 miles from the epicenter, exposure to trauma was also severe. Destruction was uniform throughout the city, with 50% of the buildings destroyed and 90% substantially damaged. Approximately 7% of the population was killed. The severity of exposure to trauma was only slightly less than that of Spitak. In Yerevan, the capital of Armenia, located 47 miles from the epicenter, damage to buildings was less extensive, and there were no deaths attributable to the earthquake. However, residents had indirect exposure by way of constant media coverage, with highly disturbing images of the death and destruction in the earthquake zone.

The initial study conducted 1.5 years after the earthquake has been described in detail elsewhere (13). Because of the multiplicity of hardships facing both the victims and the staff working in the earthquake zone, and the lack of available mental health personnel, this study focused on following a subgroup of the original group. For estimation of group size, a power analysis indicated that 21 subjects per study group was the minimum size necessary to obtain significant results (power=0.80, alpha=0.01). For the follow-up, every other untreated subject from the original study was selected. Because the number of treated subjects from the original group was small, all those who had received treatment were targeted for follow-up.

The original group from Spitak that had been previously evaluated at 1.5 years after the earthquake consisted of 63 students. Two subjects were lost at the 5-year follow-up. Thus, the group for this study consisted of 32 students (i.e., every other student from the available 61), with 12 boys of an average age of 16.6 (SD=0.5, range=16–17) and 20 girls of an average age of 16.5 (SD=0.5, range=16–17). The subjects were drawn from two schools from the two main regions of the city. None of the subjects in this group had received psychological treatment. The number of deaths of nuclear family members for the subjects from this group was 11, two of whom were from the same household.

The students from Gumri consisted of two groups: those who had received treatment and those who had not. The original treatment group consisted of 38 students. Two were lost at the 5-year follow-up, leaving 36, with 14 boys of an average age of 16.2 (SD=0.4, range=15–17) and 22 girls of an average age of 16.3 (SD=0.5, range=16–17). The number of deaths of nuclear family members for the subjects in this group was 10, each from a different family. The untreated group originally consisted of 56 students. Two subjects were lost to follow-up. Thus, the group for the present study consisted of 27 students (i.e., every other student from the available 54 students), with 11 boys of an average age of 16.4 (SD=0.5, range=16–17) and 16 girls of an average age of 16.4 (SD=0.5, range=16–17). The number of deaths of nuclear family members for this group was four, each from a different family. The schools in Gumri were selected from the four major regions of the city, where morbidity, mortality, damage, and destruction had been pervasive and uniform. The students who received treatment were from two Gumri schools that were chosen for intervention because of their proximity to our clinic. Because of a lack of clinical resources, we could not provide intervention in the other schools that were screened at 1.5 years after the earthquake in Gumri. Thus, the students in these schools served as comparison subjects for the present study. The fourth group was from Yerevan, the capital of Armenia, located 47 miles from the epicenter. Only one of the original 60 students was lost to follow-up. This group consisted of 30 students (i.e., every other student from the available 59), with 14 boys of an average age of 15.6 (SD=0.5, range=15–16) and 16 girls of an average age of 15.7 (SD=0.5, range=15–16). None of these students had received psychological treatment. There were no deaths of nuclear family members for the subjects in this group.

The principals of the schools that participated gave their approval for the study. The parents were informed of the evaluations and gave written informed consent for their child's participation. The adolescents who participated gave their assent. None declined to participate.

Instruments

Posttraumatic stress symptoms were evaluated by using the 20-item self-report CPTSD-R. Ratings of symptom frequency over the previous month are made on a 5-point Likert scale. The psychometric properties of this instrument have been reported elsewhere, indicating that among Armenian students in the earth-
PTSD AND DEPRESSION AFTER DISASTER

TABLE 1. Child PTSD Reaction Index (CPTSD-RI) Scores Depression Self-Rating Scale (DSRS) Scores and Among Adolescents Differentially Exposed to the 1988 Spitak Earthquake

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spitak Students Who Did Not Receive Treatment (N=27)</th>
<th>Yerevan Students (N=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPTSD-RI scores</td>
<td>Mean SD</td>
<td>Mean SD</td>
</tr>
<tr>
<td>Time 1^a</td>
<td>53.0^b 8.4</td>
<td>41.1^b 7.8</td>
</tr>
<tr>
<td>Time 2^d</td>
<td>47.6 14.2</td>
<td>35.7 11.8</td>
</tr>
<tr>
<td>Change in scores</td>
<td>-5.4 14.2</td>
<td>-5.4 10.9</td>
</tr>
<tr>
<td>DSRS scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1^e</td>
<td>18.7 3.3</td>
<td>14.0 5.2</td>
</tr>
<tr>
<td>Time 2^f</td>
<td>20.0 6.6</td>
<td>16.8 6.1</td>
</tr>
<tr>
<td>Change in scores</td>
<td>1.3 6.2</td>
<td>2.8 6.7</td>
</tr>
</tbody>
</table>

^a Significant difference of across-group comparisons of CPTSD-RI scores at 1.5 years: Spitak students > Gumri students who did not receive treatment > Yerevan students (p<0.001).
^b Significant difference of within-group comparisons of CPTSD-RI and DSRS scores between 1.5 and 5 years (p<0.05).
^c Significant difference of within-group comparisons of CPTSD-RI and DSRS scores between 1.5 and 5 years (p<0.01).
^d Significant difference of across-group comparisons of CPTSD-RI scores at 5 years: Spitak students > Gumri students who did not receive treatment > Yerevan students (p<0.001).
^e Significant difference of across-group comparisons of DSRS scores at 1.5 years: Spitak students > Gumri students who did not receive treatment and Yerevan students (p<0.001).
^f Significant difference of across-group comparisons of DSRS scores at 5 years: Spitak students > Yerevan students (p<0.001).

The trauma/grief-focused treatment incorporated therapeutic techniques from the literature (14–18). An eclectic model of treatment was used that included psycho-education with interpersonal and cognitive-behavior therapy techniques. Therapy addressed five major areas related to trauma that included the following:

1. Reconstruction of earthquake-related experiences immediately before, during, and after the earthquake. This entailed co-construction of a trauma narrative reviewing features of the earthquake experience, with clarification of distortions and misattributions.

2. Identification of trauma reminders, with links made to aspects of traumatic experiences. The students were assisted with increasing tolerance for expectable reactivity to reminders, increasing support-seeking behavior during and after exposure to reminders, and avoiding unnecessary exposure to distressing reminders.

3. Identification of postdisaster stresses and adversities. The students received guidance to help them cope better with changes and losses at home and at school. They were assisted in decreasing maladaptive and avoidant behaviors.

4. Assistance with grief resolution by helping bereaved students reconstitute a nontraumatic mental representation of a deceased person so as to be able to engage in positive reminiscing. The students were helped to identify and engage other individuals who could provide advice, counseling, or companionship.

5. Identification of missed developmental opportunities due to loss of family members or traumatic avoidance. The students were assisted in reengaging in activities that promoted normal developmental progression.

**Statistical Analyses**

Skewness coefficients on all groups indicated a normal distribution of data, thereby permitting the use of parametric tests. Even though the ages of the subjects fell within a narrow range, an analysis of covariance was performed with age as a covariate. The results indicated that there was no age-by-city or time effect. Consequently, age groups were collapsed for further analyses. Analysis of variance (ANOVA) was used to analyze differences among the three groups (Spitak, Gumri, and Yerevan) that did not receive treatment on the CPTSD-RI total score, as well as B, C, and D subcategory scores of DSRS scores, and the change in scores between 1.5 and 5 years for both scales. Additionally, repeated-measures ANOVA was used to examine within-group changes over time for these variables. To examine differences on the distress measures between the two Gumri groups (treatment versus no treatment), t tests were used to analyze each of the distress measures. Sex differences were also analyzed with the t test.

**Results**

**Recovery in Untreated Groups**

At 1.5 years after the earthquake, mean CPTSD-RI scores differed significantly among the three groups (F=34.6, df=2, 88, p<0.0001) (Table 1). Bonferroni post hoc tests revealed a dose-of-exposure effect, with the Spitak group scoring significantly higher than the other two groups (p<0.001). The next highest was the score from the Gumri group, followed by the Yerevan group (p=0.02). The CPTSD-RI scores of both earthquake groups (Spitak and Gumri) were above the cutoff of 40 for a diagnosis of PTSD.

At 5 years after the earthquake, there was also a group effect (F=20.4, df=2, 88, p<0.0001). Bonferroni post hoc tests again indicated that the Spitak group scored higher than the other two groups (p≤0.001) and remained above the cutoff of 40. The Gumri students who did not receive treatment scored significantly higher than the Yerevan group (p<0.05), with both scores below the cutoff. Within-group comparisons of scores showed a significant decrease in CPTSD-RI scores in all three groups (p<0.05). The changes in scores between 1.5 and 5 years among the three groups were not significantly different from one another.

The combined group analysis of mean CPTSD-RI scores showed that girls scored significantly higher at both time
periods: 1.5 years, 52 girls: mean=46.1 (SD=11.4); 37 boys: mean=39.2 (SD=11.0) (t=2.8, df=87, p=0.005); 5 years, 52 girls: mean=40.7 (SD=14.5); 37 boys: mean=32.3 (SD=14.0) (t=2.7, df=87, p < 0.01). Mean change in CPTSD-RI scores between 1.5 and 5 years did not differ significantly between the girls and the boys.

At 1.5 years, DSRS scores indicated a significant group effect (F=13.33, df=2, 87, p <0.001) (Table 1), with the Spitak group scoring significantly higher than both of the other groups, which did not differ from one another. The scores of the Spitak students surpassed the cutoff of 17 for clinical depression, whereas those of the Gumri students who did not receive treatment and the Yerevan students fell below the cutoff. At 5 years after the earthquake, there was also a significant group effect (F=8.67, df=2, 88, p<0.001). Bonferroni post hoc tests indicated that the Spitak group scored significantly higher than the Yerevan group (p<0.001) but not the Gumri group. The DSRS scores for the Spitak group remained above the cutoff for clinical depression, whereas the scores for the Gumri group increased to just below the cutoff of 17. Within-group analyses of DSRS scores indicated that even though there was an increase in scores in the three groups, only the Gumri students who did not receive treatment had scores that increased significantly (F=4.54, df=1.27, p<0.05). The mean change in DSRS scores between 1.5 and 5 years did not differ significantly across the three groups.

The combined group analysis of mean DSRS scores showed that at both time intervals, the girls had higher scores than the boys, but the difference reached significance only 5 years after the earthquake: 1.5 years, 52 girls: mean=16.0 (SD=5.6); 36 boys: mean=14.2 (SD=4.5); 5 years, 52 girls: mean=18.5 (SD=5.9); 37 boys: mean=14.9 (SD=6.4) (t=2.7, df=87, p<0.01). The mean change in DSRS scores between 1.5 and 5 years did not differ significantly between the girls and the boys.

**Outcome of Treatment**

At 1.5 years, mean CPTSD-RI scores did not differ significantly between the two groups (Figure 1): 36 Gumri students who received treatment: mean=44.3 (SD=8.6); 27 Gumri students who did not receive treatment: mean=41.2 (SD=7.8) (t=1.5, df=61, n.s.). At 5 years, however, the scores of the Gumri students who received treatment were significantly lower than those of the Gumri students who did not receive treatment: 36 Gumri students who received treatment: mean=28.1 (SD=10.3); 27 Gumri students who did not receive treatment: mean=35.7 (SD=11.8) (t=2.7, df=61, p<0.01). Within-group comparisons at 1.5 and 5 years indicated that CPTSD-RI scores decreased significantly in both groups (Gumri group with treatment: p<0.001; Gumri group without treatment: p<0.02). Mean change in CPTSD-RI scores between 1.5 and 5 years in the Gumri students who received treatment was three times greater than that of the Gumri students who did not receive treatment—36 Gumri students who received treatment: mean change=–16.3 (SD=13.0); 27 Gumri students who did not receive treatment: mean change=–5.4 (SD=11.0) (t=3.5, df=61, p<0.001). There was no significant difference in the change of CPTSD-RI scores between 1.5 and 5 years among treated girls and boys: 22 girls, mean=15.5 (SD=13.5); 14 boys, mean=17.6 (SD=12.5).

Within-group comparisons indicated that B, C, and D subcategory scores decreased significantly between 1.5
and 5 years in the Gumri students who received treatment, whereas only the B subcategory scores decreased significantly in the Gumri students who did not receive treatment: 36 Gumri students who received treatment—B subcategory score at 1.5 years: mean=1.49 (SD=0.60) (t=2.4, df=35, p=0.02); 27 Gumri students who did not receive treatment—B subcategory score at 1.5 years: mean=2.16 (SD=0.55) (t<2.2, df=36, p<0.1). The mean changes in B and C subcategory scores between 1.5 and 5 years of the 36 Gumri students who received treatment were significantly greater than in those who did not receive treatment (N=27): Gumri students who received treatment B: mean=1.20 (SD=0.91); Gumri students who did not receive treatment B: mean=0.47 (SD=0.86) (t=3.2, df=60, p<0.01); Gumri students who received treatment C: mean=0.57 (SD=0.90); Gumri students who did not receive treatment C: mean=0.14 (SD=0.81) (t=1.94, df=61, p<0.05). There was no significant difference between groups for D subcategory scores.

At 1.5 years, the mean DRS score for the Gumri students who received treatment was significantly higher than that of the Gumri students who did not receive treatment (Figure 1) —36 Gumri students who received treatment: mean=16.9 (SD=4.3); 27 Gumri students who did not receive treatment: mean=14.0 (SD=5.2) (t=2.4, df=61, p=0.02). At 5 years, there was no significant difference between the two groups: 36 Gumri students who received treatment: mean=15.2 (SD=4.0); 27 Gumri students who did not receive treatment: mean=16.8 (SD=6.1) (t=1.2, df=61, n.s.). Within-group comparisons of DRS scores indicated a tendency for decreasing DRS scores in the Gumri students who received treatment (p<0.07) and a significant increase of scores of the Gumri students who did not receive treatment (p<0.05). Of importance, the change in DRS scores between 1.5 and 5 years differed significantly between the two groups, with the Gumri students who received treatment showing improvement and the Gumri students who did not receive treatment showing worsening symptoms: 36 Gumri students who received treatment: mean change=−1.7 (SD=5.4); 27 Gumri students who did not receive treatment: mean change=2.7 (SD=6.7) (t=2.9, df=61, p<0.01). There was no sex effect among the treated subjects: 22 girls: mean=1.0 (SD=4.9); 14 boys: mean=2.7 (SD=6.2).

Discussion

Longitudinal Course of Untreated Subjects

This study extends the literature on the longitudinal course of untreated posttraumatic stress reactions among children and adolescents by presenting findings at 5 years after a disaster. The mild improvement of PTSD symptoms at 5 years in this study augments prior findings of the intractability of PTSD symptoms among untreated adolescents at the 3-year follow-up (7). At both 1.5 and 5 years, the severity of PTSD symptoms among differentially exposed adolescents followed a dose-of-exposure pattern. This type of information is important for the appropriate allocation of mental health resources after disasters.

The CPTSD-RI scores in Spitak, the city at the earthquake’s epicenter, were above the cutoff score of 40 for PTSD at 1.5 and 5 years (53.0 and 47.5, respectively). These high levels of PTSD symptoms are likely attributable to the multiplicity and severity of disaster-related traumatic experiences during and for days after the earthquake. The persistence of PTSD symptoms may be related to the unremitting severe postearthquake stresses and adversities in the recovery environment, such as impoverished living conditions and a lack of food, heat, and electricity. Furthermore, the persistence of PTSD symptoms may have been related to the high levels of comorbid depression that interfered with the resolution of PTSD symptoms (13). Finally, the persistence of symptoms may have been related to the pervasive trauma reminders. Up to the 5-year follow-up, throughout Gumri and Spitak, there were numerous destroyed buildings, makeshift homes, debris, and people on the streets with earthquake-related disabilities.

Consistent with prior studies (8, 13, 19), the girls had slightly higher levels of PTSD symptoms at both time intervals than the boys. Further studies are needed to determine if girls are more willing to endorse symptoms or are at a higher risk for posttraumatic reactions. Despite this difference, the change in the PTSD score of the girls did not differ from that of the boys.

Depressive symptoms also followed a dose-of-exposure pattern at both time intervals. At 5 years, the highest severity levels of depression were found in Spitak. The severity of depression in Spitak at both 1.5 and 5 years (18.7 and 20.1, respectively) fell above the cutoff of 17 for clinical depression. This severity pattern is commensurate with levels found among adolescents in the highly exposed areas of Nicaragua 6 months after Hurricane Mitch (8) and in Armenia 3 years after the earthquake (7). The severity and persistence of depressive symptoms may have been related to a number of factors, including the extensive deaths of family members and friends, destruction and loss of residences, and chronic and pervasive postearthquake stresses and adversities, financial hardship, and unemployment. Prior studies have also indicated that depressive symptoms may be related to the persistence of posttraumatic stress symptoms (5, 7). Finally, given the traumatic circumstances of the deaths, the persistence may have been related to unresolved grief reactions. These high levels of depression underscore the need to evaluate children and adolescents for depression and to plan for the provision of specific interventions aimed at treating depression and traumatic grief. The girls scored higher at 5 years. This finding is consistent with prior findings in Nicaragua (8).
and with the results of epidemiological studies that indicated that depressive disorders show a change in sex distribution after puberty, when the rates for girls begin to exceed those for boys (20).

**Treatment Outcome**

This study also extends the nascent treatment outcome literature on the effectiveness of brief trauma/grief-focused psychotherapy among adolescents after a disaster. The decrease in CPTSD-RI scores among the treated subjects was three times that of the untreated subjects (16.3 versus 5.4, respectively). As an example, a decrease of 16 points from an initial score of 44 on the CPTSD-RI may indicate a decrease in the occurrence of 15 PTSD symptoms from two to three times per week to once per week.

The benefits of treatment manifested themselves primarily in the improvement in B (reexperiencing) and C (avoidance and numbing) subcategory scores. The benefits in the B subcategory were most likely due to strategies used to increase tolerance for reexperiencing phenomena by repeated recounting of the traumatic experiences, cognitive reworking of earthquake-related experiences, and identification and management of psychological and physiological reactivity related to trauma reminders. The beneficial effects of therapy on C subcategory symptoms may have been due to the supportive and educational measures employed to discourage withdrawal and encourage engagement in activities with peers and family members. These findings suggest that periodic assessments of subcategory scores provide a useful metric to monitor progress and accordingly adjust the emphasis of intervention strategies. Despite the significant improvement of PTSD symptoms, treated subjects remained symptomatic 5 years after the earthquake. This indicates the need for further empirical studies that will compare the effectiveness of different modalities of treatment.

Comparisons of the change in depression scores between 1.5 and 5 years showed a significant difference, with a halting of the progression among the treated subjects and a worsening of depression among the untreated subjects. In therapy, a variety of current problems and concerns considered contributory to depression were addressed. These problems included difficulties coping with losses and interpersonal conflicts. Additionally, the students were assisted in dealing with posttraumatic stress and grief reactions, both of which may have contributed to the etiology and exacerbation of depressive symptoms.

As it was prospective, this study obviated potentially confounding methodological problems related to memory failure inherent in retrospective studies. Furthermore, the fact that the subjects were of a similar age, were not seeking treatment, and were ethnically, culturally, and religiously homogeneous reduced the possibility of confounds. The attrition rate, which is usually problematic in longitudinal studies, was low in this study group.

There are limitations to the present study. First, the measures used were self-reports and may therefore have been influenced by either the over- or underreporting of symptoms. Second, even though there was uniformity of destruction, morbidity, and mortality throughout Gumri and baseline CPTSD-RI and DSRS scores were comparable between the treatment and no-treatment comparison groups, it remains possible that factors, such as school milieu or teacher responsiveness, may have played a role in the different courses of recovery. Third, grief reactions were not measured in this study. The inclusion of a grief inventory would have allowed investigation of the extent of grief reactions and their relationship with depression. A pilot study by Cohen et al. (21) demonstrated the usefulness of a grief inventory among traumatized bereaved children.

The finding of the intractability of symptoms has important developmental ramifications. Chronic PTSD symptoms are likely to have detrimental effects on the future of these adolescents, including their conscience functioning (22), academic achievement (23), health, teenage pregnancy, and the stability they bring to family and community life (24). They may be at risk for developing other types of psychiatric disorders and for engaging in risk-taking behaviors (24). Similarly, depressive symptoms may lead to more serious future depressive disorders (25) and have a detrimental effect on their future adaptive functioning and self-esteem (26). Thus, therapy may not only have a beneficial effect on PTSD and depressive symptoms but may also avert the onset of comorbid disorders and prevent psychosocial maladaptation.

This study strongly supports the importance of initiating a comprehensive public mental health program for children and adolescents after natural disasters. The present finding of an enduring effect of treatment provided at 1.5 years after the earthquake suggests that earlier intervention, where possible, should be undertaken and evaluated. Future studies should assess the potential additional benefits of augmenting trauma/grief-focused psychotherapy with specific interventions for depression and adjunctive pharmacotherapy.

References

Frontline Treatment of Combat Stress Reaction: A 20-Year Longitudinal Evaluation Study

Zahava Solomon, Ph.D.
Rami Shklar, Ph.D.
Mario Mikulincer, Ph.D.

Objective: The purpose of the study was to evaluate the long-term (20-year) effectiveness of frontline treatment provided to combat stress reaction casualties.

Method: A longitudinal quasi-experimental design was employed. Participants were combat stress reaction casualties of the 1982 Lebanon War who received frontline treatment (N=79), comparable combat stress reaction casualties who did not receive frontline treatment (N=156), and matched soldiers who did not experience combat stress reaction (N=194). Subjects were asked which of the frontline treatment principles (proximity, immediacy, expectancy) were applied in their treatment, whether or not they returned to their unit after frontline treatment, and if so, whether they returned before or after they felt completely recovered. Outcome assessments included measures of posttraumatic and psychiatric symptoms and of social functioning.

Results: Twenty years after the war, traumatized soldiers who received frontline treatment had lower rates of posttraumatic and psychiatric symptoms, experienced less loneliness, and reported better social functioning than similarly traumatized soldiers who did not receive frontline treatment. In addition, a cumulative effect of application of frontline treatment principles was documented: the more principles applied, the stronger the effect on psychiatric outcomes.

Conclusions: Frontline treatment is associated with improved outcomes even two decades after its application. This treatment may also be effective for nonmilitary precursors of posttraumatic stress disorder.

The stress of combat is notoriously pathogenic. Soldiers are at risk for both short- and long-term psychopathology. Acute combat stress reaction, previously termed battle shock or battle fatigue, encompasses an array of reversible psychiatric and somatic symptoms and impaired functioning. Although persons with combat stress reaction may recover, combat stress reaction often crystallizes into chronic posttraumatic stress disorder (PTSD) and places casualties at risk for chronic PTSD (1).

The development of PTSD is often an evolving process and extends over time through a series of stages ranging from relatively contained distress to severe disability. In its acute phase, combat stress reaction entails considerable distress and shame, yet its damage is limited to military functioning. In its chronic phase, PTSD may be likened to cancer; it metastasizes and is associated with higher rates of psychiatric and somatic comorbidities, substance abuse, impaired functioning, and higher mortality risk (2).

As the disease evolves over time, pathological changes and debilitating comorbidity may become fixed and irreversible. Therefore, the aim in addressing the disorder is to push back the intervention to emphasize preventive rather than curative medicine. Prevention means thwarting the development of the disorder and/or taking measures to halt or slow its progress. The acute phase of combat stress reaction is often seen as a window of opportunity. Since World War I, many armies have adopted frontline treatment as the preferred intervention for combat stress reaction. This treatment was conceptualized by Salmon (3) and later rephrased by Artiss (4) in terms of three principles: 1) proximity—treatment is administered close to the front line, 2) immediacy—treatment is administered close in time to the symptoms’ onset, and 3) expectancy—the expectation is that the soldier will recover rapidly and resume functioning. This psychiatric first aid is time-limited, lasting between 48 and 72 hours. The aim is to meet the casualty’s physiological needs (e.g., food, drink, and sleep), provide temporary relief from harsh battle stressors, enable the exhausted and distressed soldier to regain some control, and decrease hyperarousal. In frontline treatment, human contact is used to reassure the soldier, encourage clarification and sharing of emotions, and humanize and legitimize fears. Frontline treatment allows expression of grief, guilt, and shame but challenges self-depreciation. It conveys to the distressed soldier the expectation of recovery and resumption of functioning.

Frontline treatment is guided by the notion that the soldier’s response is not evidence of an underlying disorder but a natural, appropriate response to the extreme stress of war. The soldier’s symptoms are viewed as stemming from the stress of combat. The intervention includes well-documented stress buffers, such as physical activity, cur-
taintment of avoidant coping strategies (e.g., reference 5), and enhanced social support (e.g., reference 6), and has the further benefit of reducing the well-documented detrimental consequences of psychiatric labeling and the adoption of the sick role.

For the most part, the effectiveness of frontline treatment has been assessed on or near the battlefield in terms of return to duty and, to a lesser degree, recurrence of psychological breakdown. The rate of return to unit after frontline treatment seems impressive. Reports from World War I (e.g., reference 3), World War II (e.g., reference 7), the Korean War (e.g., reference 4), the Vietnam War (e.g., reference 8), and the Gulf War (reference 9) indicate high rates of restored functioning and return to duty. Furthermore, follow-up studies reveal that the vast majority of men who returned to combat duty after treatment performed satisfactorily or better. Although opponents of frontline treatment argue that it is used by the military in the attempt to conserve manpower even at the cost of retraumatizing soldiers (e.g., reference 10), surveys conducted in various war zones consistently report a decline in rates of psychiatric hospitalization and very low rates of recurrence (7, 9, 11) in those who returned to their units.

Yet, despite the extensive use of frontline treatment and its apparent effectiveness on the battlefield, very little systematic research has been conducted (12). War is clearly not an ideal time to conduct evaluation research. In the heat of battle, systemic clinical trials, with random allocation of patients and adherence to double-blind protocols, are hardly feasible.

The unique circumstances in the 1982 Lebanon War allowed for an empirical assessment of the effectiveness of frontline treatment. Although frontline treatment was the intervention endorsed by the Israeli Defense Forces during the war, combat stress reaction casualties were treated either by frontline treatment or in civilian facilities inside Israel. The fact that the location and type of treatment were determined largely by logistic constraints, such as battle conditions, the nature of the terrain, and the availability of transport, provided a quasi-experimental situation. In addition, the combat stress reaction cases that reached mental health personnel in the battle zone were documented, and the records were subsequently computerized. This combination of circumstances allowed a systematic evaluation of the effectiveness of frontline treatment (13). This evaluation, carried out in 1983, revealed the effectiveness of frontline treatment in terms of relatively high rates of return to duty and relatively low rates of PTSD 1 year after the war. Use of each of the frontline treatment principles (proximity, immediacy, expectancy) was associated with better outcomes. Furthermore, rates of PTSD decreased linearly with the application of more treatment principles.

These findings showed the short-term benefits of frontline treatment. The question of whether the benefits are enduring has not been addressed, however. The current study revisits the combat stress reaction casualties of the Lebanon War 20 years later and examines the long-term effects of frontline treatment on their mental state and functioning.

Method

Subjects

This study examined a total of 429 Israeli veterans of the 1982 Lebanon War, including the following three groups: 1) frontline treatment group (N=79, or 88% of the frontline treatment group in the 1983 evaluation), consisting of veterans who received a diagnosis of combat stress reaction and received treatment on the front line in Lebanon; 2) rear echelon treatment group (N=156, or 67% of the 1983 rear echelon treatment group), consisting of combat stress reaction casualties who received their initial treatment in military clinics in Israel; and 3) comparison group (N=194, or 75% of the 1983 comparison group), consisting of soldiers who fought in Lebanon in units that had combat stress reaction casualties but who did not themselves receive a diagnosis of any physical or mental problems during the war. The participants in the 1983 evaluation who did and did not participate in the 20-year follow-up did not differ in sociodemographic background, post-traumatic or psychiatric symptoms, health status, or social functioning, as assessed in 1983.

The mean age of the study subjects was 47.25 years (SD=5.54). Asked to rank their family income, 13% ranked their family income below the national average, 62% as average, and 25% as higher than average. With regard to family status, 92.6% of all respondents were married, 2.1% were single, 5.1% were divorced, and 0.2% were widowers.

When the original study was designed in 1982, the subjects with combat stress reaction and the comparison subjects without combat stress reaction were matched for pre-military, military, and sociodemographic characteristics. It is unlikely that casualties were preselected for mode of treatment on the basis of the severity of their breakdown or any other feature that could bias the findings. A series of statistical analyses carried out for the present study showed no significant difference between the two treatment groups in age, education, economic status, living conditions, or country of origin. Nor did the groups differ in pre-military ratings of intelligence, motivation, rank, corps membership, or prewar adjustment problems or psychiatric symptoms. It is noteworthy that in operational debriefings conducted immediately after the ceasefire in 1983, the mental health officers who first examined the casualties consistently reported that allocation to treatment mode was determined by logistic constraints (see earlier discussion) rather than by any systematic screening.

Measures

Assessment of intervention. In the frontline treatment group, we assessed the application of the frontline treatment principles. Proximity was assessed on the basis of the place of treatment recorded in the soldiers’ medical files. Immediacy was assessed on the basis of subjects’ reports in 1983 about whether they received initial treatment within a day after onset of combat stress reaction symptoms. Expectancy was assessed on the basis of subjects’ reports in 1983 about whether the goal of their initial treatment had been to return them to their units.

For the frontline group, we also assessed whether subjects returned to their units after frontline treatment. Subjects were asked two questions in 1983. They were first asked, “Following treatment, did you return to your unit during the war?” (yes/no). Subjects who answered affirmatively were then asked whether they remained in their units, were assigned their old duties or new ones, or were discharged soon after they returned. Subjects who
reported having gone back to their units were defined as “re-
turned to unit,” whether they resumed their former duties or were
assigned new ones. Subjects who reported not having gone back
to their units or having gone back to their units only to be dis-
charged on arrival were defined as “not returned to unit.”
We also assessed frontline subjects’ perceptions of the timing of
their return to their units. In 1983, subjects were asked: “If you
were returned to your unit during the war, was it 1) after you com-
pletely recovered, 2) before you completely recovered, or 3) when
you were still feeling unfit?”

**Outcome measures.** To enable comparison with findings from
the 1983 assessment, we used the same measures that were used
then, including the PTSD Inventory, the Impact of Event Scale, the
SCL-90-R, the Social Network Inventory, the Problems in Social
Functioning questionnaire, and the UCLA Loneliness Scale.
The PTSD Inventory (14), a self-report scale based on DSM cri-
teria, was used to assess posttraumatic symptoms. The question-
naire consists of statements tapping both DSM-III and DSM-IV
symptom criteria. Subjects are asked to indicate on a 4-point
scale ranging from “never” to “often” the frequency with which
they experienced each of the described symptoms. The scale has
good psychometric properties (13). In the current study, Cron-
bach alphas were 0.96 for symptom intensity, 0.93 for the intru-
sion scale, 0.90 for the avoidance scale, 0.91 for the arousal scale,
and 0.94 for the number of symptoms. Concurrent validity of the
scale was also high (13) when it was compared both with widely
accepted self-report measures (15, 16) and with clinical diagnoses
of PTSD (13, 17).

The Impact of Event Scale (15) is a 15-item scale tapping instru-
ment and avoidance tendencies. Subjects are asked to indicate on
a 4-point scale from “not at all” to “often” how frequently they
experienced each reaction during the previous week. Two total
scores—intrusion and avoidance—were computed by summing
the items corresponding to each scale. This is a widely used mea-
sure in trauma studies, with proven psychometric qualities (16,
17). Cronbach alphas in the current study were 0.95 for intrusion
and 0.88 for avoidance.

The SCL-90-R (18) is a self-report measure that taps psychiatric
symptoms during the 2 weeks before the assessment. It enables
assessment of both the overall severity of psychiatric symptoms
and the severity of each of nine symptom categories: somatiza-
tion, obsessive-compulsive problems, interpersonal sensitivity,
depression, anxiety, hostility, phobic anxiety, paranoid ideation,
and psychoticism, as well as additional symptoms. General dis-
tress is rated on three indices: 1) the global severity index, which
is the mean intensity of the 90 symptoms; 2) the positive symptom
total, which is the total number of positively endorsed symptoms;
and 3) the positive symptom distress index, which is the mean in-
tensity of the positively endorsed symptoms. In the current study,
Cronbach alphas for the subscales ranged from 0.87 to 0.99.
Social support was measured by the 8-item revised Social Net-
work Inventory (unpublished 1978 paper by D. Mueller), which taps
expressive and instrumental support. Consistent with previous
studies (19), Cronbach alpha in the current study was high at 0.91.
The Problems in Social Functioning questionnaire (20) is a 33-
item self-report questionnaire examining problems in different
areas of social and occupational functioning in the previous year.
In the current study, reliability coefficients for social and occupa-
tional functioning were acceptable (alpha=0.76 and alpha=0.82,
respectively).

Feelings of loneliness were assessed with the revised UCLA
Loneliness Scale (21). This scale consists of 20 items, 10 of which
reflect satisfaction with social relationships and 10 of which re-
fect dissatisfaction. The original scale was found to have good
psychometric qualities (21). Similar results were obtained in the
current study (alpha=0.87).

Stressful life events were assessed by a checklist of 20 life events
experienced since the war, including family (e.g., bereavement),
work (e.g., unemployment), and personal (e.g., severe illness)
life events.

**Procedure**
The names and addresses of the participants in the 1983 study
were obtained from updated Army records, and the persons were
contacted by phone. Their participation in the present study was
requested after the study aims were explained. Those who agreed
to participate were offered the choice of meeting in their homes
or at another location of their choice to complete the question-
naire. Before administration of the questionnaire, subjects signed
an informed consent form. Data were collected in 2002.

**Results**

**Long-Term Effects of Frontline Treatment**
First, we examined the differences in PTSD rates in the
three groups. Results revealed a significant group differ-
ence ($\chi^2=33.01, df=2, p<0.01)$. The rate of PTSD in the com-
parison group (13.9% [N=27]) was significantly lower than in both
the frontline treatment group (30.4% [N=24]) ($\chi^2=10.15, df=1, p<0.01$) and the rear echelon treatment group
(41.0% [N=64]) ($\chi^2=33.34, df=1, p<0.001$). However, al-
though combat stress reaction casualties who received
frontline treatment had a lower PTSD rate (30.4% [N=24])
than combat stress reaction casualties referred to rear ech-
elon treatment (41.0% [N=64]), this difference was not sig-
nificant.

In addition, we calculated the relevant odds (risk) ratios
for development of PTSD in the three groups. For the front-
line treatment group versus the comparison group, the
odds ratio was 2.70 (95% confidence interval [CI]=1.44–
5.06; Wald $\chi^2=9.58, df=1, p<0.01$). For the rear echelon
treatment group versus the comparison group, the odds
ratio was 4.30 (95% CI=2.57–7.21; Wald $\chi^2=30.63, df=1,
p<0.001$). For the frontline treatment group versus the rear
echelon treatment group, the odds ratio was 1.59 (95% CI=
0.90–2.84; Wald $\chi^2=2.52, df=1, p<0.05$).

Second, we examined differences in the three groups’
levels of current psychopathology and social functioning.
For this purpose, we conducted one-way multivariate and
univariate analyses of variance (ANOVAs). The dependent
variables were intensity and number of posttraumatic
symptoms, severity of the three clusters of posttraumatic
symptoms, the Impact of Event Scale intrusion and avoid-
ance subscale scores, the three global distress scores (glo-
bal severity index, positive symptom total, and positive
symptom distress index), loneliness score, perceived level
of social support, level of interpersonal functioning, and
level of occupational functioning. Table 1 presents the
means, standard deviations, and F ratios from the analy-
ses. The multivariate ANOVA of scores for the posttrau-
matic symptom clusters of the PTSD Inventory revealed a
significant difference among study groups ($F=15.38, df=6,
846, p<0.01$). As Table 1 shows, univariate ANOVAs and
Duncan’s post hoc tests indicated that the intensity and
Frontline Treatment of Combat Stress

TABLE 1. Scores on Outcome Measures at 20-Year Follow-Up of Israeli Veterans Who Received Frontline Treatment or Rear Echelon Treatment for Combat Stress Reaction in 1982 and Comparison Veterans Without Combat Stress Reaction

<table>
<thead>
<tr>
<th>Measure</th>
<th>Frontline Treatment Group (Group 1) (N=79)</th>
<th>Rear Echelon Treatment Group (Group 2) (N=194)</th>
<th>Comparison Group (Group 3) (N=194)</th>
<th>F (df=2, 326)</th>
<th>Post Hoc Comparisonsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTSD Inventory score</td>
<td>Mean 1.81 SD 0.78</td>
<td>Mean 2.02 SD 0.84</td>
<td>Mean 1.36 SD 0.49</td>
<td>41.28*</td>
<td>1&lt;2; 1, 2&gt;3</td>
</tr>
<tr>
<td>Intensity of symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrusion severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperarousal severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact of Event Scale score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL-90-R Global severity index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive symptom total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive symptom distress index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCLA Loneliness Scale score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems in Social Functioning score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Duncan’s post hoc test (p<0.01).

*p<0.0001.

number of posttraumatic symptoms and the severity of intrusion and hyperarousal were significantly lower in combat stress reaction casualties who received frontline treatment than in those referred to the rear for treatment. As for the avoidance cluster, a similar although nonsignificant difference was observed. The two combat stress reaction groups reported significantly more PTSD symptoms and more severe PTSD symptoms than the comparison group. The multivariate ANOVAs of the Impact of Event Scale scores also revealed a significant difference among study groups (F=16.57, df=4, 848, p<0.01). Univariate ANOVAs and Duncan’s post hoc tests indicated that combat stress reaction casualties who received frontline treatment had a significantly lower Impact of Event Scale intrusion score than those in the rear echelon group, as well as a lower, although not significantly lower, avoidance score. The comparison group had significantly lower scores than the two combat stress reaction groups on both Impact of Event Scale subscales.

A significant multivariate effect of study group was also found in the three global SCL-90-R scores (F=8.10, df=6, 760, p<0.01). Univariate ANOVAs and Duncan’s post hoc tests indicated that combat stress reaction casualties who received frontline treatment reported lower levels of global distress than combat stress reaction casualties who did not receive this intervention, but these differences were not significant. The two combat stress reaction groups endorsed significantly more distress on all three measures than the comparison group (Table 1).

Similar results were noted for the four social functioning measures (loneliness, perceived social support, interpersonal functioning, occupational functioning) (F=8.41, df=8, 832, p<0.01). Univariate ANOVAs and Duncan’s post hoc tests indicated that combat stress reaction casualties who received frontline treatment reported significantly lower levels of loneliness, higher levels of perceived social support, and better interpersonal functioning than combat stress reaction casualties who did not receive this intervention. With regard to occupational functioning, the frontline treatment group reported better functioning than the other combat stress reaction group, but the difference was not statistically significant. Again, the two combat stress reaction groups were found to be significantly more impaired on all social functioning measures than the comparison group.

The Long-Term Effects of Frontline Treatment

To assess the cumulative effect of the application of the various principles of frontline treatment, we examined the relationships between the application of the treatment principles and each outcome measure. The findings showed that when all three treatment principles were applied, the PTSD rate was 25%; when two principles were applied, the rate was 32.6%; and when one principle was applied, the rate was 38.5%. The highest PTSD rate, 47.9%, was found among soldiers who had received none of the treatment principles. Although the differences did not reach significance, the percentages suggest a clear trend. One-way multivariate and univariate ANOVAs revealed a significant relationship between loneliness and the number of treatment principles applied (F=3.21, df=2, 76, p<0.05).

Return to Unit

We also assessed the relationships between return to unit and the other outcome measures. No significant differences were found in any of the outcome measures between combat stress reaction casualties who returned to duty and those who did not.
Perceived Timeliness of Return to Unit

The relationship between perception of the time of return to unit as appropriate and present mental state was also assessed. The results indicated that combat stress reaction casualties who received frontline treatment and who returned to duty before they felt entirely recovered currently had significantly more posttraumatic symptoms (50%) than those who returned after they felt fit (15%) ($\chi^2=4.08$, df=1, $p<0.05$). Those who returned to duty before they felt fit also reported more problems in occupational functioning than those who were returned after they felt fit (mean=0.61, SD=0.36, versus mean=0.26, SD=0.26) ($t=2.98$, df=29, $p<0.01$) and reported more problems in social functioning than those who returned to duty after they felt fit (mean=0.40, SD=0.30, versus mean=0.15, SD=0.30) ($t=2.29$, df=29, $p<0.05$).

Contribution of Type of Treatment and Postwar Stressful Life Events

Since 20 years elapsed between the Lebanon War and the current evaluation, we also assessed the relationship between treatment mode and life events that occurred in the interval. A one-way ANOVA revealed a significant group difference in the number of life events experienced ($F=6.05$, df=2, 425, $p<0.01$). Duncan’s post hoc tests ($p<0.01$) showed that the comparison group reported significantly fewer life events (mean=4.12, SD=2.41) than both the frontline treatment group (mean=4.90, SD=2.75) and the rear echelon treatment group (mean=5.04, SD=2.79) but that the two treatment groups did not differ significantly in the number of life events they reported.

Multiple regression analysis was used to assess the unique contribution of study group (frontline treatment group versus rear echelon treatment group and frontline treatment group versus comparison group) and number of life events since the war to the severity of current PTSD (number of PTSD symptoms endorsed). These variables explained 31.4% of the variance in number of PTSD symptoms ($F=62.44$, df=3, 424, $p<0.001$). Each of the three predictors made a significant unique contribution: life events (beta=0.40, $p<0.001$), frontline treatment versus rear echelon treatment (beta=-0.12, $p<0.05$), and frontline treatment versus comparison group (beta=-0.23, $p<0.001$). Participants who reported a higher number of life events also endorsed a higher number of posttraumatic symptoms. More PTSD symptoms were endorsed by combat stress reaction casualties who did not receive frontline treatment than by the other participants.

Discussion

This study assessed the long-term effectiveness of frontline treatment 20 years after its application in the 1982 Lebanon War. The results reveal that traumatized soldiers who received frontline treatment had lower rates of posttraumatic and psychiatric symptoms, experienced less loneliness, and reported better interpersonal functioning than similarly traumatized soldiers who did not receive frontline treatment. In addition, the findings point to a cumulative effect of the application of frontline treatment principles. We found that when more principles of frontline treatment were applied, psychiatric outcomes were better, although this result did not reach statistical significance. On the other hand, no differences were found in the symptoms or functioning of soldiers with combat stress reaction who returned to their units after receiving frontline treatment and those who did not return to their units. Time of return to unit, however, was related to outcome. Combat stress reaction casualties who received frontline treatment and who felt, in 1983, that they had been returned to their units after they had completely recovered fared significantly better than similarly traumatized men who felt that they had been returned to their units before they had completely recovered.

Several explanations may be offered for the effectiveness of frontline treatment. First, frontline treatment may provide the distressed soldier with temporary relief in relative safety and may thus reduce hyperarousal, which has been found to be associated with risk for chronic PTSD (22, 23), and may help the soldier to see his breakdown as a normal reaction to the pressure of combat and as one that he can overcome. These messages might also help to restore his sense of self-efficacy after the breakdown, as would the demands for activity (e.g., keeping clean, participation in sports, etc.) that are an integral part of frontline treatment.

Frontline treatment may also further recovery by enabling the casualty to continue to enjoy the social support of his comrades and commanders, who can readily visit him during his treatment. This element may help to restore the sense of affiliation that was disrupted by the combat stress reaction (25). Furthermore, social support has consistently been shown to aid recovery from psychiatric crisis (e.g., references 6, 20, 26).

Another possibility is that frontline treatment discourages avoidant coping, which has been shown to predict PTSD among soldiers with combat stress reaction (e.g., reference 5). Unlike soldiers in the rear echelon treatment group, soldiers treated on or near the front remain exposed to combat stimuli similar to those that brought about their breakdown. Their gradual reexposure in a supportive and reassuring situation may lead to desensitization, which enables them to cope with the stimuli and reduces their need to avoid them.
Finally, frontline treatment may help to stem the cascading loss of resources that occurs with combat stress reaction (27). Just as traditional first aid stops hemorrhage and saves the wounded soldier’s life, frontline treatment can be seen as psychological first aid that promotes coping and self-efficacy and minimizes deterioration into the sick role and incapacity.

The main methodological limitation of this study is its quasi-experimental design, which did not meet the strict criteria for controlled clinical trials, such as random allocation of subjects and double-blind protocols. However, such procedures are impossible in combat, and allocation to treatment mode was based on logistic considerations, not severity of the acute reaction.

Nevertheless, the study findings demonstrate the effectiveness of frontline treatment for combat stress reaction even 20 years after combat. Most of the short-term benefits of frontline treatment that were documented two decades ago endured the passage of time. Although frontline treatment did not prevent the development of PTSD in all cases, it did seem to have reduced the risk. In light of the psychological and economic toll of PTSD (2), this finding is potentially of considerable clinical significance.

Our findings support the view that the acute phase of traumatization is a critical period and that early intervention should occur during this window of opportunity to prevent the crystallization of combat stress reaction into entrenched PTSD (28). They also suggest that frontline treatment principles, especially immediacy and expectancy, might be adopted to the treatment of acute stress reaction stemming from nonmilitary traumatogenic events, such as rape, severe illness, accidents, or natural disasters. Such interventions should be accompanied by systematic evaluation research.

Received July 11, 2004; revision received Oct. 19, 2004; accepted Oct. 22, 2004. From Adler Research Center, Tel Aviv University; and the Department of Psychology, Bar Ilan University, Ramat Gan, Israel. Address correspondence and reprint requests to Dr. Solomon, Adler Research Center, Tel Aviv University, P.O. Box 69978, Ramat Aviv, Israel; solomon@post.tau.ac.il (e-mail).

Supported by the Rehabilitation Division of the Israeli Ministry of Defense.

The authors thank the study participants, Anat Damon for statistical assistance, and Danny Horesh for assistance with the preparation of the manuscript.

References

22. Yehuda R, McFarlane AC, Shalev AY: Predicting the development of posttraumatic stress disorder from the acute response to a traumatic event. Biol Psychiatry 1998; 44:1305–1313
25. Marlowe DH: Cohesion, Anticipated Breakdown, and Endurance in Battle: Considerations for Severe and High Intensity Combat. Washington, DC, Walter Reed Army Institute of Research, Neuropsychiatry Division, Department of Military Psychiatry, 1979
Altered White Matter/Gray Matter Proportions in the Striatum of Patients With Schizophrenia: A Volumetric MRI Study

Chiharu Tamagaki, M.D., Ph.D.
Göran C. Sedvall, M.D., Ph.D.
Erik G. Jönsson, M.D., Ph.D.
Gaku Okugawa, M.D., Ph.D.
Håkan Hall, Ph.D.
Stefan Pauli, M.D.
Ingrid Agartz, M.D., Ph.D.

Objective: Anatomical structures of the striatum were studied in 58 patients with schizophrenia and 56 healthy comparison subjects of both genders matched for age and handedness.

Method: Magnetic resonance imaging scans were used to measure gray matter, white matter, and CSF volumes of the caudate, putamen, and nucleus accumbens in the left and the right hemispheres.

Results: White matter/gray matter ratios of the striatal structures were significantly lower in patients than in healthy subjects. In patients, relative white matter volumes in the caudate and nucleus accumbens were reduced, whereas gray matter in the putamen was increased. The total accumbens volume did not differ by diagnosis, but left side accumbens was larger than right in the healthy subjects. The proportion of white matter was greater in women in both the patient and healthy comparison groups. Total caudate and putamen volumes demonstrated no differences due to diagnosis or laterality, but a negative correlation was found in patients between white matter volumes and increasing age. There were no significant correlations among total striatal volumes, white matter/gray matter ratios, age at onset of illness, or illness duration. An estimate of lifetime neuroleptic consumption was positively correlated with right gray matter volume of the putamen in male schizophrenia patients who received typical neuroleptics.

Conclusions: The proportion of white matter to gray matter tissue volumes of the caudate, putamen, and nucleus accumbens is altered in medicated chronic schizophrenia patients, but the total volumes are unchanged.

Since striatal structures play an important role in cognitive function and are highly related to neuroleptic treatment, attempts have been made to investigate cortical-striatal circuits in relation to cognitive deficits in schizophrenia (1). Quantitative magnetic resonance imaging (MRI) studies have focused on alterations of the striatal regions, i.e., the caudate, putamen, and nucleus accumbens. The possibility that striatal regions may be altered by typical neuroleptic medication has been suggested. However, volumetric MRI studies in the striatum of schizophrenia patients have yielded discrepant results. Many studies have reported enlargement of the caudate (2–6) and the putamen (6–8) in patients medicated with typical neuroleptics. Other investigators have found that the caudate volumes were not altered in longitudinal follow-up studies and suggested that the caudate was not changed by either neuroleptic medication or the disease process itself (9, 10). Furthermore, quantitative neuropathological studies comparing striatal structures in schizophrenia have yielded inconsistent results. In postmortem studies, Bogerts et al. (11, 12) and Falke et al. (13) reported that the volumes of the caudate and putamen did not differ between patients with schizophrenia and healthy subjects, whereas other studies have reported increased volume of striatal structures in schizophrenia (14–16). A rat study using dopamine D2 and dopamine D1 ligand binding showed that there were no significant differences in striatal volumes between the animals treated with haloperidol and untreated control rats (17). Thus, the question of whether striatal structures are changed or not in schizophrenia has not been fully answered.

Despite the important role of the nucleus accumbens in cognitive function (18, 19) as well as in reward function, and the specific dopamine D3 receptor distribution to this region, the nucleus accumbens has not been the focus of previous volumetric MRI studies.

In the present study, we measured for the first time gray matter, white matter, and CSF tissue class volumes from segmented magnetic resonance imaging (MRI) scans of the caudate, putamen, and nucleus accumbens in right and left hemispheres of the same subjects. Gray and white matter proportions as well as the total volumes of each of the striatal structures were compared between patients with schizophrenia and healthy subjects of both genders.

Method

Fifty-eight patients (36 men and 22 women) diagnosed with schizophrenia according to DSM-IV and 56 healthy comparison subjects (34 men and 22 women), all Caucasians, were included.
TABLE 1. Demographic and Clinical Characteristics of Patients With Schizophrenia and Healthy Comparison Subjects, by Gender

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Schizophrenia Patients</th>
<th>Healthy Comparison Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (N=36)</td>
<td>Female (N=22)</td>
</tr>
<tr>
<td>Right-handed</td>
<td>32  89</td>
<td>19  86</td>
</tr>
<tr>
<td>Age (years)</td>
<td>40  7</td>
<td>41  8</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>180 7</td>
<td>176 6</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>90  2</td>
<td>75  3</td>
</tr>
<tr>
<td>Body mass index (kg/m²)c</td>
<td>28  6</td>
<td>27  5</td>
</tr>
<tr>
<td>Infracraniare volume (cm³)d</td>
<td>1,537 138</td>
<td>1,378 110</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>26  5</td>
<td>26b 6</td>
</tr>
<tr>
<td>Illness duration (years)</td>
<td>15  7</td>
<td>15b 10</td>
</tr>
<tr>
<td>Estimated total lifetime neuroleptic consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(in haloperidol equivalents)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical neuroleptics</td>
<td>382 100</td>
<td>275 177</td>
</tr>
<tr>
<td>Atypical neuroleptics</td>
<td>354 147</td>
<td>276 186</td>
</tr>
</tbody>
</table>

a Gender effect per two-way ANOVA (p<0.01).
b Data unavailable for one subject.
c Diagnostic effect per two-way ANOVA (p<0.01).

in the study. All subjects lived within the area of northern Stockholm, Sweden. They were recruited at the Department of Clinical Neuroscience, Karolinska Hospital, Stockholm, Sweden. All subjects gave written informed consent for participation in this research, which was approved by the institutional review board (Research Ethics Committee) of the Karolinska Institutet. Exclusion criteria for the healthy subjects were current or past treatment for a psychiatric disorder or psychotic disorder in a first-degree relative. For all subjects, a history of alcoholism or drug addiction, head trauma with loss of consciousness for more than 5 minutes, or a history of somatic disorders affecting brain function were exclusion criteria.

Subject characteristics are presented in Table 1. The patients and the healthy subjects were matched by age and gender as closely as possible. The mean age at the time of scanning was 40.0 years (SD=7.3) for patients and 38.2 years (SD=8.0) for healthy subjects. There were no significant differences in gender and handedness between the groups. All participants were assessed with the Structured Clinical Interview for DSM-III-R (20). In addition, case notes were evaluated for diagnoses according to DSM-III-R and DSM-IV. The mean age at onset of illness was 25.4 years (SD=5.3) for the men and 24.8 years (SD=5.6) for the women. The mean duration of illness was 14.4 years (SD=6.8) and 14.1 years (SD=9.7) for men and women, respectively. At the time of investigation, 52 of the 58 patients were receiving neuroleptic medication. Twenty-five patients (10 women and 15 men) were being treated with typical neuroleptics (haloperidol, perphenazine, levomepromazine, thioridazine, or thiothixene). Twenty-seven patients (eight women and 19 men) were being treated with atypical neuroleptics (risperidone, olanzapine, or clozapine). An approximate estimate of total lifetime consumption of typical and atypical neuroleptics in haloperidol equivalent units (21) was calculated by multiplying duration of illness by the haloperidol equivalent units of current daily neuroleptic dose.

Tissue Segmentation

The segmentation was performed on Silicon Graphics O₂ computers, operative system: IRIX 6.5 (Silicon Graphics, Mountain Valley, Calif.) at the Psychiatry Section, Karolinska Hospital. The segmentation procedure classifies the imaged volume into gray matter, white matter, CSF, nonclass, and venous blood tissue class volumes using BRAINS (22, 23), an image analysis software program suite developed at the Iowa Mental Health Clinical Research Center. This is a semiautomatic system that utilizes MR image intensity. The intracranial volume is automatically obtained. Detailed information on the segmentation procedure and an evaluation of validity has been given by Harris et al. (24). Reproducibility and reliability of the segmentation procedure have also been ascertained previously by our research group (25, 26).

Tracing Methods

The caudate, putamen, and nucleus accumbens were delineated on coronal slices. Traces of the caudate and the putamen were automatically generated using an artificial neuronal network on continuous coronal slices (thickness: 1 mm) in the reconstructed image using the BRAINS software. A detailed description of the artificial neuronal network has been made by Magnotta et al. (27). In the current study, the traces of the artificial neuronal network of the caudate and the putamen were edited manually by following the trace guidelines, available from the University of Iowa Mental Health Clinical Research Center (http://iowa-mhcrc.psychiatry.uiowa.edu/mhcrc/IPLpages/manual_tracing.htm). Briefly, the caudate was delineated from the most anterior coronal section to the most posterior where the caudate structure was visible to the naked eye. The caudate tail was cut at the center of the mamilary body. The medial boundary was the lateral ventri-
The lateral boundary was the internal capsule. The nucleus accumbens was excluded from the caudate measurements.

The tracing of the putamen started from the anterior coronal slice on which the structure was first visible then continued posteriorly on coronal sections until it disappeared. The medial boundary was the internal capsule and the globus pallidus.

The delineation of the nucleus accumbens was done manually in coronal sections. The nucleus accumbens is a cell mass that is the most rostroventromedial part of the caudate-putamen complex (28). The posterior and anterior borders of the nucleus accumbens are hard to reliably delineate. In the anterior aspect, artifacts of the lateral ventricles are seen at the inferior part of the caudate and are hard to differentiate from the boundary of the nucleus accumbens. In the posterior aspect, it is difficult to discriminate the boundary between the nucleus accumbens and the substantia innominata (29) because of their contiguity. Therefore, we defined the posterior boundary of the nucleus accumbens as the slice anterior to the caudate separated from the nucleus accumbens (Figure 1). The delineation was continued anteriorly until the structure disappeared. Gray matter, white matter, and CSF tissue class volumes within the striatal volumes in both the right and the left hemispheres were measured in milliliters using the BRAINS software. The tissue class volumes of gray and white matter but not the CSF were added within each of the striatal structures to obtain the volume of each of the striatal structures. All measurements were completed by a single operator. The operator was blind to subject identity and diagnosis.

**Reliability of Measurements**

Ten scans were randomly selected for interoperator reliability. Two independent operators who were specialists in psychiatry with at least 1 year of postdoctoral training in working with the current image analysis program and who were blind to subject diagnosis and identity performed the tracing of the striatal structures. The volumes of left and right side for each structure were added to compute the intraclass coefficients (ICCs) (30) used to establish interoperator reliability for the entire striatum (ICC=0.93), caudate (ICC=0.96), putamen (ICC=0.85), and nucleus accumbens (ICC=0.79).

**Statistical Analysis**

Before conducting parametric comparisons, data were checked by the Kolmogorov-Smirnov test, which is an established method for checking normal distribution. Differences in age, height, weight, and body mass index among groups were analyzed by using analyses of variance (ANOVAs). Handedness was compared using chi-square tests with chi-square value for the basic analyses of group differences. Relative brain volume measures were used for all group comparisons. The absolute striatal gray matter, white matter, and absolute total striatal structure volumes were adjusted for interindividual differences in intracranial volume by dividing each striatal structure by intracranial volume multiplied by 1,000 to obtain the relative volumes. Two-way ANOVAs were performed on the relative striatal volumes, with gender and diagnostic groups (schizophrenia patients versus healthy subjects) as independent factors. Repeated ANOVAs with side as the repeated factor were performed to compare measurements of left and right side structures between the groups. We also investigated the ratios of white matter to gray matter of the striatal structures by two-way ANOVAs, with gender and diagnostic groups as independent factors. Since the left and right striatal structures were correlated in both patients and healthy subjects, left and right side measurements were summed up in the ratio comparison. Spearman’s rank correlation test was used for the statistical analyses between the striatal volumes and age at MRI, age at onset of schizophrenia, and the duration of illness. We examined the
correlations between estimated lifetime neuroleptic consumption and the relative volumes of gray and white matter in each of the striatal structures by using linear regression analysis. Because of the number of tests performed, a conservative alpha level of 0.01 was used.

Results

There were no significant differences in gender, handedness, or distribution of age between the schizophrenia and comparison groups. The diagnostic groups did differ in body mass index, which was higher in the patients than the healthy subjects (Table 1). There were gender differences in height and weight, with men being taller and heavier than women. The intracranial volume did not significantly differ between the diagnostic groups, but was significantly different between genders (F=51.78, df=1, 112, p<0.001), with smaller volumes seen in the women. The schizophrenia men and women did not differ with regard to age at onset, illness duration, or estimated lifetime neuroleptic consumption.

Striatal Volumes and Volume Ratios

The absolute volumes are presented in Table 2. Absolute volumes were not compared across groups. The comparison of the relative total volumes of the caudate, putamen, and nucleus accumbens revealed no significant group differences across diagnostic or gender groups. As demonstrated in Table 3, relative white matter volumes of the caudate and nucleus accumbens were significantly smaller in the patients (p<0.001), whereas relative gray matter in the putamen was significantly larger in the patients (p<0.01). The results were equal across hemispheres. A significant gender effect was demonstrated for the relative white matter of the nucleus accumbens on the left and right side, with larger volumes in women than men (p=0.01). There were no significant diagnosis-by-gender interaction effects for any of the analyses.

As seen in Table 4, the white matter/gray matter ratios of the caudate, the putamen, and the nucleus accumbens were highly significantly lower in patients compared with healthy subjects (p<0.0001). There were no significant gender or diagnosis-by-gender interaction effects in the ratio comparisons.

Repeated ANOVA revealed a laterality difference in the nucleus accumbens of the healthy subjects only, in absolute as well as relative volumes, left being larger than right (absolute volume: F=11.55, df=1, 110, p=0.0009; relative volume: F=10.37, df=1, 110, p<0.002). There were no laterality differences in the caudate and the putamen in any of the subject groups.

There were negative correlations between age at MRI and white matter volume in the caudate (F=9.73, df=1, 56, p<0.003) and putamen (F=8.71, df=1, 56, p<0.005) in patients but not in healthy subjects. There were no other significant correlations between any of the other striatal tissue class volumes and age at MRI. Spearman rank correlation tests revealed that the striatal volumes were not correlated with age at onset of illness or illness duration. There was a significant correlation between the size of the relative right gray matter volume of the putamen and the estimated lifetime consumption of typical neuroleptics in male schizophrenia patients (F=10.18, df=1, 13, p<0.008). There were no other significant correlations between segmented striatal volumes or white matter/gray matter ratios and estimated lifetime neuroleptic consumption (atypical or typical, alone or combined).
cell death of oligodendroglia in the caudate of schizophrenia, findings in the striatum have not been reported. An number of different brain regions (34, 35), although specific brain imaging has indicated disruption of the white matter in a number of different brain regions (34, 35), although specific findings in the striatum have not been reported. An ultrastructural study by electron microscopy demonstrated cell death of oligodendroglia in the caudate of schizophrenic patients (36). The present white matter reduction in the caudate was in agreement with a study from Japan that used a similar method to investigate the caudate nucleus (37). Taken together, those and the present findings indicate that the white matter reductions seen in the caudate and nucleus accumbens underlie some of the pathophysiological changes observed in patients with schizophrenia. Abnormal myelination might affect connectivity in cortical-striatal circuits and thus contribute to the functional impairment in schizophrenia.

Despite the alterations of the proportion of the two main tissue types in the striatal structures, the total volumes were unchanged. The changes may represent neurodevelopmental abnormalities, pathophysiological effects of chronic disease, a differential effect of neuroleptic treatment in the striatum, or effects of some hitherto unknown factors, alone or combined. The gender effect for the white matter of the nucleus accumbens on both sides with larger volumes in women is interesting, since it cannot be ascribed to a neuroleptic effect or to the disease as such. We found a correlation between lifetime estimate of typical neuroleptic consumption—but not with atypicals nor typical/atypical combinations—and the gray matter of the right putamen in male schizophrenia patients. We used an approximate estimate of lifetime neuroleptic consumption, and the findings might be consistent with the hypothesis that the increase in gray matter seen in the putamen is associated with typical neuroleptic medication. One may assume that subjects with a long treatment history earlier in life would have received typical neuroleptics but later been switched to the atypical antipsychotics (i.e., the esti-

### TABLE 3. Relative Striatal Gray and White Matter Volumes in the Left and Right Hemisphere of Schizophrenia Patients and Healthy Comparison Subjects, by Gender

<table>
<thead>
<tr>
<th>Striatal Area</th>
<th>Relative Volume (ml)</th>
<th>Diagnosis Effect</th>
<th>Gender Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Schizophrenia Patients</td>
<td>Healthy Comparison Subjects</td>
<td>Analysis&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Male (N=36)</td>
<td>Female (N=22)</td>
<td>Mean</td>
</tr>
<tr>
<td>Caudate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>1.65</td>
<td>0.24</td>
<td>1.75</td>
</tr>
<tr>
<td>Gray</td>
<td>1.32</td>
<td>0.20</td>
<td>1.36</td>
</tr>
<tr>
<td>White</td>
<td>0.33</td>
<td>0.11</td>
<td>0.39</td>
</tr>
<tr>
<td>Right</td>
<td>1.61</td>
<td>0.23</td>
<td>1.71</td>
</tr>
<tr>
<td>Gray</td>
<td>1.22</td>
<td>0.20</td>
<td>1.28</td>
</tr>
<tr>
<td>White</td>
<td>0.40</td>
<td>0.10</td>
<td>0.43</td>
</tr>
<tr>
<td>Putamen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>3.07</td>
<td>0.33</td>
<td>3.23</td>
</tr>
<tr>
<td>Gray</td>
<td>1.93</td>
<td>0.27</td>
<td>1.97</td>
</tr>
<tr>
<td>White</td>
<td>1.14</td>
<td>0.19</td>
<td>1.27</td>
</tr>
<tr>
<td>Right</td>
<td>3.07</td>
<td>0.34</td>
<td>3.22</td>
</tr>
<tr>
<td>Gray</td>
<td>1.82</td>
<td>0.26</td>
<td>1.81</td>
</tr>
<tr>
<td>White</td>
<td>1.25</td>
<td>0.19</td>
<td>1.34</td>
</tr>
<tr>
<td>Nucleus accumbens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>0.13</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>Gray</td>
<td>0.13</td>
<td>0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>White</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Right</td>
<td>0.11</td>
<td>0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>Gray</td>
<td>0.11</td>
<td>0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>White</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<sup>a</sup> Absolute volume divided by intracranial volume multiplied by 1,000.

<sup>b</sup> ANOVA; there were no significant interaction effects.

### Discussion

To our knowledge, this is the first study to investigate and report changes in gray matter and white matter tissue proportions in the striatum (i.e., the caudate, putamen, and nucleus accumbens) of the same subjects suffering from schizophrenia. We found the white matter/gray matter ratios of all three structures to be altered in the comparison between patients with chronic medicated schizophrenia and healthy subjects. The ratios were highly significantly lower in the patients for all three structures. When comparing the relative white and gray matter volumes across groups we could confirm that the deviant ratios in the patients were due to smaller white matter volumes in the caudate and nucleus accumbens, whereas in the putamen larger gray matter relative to white matter volumes were demonstrated. The changed proportions between the gray and the white matter volumes were symmetrical across hemispheres. White matter reduction within brain structures suggests a loss of myelin, which is formed and maintained by oligodendroglia (31), while the gray matter increase in the striatum may reflect neuronal changes. In schizophrenia, there is evidence for white matter abnormalities (32, 33). In gene expression studies of schizophrenia, genes associated with myelin abnormalities have been found to be down regulated (32). MR diffusion tensor imaging has indicated disruption of the white matter in a number of different brain regions (34, 35), although specific findings in the striatum have not been reported. An ultrastructural study by electron microscopy demonstrated cell death of oligodendroglia in the caudate of schizophrenic patients (36). The present white matter reduction in the caudate was in agreement with a study from Japan that used a similar method to investigate the caudate nucleus (37). Taken together, those and the present findings indicate that the white matter reductions seen in the caudate and nucleus accumbens underlie some of the pathophysiological changes observed in patients with schizophrenia. Abnormal myelination might affect connectivity in cortical-striatal circuits and thus contribute to the functional impairment in schizophrenia.

Despite the alterations of the proportion of the two main tissue types in the striatal structures, the total volumes were unchanged. The changes may represent neurodevelopmental abnormalities, pathophysiological effects of chronic disease, a differential effect of neuroleptic treatment in the striatum, or effects of some hitherto unknown factors, alone or combined. The gender effect for the white matter of the nucleus accumbens on both sides with larger volumes in women is interesting, since it cannot be ascribed to a neuroleptic effect or to the disease as such. We found a correlation between lifetime estimate of typical neuroleptic consumption—but not with atypicals nor typical/atypical combinations—and the gray matter of the right putamen in male schizophrenia patients. We used an approximate estimate of lifetime neuroleptic consumption, and the findings might be consistent with the hypothesis that the increase in gray matter seen in the putamen is associated with typical neuroleptic medication. One may assume that subjects with a long treatment history earlier in life would have received typical neuroleptics but later been switched to the atypical antipsychotics (i.e., the esti-
The subject material is not fully carried out in the nucleus accumbens before. The investigation also has not been carried out in the nucleus accumbens among subject groups. This kind of significant differences in proportions between functional deficits related to modulating gating of information flow and processing of information within the thalamocortical circuitry (19, 38, 39). Our findings of white matter reduction in the nucleus accumbens in schizophrenia patients (38) confirm a relationship between the segmented striatal structures and illness duration or age at onset. The recent Japanese study also found no correlation between duration of illness and segmented caudate volumes (37).

Contrary to the case of the caudate and the putamen, the nucleus accumbens has only been sparsely investigated in volumetric MRI studies. To our knowledge, there is only one previous volumetric MRI study of the nucleus accumbens, and in that study no significant differences were found in first-episode schizophrenia patients relative to healthy subjects (38). The result of our study confirms that the total volume of the nucleus accumbens is not altered in medicated schizophrenia patients. We found laterality effects in the healthy subjects (larger left side) and greater white matter proportions in women. The nucleus accumbens has been proposed as a critical station for cognitive deficits related to modulating gating of information flow and processing of information within the thalamocortical circuitry (19, 38, 39). Our findings of white matter reduction in the nucleus accumbens in schizophrenia may be related to such cognitive deficits.

The strengths of this study are that we have used a method that is well validated (24, 25). In several studies, it has been demonstrated to have a high interoperator as well as test-retest reliability. Semiautomatically detected striatal volumes have been combined with manual editing/delineation in each MR section. The novelty is that we have segmented MR images into different tissue types that were compared across groups and that we found highly significant differences in proportions between functionally different tissues among subject groups. This kind of investigation also has not been carried out in the nucleus accumbens or putamen before. The subject material is comparatively large and carefully clinically characterized. Both women and men were included. Our subjects are very moderate alcohol consumers, and we have ruled out significant effects of alcohol on the brain volumes in this subject material (unpublished 2005 study).

The limitations of this study are that we were not able to compare tissue class volumes of the striatal structures in initially drug-naive patients in a longitudinal prospective design. We also did not subgroup the patients according to validated information on lifetime neuroleptic medication. The measure of lifetime neuroleptic consumption that we used is an approximation, and we do not have a measure of its validity.

Conclusions

The proportion of white matter to gray matter tissue volumes of the caudate, putamen, and nucleus accumbens is altered in medicated chronic schizophrenia patients, but the total volumes are unchanged. Illness duration appears not to be related to this finding. Further studies are warranted to determine the underlying mechanisms of these changes in schizophrenia.

References


Received April 16, 2004; revision received Jan. 25, 2005; accepted Feb. 4, 2005. From Human Brain Informatics, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; the Department of Neuropsychiatry, Kansai Medical University, Moriguchi, Japan; and the Institute of Psychiatry, University of Oslo. Address correspondence and reprint requests to Dr. Agartz, Institute of Psychiatry, University of Oslo, PB 85 Vinderen, 0319 Oslo, Norway; Ingrid.Agartz@medisin.uio.no (e-mail).

Supported by the Wallenberg Foundation and the Swedish Research Council (grant 2003:5845). The authors thank research nurse Monica Hellberg for her assistance.

TABLE 4. Striatal White Matter/Gray Matter Ratios in Both Hemispheres of Schizophrenia Patients and Healthy Comparison Subjects, by Gender

<table>
<thead>
<tr>
<th>Area</th>
<th>Schizophrenia Patients</th>
<th>Healthy Comparison Subjects</th>
<th>Analysisa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (N=36)</td>
<td>Female (N=22)</td>
<td>Male (N=34)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Caudate</td>
<td>2.91</td>
<td>0.73</td>
<td>3.16</td>
</tr>
<tr>
<td>Putamen</td>
<td>6.46</td>
<td>1.17</td>
<td>6.88</td>
</tr>
<tr>
<td>Nucleus</td>
<td>0.27</td>
<td>0.15</td>
<td>0.34</td>
</tr>
</tbody>
</table>

a ANOVA; there were no significant interaction effects.


Lateral and Medial Hypofrontality in First-Episode Schizophrenia: Functional Activity in a Medication-Naive State and Effects of Short-Term Atypical Antipsychotic Treatment

Beth E. Snitz, Ph.D.
Angus MacDonald III, Ph.D.
Jonathan D. Cohen, M.D., Ph.D.
Raymond Y. Cho, M.D.
Theresa Becker, B.S.
Cameron S. Carter, M.D.

Objective: The dorsolateral prefrontal cortex and the anterior cingulate cortex are critical components of the brain circuitry underlying executive control. The objective of this study was to investigate control-related dorsolateral prefrontal cortex functioning and conflict-related anterior cingulate cortex functioning in a group of never medicated first-episode schizophrenia patients to determine whether both regions show dysfunction at illness onset. A second objective was to assess short-term effects of atypical antipsychotic medication on dorsolateral prefrontal cortex and anterior cingulate cortex functioning.

Method: First-episode schizophrenia patients (N=23) and healthy comparison subjects (N=24) underwent event-related fMRI and performed a cognitive task designed to functionally dissociate the two regions. Four weeks after initiation of pharmacotherapy for patients, a subset of 11 patients and 16 comparison subjects underwent a repeat assessment.

Results: At baseline, patients exhibited hypoactivation in the dorsolateral prefrontal cortex and anterior cingulate cortex. After 4 weeks of antipsychotic treatment, the patients demonstrated improved functioning in the anterior cingulate cortex but not in the dorsolateral prefrontal cortex.

Conclusions: These findings confirm the presence of dorsolateral prefrontal cortex dysfunction early in the course of schizophrenia and suggest that anterior cingulate cortex functioning may be altered at illness onset as well. Results also suggest that anterior cingulate cortex functioning may be especially sensitive to remedial antipsychotic treatment effects. These findings are consistent with an emerging literature documenting short-term benefits of atypical antipsychotic medication for the neural circuitry underlying cognitive deficits in schizophrenia.

Deficits in the executive control of cognition have long been a major research focus in schizophrenia. Functional neuroimaging studies have documented relationships between impairments in executive control and abnormal cortical activation in several regions of the prefrontal cortex. An extensive literature has linked impairments in strategic processes, such as representing and maintaining goals and allocating attentional resources, to dysfunction of the dorsolateral prefrontal cortex (1–9). A second frontal area reported to show altered activation in schizophrenia is the anterior cingulate cortex (10–13) on the medial surface of the frontal lobes. This region is posited to have a critical role in evaluative functions such as self-monitoring of performance (14). Although much has been learned about relationships between these components of executive control and their neural underpinnings, many questions remain concerning their respective roles in the etiology and course of schizophrenia. The present study examined dorsolateral prefrontal cortex and anterior cingulate cortex activity in never-medicated first-episode schizophrenia patients. We used a novel task designed to functionally dissociate the two regions and sought whether dysfunction of either or both of these regions would be observable at illness onset.

The failure of patients to activate the dorsolateral prefrontal cortex is among the most consistent findings in schizophrenia (1, 2, 5–7, 9). Functional magnetic resonance imaging (fMRI) studies of first-episode patients have shown dorsolateral prefrontal cortex hypoactivation (2–4), although hyperactivation has also been reported (8). The executive control function of interest in the present study was context processing (i.e., the ability to use context information to guide task-appropriate behavior), which robustly activates the dorsolateral prefrontal cortex in healthy subjects (2, 5, 15). Specific context-processing deficits have been reported in schizophrenia patients over a range of clinical states (16, 17), as well as in their healthy first-degree relatives (18). These deficits are associated with disorganization symptoms of the illness (4, 7, 17, 19) and are accompanied by reduced dorsolateral prefrontal...
cortex activity (2, 4, 5). We therefore predicted hypoactivation in the dorsolateral prefrontal cortex would be associated with context-processing deficits in this medication-naive, first-episode schizophrenia patient group.

With regard to functional alterations in the anterior cingulate cortex, neuroimaging studies in healthy participants have suggested that the anterior cingulate cortex is responsive to the occurrence of cognitive conflict and signals the need for recruitment of control processes to resolve conflict (14, 20). Schizophrenia patients show reduced error-related anterior cingulate cortex activity and less of a performance adjustment after error commission (11) or associated with response conflict (20), suggesting that an internal monitoring function of the anterior cingulate cortex is impaired in schizophrenia. To our knowledge, anterior cingulate cortex functioning in first-episode patients has not been reported, so it is unknown whether conflict-related anterior cingulate cortex dysfunction is present before possible effects of medication or illness chronicity arise.

Addressing the question of antipsychotic medication effects on dorsolateral prefrontal cortex and anterior cingulate cortex activity was another goal of the current study. Few neuroimaging studies have directly examined changes in functional activation due to treatment. In a recent PET study, Lahti et al. (21) found that clozapine, but not haloperidol, normalized the reduced anterior cingulate cortex activation pattern observed after medication withdrawal. Honey et al. (22) reported increased activation in the right dorsolateral prefrontal cortex and supplementary motor area/cingulate gyrus in chronic patients after substitution of risperidone for typical antipsychotics. Reviews of second-generation atypical antipsychotics agree that they may offer modest benefits for impaired cognition in schizophrenia (23, 24). If so, one would expect to observe corresponding changes between on- and off-medication states in functional brain activation. We tested this hypothesis in the current study, directly comparing dorsolateral prefrontal cortex and anterior cingulate cortex activation during the unmedicated state to that after 4 weeks of stabilization with second-generation antipsychotic medication.

Method

Participants

Participants at baseline were 23 medication-naive patients with first-episode schizophrenia and 24 healthy comparison subjects. Of these participants, a subset of 11 patients and 16 comparison subjects were scanned again after 4 weeks. All participants were part of a larger study of first-episode psychosis and were recruited if they were experiencing any type of psychotic symptom and it was their first psychiatric hospitalization or contact with outpatient psychiatric services. After baseline assessments, patients were treated naturallyistically by their treating psychiatrists and followed longitudinally. Diagnostic confirmation occurred 6 months after index hospitalization through case conferences based on chart review and assessment with the Structured Clinical Interview for DSM-IV (SCID). Two of the patients had limited previous exposure to neuroleptics before baseline, but exclusion criteria were enforced by limiting total lifetime continuous treatment to 2 weeks or less with no more than three doses of oral neuroleptics during the month preceding study entry. Four patients had previous exposure to antidepressants, and two had been exposed to anxiolytics. All psychotropic medications were washed out at least 72 hours before the baseline scan. After baseline, five patients dropped out of the study; two were excluded from follow-up analysis because of poor medication compliance, and one was excluded because of poor comprehension of the task during the second scan.

Healthy comparison participants, recruited through community advertisements, were evaluated with the nonpatient version of the SCID and were excluded for any lifetime history of an axis I disorder or family history of psychotic disorders. Exclusion criteria for both patients and comparison subjects were the following: 1) age older than 50 or younger than 12, 2) DSM-IV mental retardation, 3) substance dependence within the past 6 months or substance abuse within the past month, or 4) lifetime history of significant neurologic disorder or head trauma or current medical condition that could influence CNS function or structure. See Table 1 for summary of demographic and clinical variables.

All participants provided written informed consent after the procedures had been fully explained. Additional safeguards to ensure that informed consent for patients was properly obtained.

### TABLE 1. Demographic and Clinical Characteristics of Medication-Naive, First-Episode Schizophrenia Patients and Healthy Comparison Subjects

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Schizophrenia Patients (Baseline N=23)</th>
<th>Healthy Subjects (Baseline N=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>70</td>
</tr>
<tr>
<td>Right-handed</td>
<td>19</td>
<td>83</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>23.0</td>
<td>5.9</td>
</tr>
<tr>
<td>Parental education (years)</td>
<td>15.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Education (years)</td>
<td>12.9</td>
<td>3.1</td>
</tr>
<tr>
<td>Antipsychotic medication at 4 weeks (N=11)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reality distortion</td>
<td>19.6</td>
<td>5.8</td>
</tr>
<tr>
<td>Poverty</td>
<td>18.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Disorganization</td>
<td>10.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Reality distortion at baseline</td>
<td>10.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Disorganization</td>
<td>15.8</td>
<td>5.7</td>
</tr>
<tr>
<td>Disorganization at baseline</td>
<td>8.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Length of medication treatment (days)</td>
<td>31</td>
<td>20.4</td>
</tr>
<tr>
<td>Antipsychotic medication at 4 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risperidone</td>
<td>7</td>
<td>1–4</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>3</td>
<td>5–10</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>1</td>
<td>300</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose/Dose Range (mg/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Ratings from the Brief Psychiatric Rating Scale, the Scale for the Assessment of Negative Symptoms, and the Scale for the Assessment of Positive Symptoms were used to calculate symptom severity along the three major factors according to Barch et al. (17).<sup>p<0.05.</sup>
TABLE 2. Response Times and Error Proportions of First-Episode Schizophrenia Patients and Healthy Comparison Subjects on a Cognitive Task (Preparing to Overcome Prepotency) at Baseline and After 4 Weeks

| Measure                  | Baseline |       |       |       |       |       |       |       |       |       | 4 Weeks |       |       |       |       |       |       |       |       |       |       |       |       |       |
|--------------------------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                          | Green Square Cue (Ipsilateral Trials) |       |       |       |       |       |       |       |       |       | Green Square Cue (Contralateral Trials) |       |       |       |       |       |       |       |       |       |       |       |       |       |
|                          | Mean     | SD    |       |       |       |       | Mean  | SD    |       |       | Mean    | SD    |       |       |       |       |       |       |       |       |       |       |       |       |
| Healthy comparison subjects | Error proportion | 0.03  | 0.03  | 0.14  | 0.11  | 0.03  | 0.03  | 0.17  | 0.12  |       | 0.06    | 0.07  | 0.15  | 0.14  | 0.03  | 0.02  | 0.12  | 0.07  |       |       |       |       |       |
|                          | Response time (msec) | 572.3 | 185.3 | 622.3 | 193.5 | 510.1 | 134.2 | 522.9 | 151.5 |       | 727.6   | 191.5 | 786.3 | 184.5 | 648.8 | 174.4 | 720.5 | 172.0 |       |       |       |       |       |
| Schizophrenia patients   | Error proportion |       |       |       |       |       |       |       |       |       |         |       |       |       |       |       |       |       |       |       |       |       |       |       |
|                          | Response time (msec) |       |       |       |       |       |       |       |       |       |         |       |       |       |       |       |       |       |       |       |       |       |       |       |

a Subjects were instructed to press a button following presentation of a right- or left-pointing arrow, using their right or left hand depending on the preceding cue. A green square cue meant the subject was to respond with the hand that corresponded with the direction of the arrow probe (i.e., button press with right hand for right-pointing arrow). A red square cue required the subject to use the opposite hand (i.e., button press with left hand for right-pointing arrow). In 70% of the trials, a green square cue was presented.

Cognitive Task

Participants viewed a centrally presented red or green square at the beginning of each trial. An arrow probe appeared 7 seconds later in the same location pointing either to the left or to the right. The green square cue required subjects to respond to the probe with a button press using the ipsilateral hand (i.e., the hand on the same side toward which the arrow was pointing), whereas the red square cue required a response with the contralateral hand, a reversal of the usual stimulus-response mapping. This task, Preparing to Overcome Prepotency, exploits the Simon spatial incompatibility effect (25) in which an imperative stimulus appearing opposite to the response side results in slower response times than when stimulus and response appear on the same side. The Simon effect was augmented by an expectancy manipulation, such that same-sided responding was more frequent (green square cues were presented in 70% of the trials). During the period after a red square cue and before the arrow probe, subjects had to recall and maintain the instructions associated with the cue in order to successfully overcome the prepotent response tendency and respond with the contralateral hand. In this period we predicted a greater activation of the dorsolateral prefrontal cortex with red square cues relative to green square cues. At the time of response to the arrow probe, it was predicted that the conflict inherent in overriding the more automatic response would engage the anterior cingulate cortex.

The Preparing to Overcome Prepotency task was presented in five blocks of 24 trials. Stimulus durations were 500 msec, with a 7-second interval between cue and probe and an 11.5-second interval following the probe to allow for resolution of the hemodynamic response. Subjects practiced the task before being scanned and were instructed to respond as quickly and accurately as possible.

Neuroimaging Methods

Acquisition. Functional scans were acquired using a 3-T GE Signa whole body scanner with a standard head coil. Twenty-eight contiguous axial slices (thickness=3.2 mm) with 3.125 mm2 in-plane resolution were obtained beginning 12.8 mm below the anterior commissure-posterior commissure line. Scans used a single-shot T2*-weighted spiral scanning pulse sequence (TR=1500 msec, TE=18 msec, flip angle=70°, field of view=20 cm), allowing full image acquisition every 1.5 seconds. Thirteen scans were acquired during each 19.5-second trial. Structural images were obtained before and in the same plane as functional images by using a standard T1-weighted pulse sequence. High-resolution anatomical images were collected before functional images.

Preprocessing. Functional images were reconstructed and movement was corrected by using Automated Image Registration (26). A maximum movement criterion of 6 mm or degrees in any direction was applied, which resulted in two comparison subjects and two patients being excluded because of excessive movement. Following these exclusions there were no differences between groups in average movement in any direction (pitch, roll, yaw rotations, or x, y, z plane shift for absolute and incremental movement; baseline: Wilks's lambda=0.66 [F=1.27, df=12, 30, p=0.28]; 4 weeks: Wilks's lambda=0.53 [F=0.97, df=12, 14, p=0.50]). A six-parameter rigid-body algorithm (26) was used to register each participant's structural T1-weighted image to their own high-resolution spoiled gradient-recall acquisition (SPGR) image. A 30-parameter nonlinear warping algorithm (27) was used to align each participant's SPGR image to the Montreal Neurological Institute (MNI) single-subject high-resolution anatomical reference brain (28). Both sets of parameter estimates were applied to the functional T2*-weighted images to bring all participant data into alignment with the MNI reference brain. The data were then smoothed in three dimensions using an 8-mm full width at half maximum kernel to accommodate individual differences in brain morphology.

Statistical Analysis

Given the a priori hypotheses involving the dorsolateral prefrontal cortex and anterior cingulate cortex, we first conducted confirmatory analyses using previously identified functional regions of interest that showed control-related activation within the dorsolateral prefrontal cortex (Brodman's area 9, Talairach coordinates: x=-41, y=18, z=28; 37 voxels [15]) and conflict-related activation within the anterior cingulate cortex (Brodman's area 32, Talairach coordinates: x=0, y=15, z=41; 23 voxels [29]). These regions were brought into the common MNI reference space by using the same aforementioned transformation algorithms. Time series for each region of interest were then generated for each subject, and average change in activation from baseline for the cue period (scans 1–5) and the probe period (scans 6–13) were examined separately. Following the confirmatory analyses, exploratory analyses of the entire image were performed by using voxel-wise analyses of variance (ANOVA)s with subject as a random factor, group as a between-subject factor (comparison subjects versus schizophrenia patients), and cue (green versus red) and scan (1–5 [cue period] versus 6–13 [probe period]) as within-subject factors. fMRI data were analyzed for correct trials only, ensuring analysis of on-task performance. Voxel-wise statistical maps were thresholded for significance at p<0.005 (uncorrected) using an 8-voxel contiguity criterion (30). Correlations between individual differences in brain activity, symptoms, and task performance were also examined.
Results

Task Performance

Table 2 shows basic performance data on the Preparing to Overcome Prepotency task. In a series of repeated-measures ANOVAs (two performance measures and two time points), main effects of condition were observed (all F>5.25, all p<0.03), indicating slower response times and higher error rates across groups on red cue (contralateral) trials compared with green cue (ipsilateral) trials. For response times at both time points, main effects of group were also observed (F>4.70, p<0.04), reflecting slower overall response times in patients. None of the ANOVAs revealed significant group-by-condition interactions (F<2.0, p>0.18).

Functional Neuroimaging at Baseline

Dorsolateral prefrontal cortex. Figure 1 illustrates the time series for dorsolateral prefrontal cortex activity following the cue presentation in the Preparing to Overcome Prepotency task. ANOVA revealed a significant main effect of scan (F=4.04, df=4, 180, p<0.005), a significant group-by-condition interaction (F=8.68, df=1, 45, p<0.01), and a significant group-by-condition-by-scan interaction (F=3.75, df=4, 180, p<0.01). These results reflect increasing activity within the dorsolateral prefrontal cortex in healthy subjects following the red square cue, which signaled the need to overcome the prepotent response to the upcoming probe. In schizophrenia patients, no increased dorsolateral prefrontal cortex activity was observed following ei-
ther cue. Within-group analyses confirmed a significant cue-by-scan interaction (F=3.88, df=4, 92, p<0.01) for comparison subjects but not for patients (F=1.41, df=4, 88, p=0.23).

Anterior cingulate cortex. Figure 1 also shows the time series for anterior cingulate cortex activity following the arrow probe. ANOVA revealed main effects of condition (F=10.01, df=1, 45, p<0.005) and scan (F=37.91, df=7, 315, p<0.001), with a significant condition-by-scan interaction (F=2.71, df=7, 315, p=0.05) and a significant group-by-condition-by-scan interaction (F=2.10, df=7, 315, p<0.05), suggesting that conflict-related modulation of anterior cingulate cortex activity was reduced in patients relative to comparison subjects. Within-group ANOVAs indicated that the condition-by-scan interaction term was significant for healthy subjects (F=2.76, df=7, 161, p<0.01) but not for patients (F=1.95, df=7, 154, p=0.07).

Functional Neuroimaging at 4 Weeks

Dorsolateral prefrontal cortex. Figure 2 displays the time series for dorsolateral prefrontal cortex activity following the cue presentation for both groups after 4 weeks during which the patients had been receiving atypical antipsychotic medication. Between-group ANOVAs revealed a significant main effect of scan (F=6.89, df=4, 100, p<0.001) and a condition-by-scan interaction (F=3.37, df=1, 100, p<0.05). The group-by-condition-by-scan interaction observed at baseline was no longer significant at 4 weeks (F=0.67, df=4, 100, p=0.60), indicating no group difference in the extent to which dorsolateral prefrontal cortex activation was modulated by degree of cognitive control required.

*Baseline time series are presented only from the subset of N=11 patients with 4-week scans.*
**Anterior cingulate cortex.** Figure 2 also illustrates anterior cingulate cortex activity during the probe period. ANOVA resulted in main effects of condition (F=5.96, df=1, 25, p<0.05) and scan (F=25.72, df=7, 175, p<0.001), with a significant condition-by-scan interaction (F=4.19, df=7, 175, p<0.001) indicating that both groups showed similar conflict-related modulation of anterior cingulate cortex activity.

**Effect of Medication**

In order to more directly examine the effect of medication on dorsolateral prefrontal cortex and anterior cingulate cortex functioning, time-series data were analyzed only for the subset of patients who returned for follow-up testing (N=11). Baseline dorsolateral prefrontal cortex and anterior cingulate cortex activity for these patients is shown in the first panel in Figure 2. For dorsolateral prefrontal cortex activity, a two-by-two-by-five within-group ANOVA, with run (baseline versus 4-week), condition (red versus green square cue) and scan number (1 through 5) as within-subject factors resulted in no significant main effects or interactions (all p>0.25).

For anterior cingulate cortex activity, within-group ANOVAs revealed main effects of run (F=5.36, df=1, 10, p<0.05) and scan (F=13.26, df=7, 70, p<0.001), with a significant run-by-condition interaction (F=5.29, df=1, 10, p<0.05) and a significant run-by-condition-by-scan interaction (F=2.53, df=7, 70, p<0.05). The latter result reflects a greater anterior cingulate cortex response to trials with red square cues than green square cues at 4 weeks compared with baseline.

**Exploratory Analysis**

A within-group exploratory analysis across the entire image of the baseline data showed a network of control-related activation (red square cue > green square cue) in both groups during the cue period in lateral frontal regions (left Brodmann's area 10), medial frontal regions (Brodmann's area 8, rostral Brodmann's area 32, and Brodmann's area 24/23), and parietal regions (Brodmann's area 40; right in healthy subjects, left in patients). During the arrow probe period, both groups activated medial frontal regions (right Brodmann's area 9/10) and right primary motor and superior parietal regions (left Brodmann's area 7/40). Healthy comparison subjects activated the caudal anterior cingulate cortex (Brodmann's area 32). A between-group analysis was consistent with the confirmatory analysis for cue activity, with comparison subjects showing significantly greater left dorsolateral prefrontal cortex activity regions (Brodmann's area 8 and Brodmann's area 9) than patients. Bilateral superior parietal regions (Brodmann's area 7 and Brodmann's area 40/7) and right inferior parietal regions (Brodmann's area 7) showed a similar pattern. Although patients failed to show significant anterior cingulate cortex activity during the probe, no significant between-group differences were observed in this region. Patients did show reduced activation in the left inferior frontal gyrus (Brodmann's area 44), supplementary motor area (Brodmann's area 6), right postcentral gyrus (Brodmann's area 3), and left superior parietal cortex (Brodmann's area 7).

**Correlations Between Brain Activation, Symptoms, and Task Performance**

Correlations between maximum dorsolateral prefrontal cortex and anterior cingulate cortex activation during red square trials and symptoms were examined in patients. At baseline, patients with greater disorganization symptoms had lower dorsolateral prefrontal cortex activation (r=–0.58, p<0.01), whereas the correlations with the other two symptom dimensions were not significant (reality distortion: r=0.16; poverty symptoms: r=–0.27). Conflict-related activation in the anterior cingulate cortex during the probe period was associated with both disorganization symptoms (r=–0.54, p<0.05) and poverty symptoms (r=–0.50, p<0.05) but not with reality distortion (r=0.20). At 4 weeks, no significant correlations were found between activation and symptoms. Furthermore, there were no significant correlations between change in symptom ratings and change in maximum dorsolateral prefrontal cortex or anterior cingulate cortex activation from baseline to 4 weeks.

Correlations in comparison subjects at baseline between response time interference and maximum fMRI signal for red square trials were consistent with those reported in MacDonald et al. (15), who used a similar task design (cue-period dorsolateral prefrontal cortex activity: r=–0.35, p<0.05, one-tailed; probe-period anterior cingulate cortex activity: r=0.33, p=0.06, one-tailed). In patients at baseline, neither of these correlations was significant nor were they significant at 4 weeks in either subject group.

**Discussion**

We used a novel task designed to probe independent functioning of the dorsolateral prefrontal cortex and anterior cingulate cortex in medication-naive, first-episode schizophrenia patients. Healthy subjects showed increased activity in the left dorsolateral prefrontal cortex following a cue that signaled the upcoming need to overcome a prepotent response, whereas patients failed to activate this region. These results indicate dorsolateral prefrontal cortex dysfunction early in the course of schizophrenia, before the initiation of medication, consistent with previous reports (1–4, 6). Moreover, patients rated as having severe disorganization symptoms tended to have lower dorsolateral prefrontal cortex activation, consistent with earlier work showing similar associations between disorganization symptoms and both hypofrontality (4, 7) and impaired context processing (17, 19).

We also observed a decrease in anterior cingulate cortex activation in patients relative to comparison subjects, an initial finding indicating, as with the dorsolateral prefrontal cortex, functional impairment at illness onset before long-term exposure to neuroleptic medication. This group

HYPOFRONTALITY IN FIRST-EPISODE SCHIZOPHRENIA

difference was not detected, however, in the exploratory analysis for conflict-related activity in this region, suggesting the effect may not be as large as that observed in the dorsolateral prefrontal cortex (which did show parallel results in both the confirmatory and exploratory analyses). At baseline, decreased anterior cingulate cortex peak activation was associated with greater symptom expression, specifically with poverty and disorganization symptoms. The latter association is consistent with previous findings that anterior cingulate cortex functional abnormalities are associated with disorganization symptoms (31, 32). To what degree observed associations between disorganization symptoms and dorsolateral prefrontal cortex and anterior cingulate cortex dysfunction share common variance is unknown but may have implications for the role of impaired cognitive control at the systems level in behavioral and conceptual disorganization in schizophrenia.

Effects of Antipsychotic Medication

A subset of patients and healthy participants underwent repeat fMRI after the patients had received 4 weeks of atypical antipsychotic treatment. There was no significant improvement in dorsolateral prefrontal cortex function over this interval. In comparison, Honey et al. (22) indicated increased right prefrontal activation during a working memory task with substitution of risperidone for typical antipsychotic medications. However, the average time that patients were treated before the repeat scan (6 weeks) was longer than in the present study (31 days). We are currently conducting a further study with an 8-week stabilization period to ensure that medication effects on cognition and brain functioning, if present, have sufficient time to emerge.

We did observe a significant medication effect on anterior cingulate cortex functioning during the response period of the task. These results are consistent with those of Honey et al. (22) and Lahti et al. (21), who both reported greater supplementary motor area/cingulate (Brodmann's area 6/32) activity with atypical relative to typical antipsychotics. Thus, the present findings contribute to an emerging literature whereby anterior cingulate cortex functioning appears to be among the most consistently sensitive to treatment effects. The mechanism by which short-term atypical antipsychotic regimens enhance task-related anterior cingulate cortex functioning is unknown, but a likely candidate is increased cortical dopamine availability (33), since there is evidence for modulation of impaired anterior cingulate cortex activation by experimental dopaminergic manipulation in schizophrenia (34).

Limitations of the present study include a small number of first-episode patients for the analysis of treatment effects, a relatively brief duration of treatment, and treatment with different atypical antipsychotics with different receptor binding profiles. Since the first two factors would result in lower power to detect treatment effects and the latter in increased variability, our reported negative finding with regard to a treatment effect on dorsolateral prefrontal cortex functioning should be interpreted with some reservation, whereas the treatment effect observed for anterior cingulate cortex functioning can be interpreted with more confidence. Another general limitation was the lack of clear behavioral effect in the scanner that would reflect a context-processing deficit (although some would consider this an advantage, since performance was not different across groups). In contrast, we found that data from the AX version of the Continuous Performance Test in these same participants at baseline indicated a specific context-processing deficit: patients had greater BX errors than healthy subjects (a high context-processing demand condition) but a similar number of AY errors (a task-difficulty control condition; see Barch et al. [17] for task details). We infer from these divergent behavioral results that the Preparing to Overcome Prepotency task is less sensitive to group differences in context processing than the AX version of the Continuous Performance Test possibly because of the long interval between cue and probe used in the scanner. This long delay might have enabled subjects to use a different strategy to maintain a level of performance comparable to that of healthy subjects, such as encoding and rehearsing the cue as an item independent of context, and then retrieving the appropriate stimulus-response mapping at the onset of the probe.

In conclusion, the current study found evidence for hypofrontality of both the dorsolateral prefrontal cortex and anterior cingulate cortex at illness onset. Impairments in frontal lobe functions were associated with core clinical features of schizophrenia before the initiation of antipsychotic medication. Short-term medication effects on conflict-related anterior cingulate cortex activation were observed, whereas medication effects on control-related dorsolateral prefrontal cortex activation were not significant. These findings contribute to an emerging picture in the literature of the demonstrated short-term benefit of atypical antipsychotics to the brain circuitry that underlies impaired cognition in schizophrenia.

Presented in part at the ninth annual meeting of the Organization for Human Brain Mapping, New York, June 18-22, 2003. Received Nov. 1, 2004; revision received Feb. 16, 2005; accepted March 7, 2005. From the Department of Psychiatry, University of Pittsburgh Medical School, Pittsburgh; the Department of Psychology, University of Minnesota, Minneapolis; the Department of Psychology, Princeton University, Princeton, N.J.; and the Department of Psychiatry, University of California, Davis. Address correspondence and reprint requests to Dr. Carter, UC Davis Imaging Research Center, 4701 X St., Sacramento, CA 95817; cameron.carter@ucdmc.ucdavis.edu (e-mail).

Supported by funds received from NIMH (MH-59883 and MH-45156) and NIH/NCCR/GCRC grant M01 RR-00056.

The authors thank Drs. Matcheri S. Keshavan, Gretchen Haas, and Debra Montrose and the clinical core staff of the Conte Center for the Neuroscience of Mental Disorders (MH-45156, D. Lewis, principal investigator) for their assistance in diagnostic and psychopathological assessments. The authors also thank Eric Yablonsky, Avram Holmes, and Jean Miewald for their assistance in data collection and management.
References


2. Barch DM, Carter CS, Braver TS, Sabb FW, MacDonald AW, Noll DC, Cohen JD: Selective deficits in prefrontal cortex function in medication naive patients and schizophrenia. Arch Gen Psychiatry 2001; 58:280–288


5. MacDonald A, Carter C: Event-related fMRI study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. J Abnorm Psychol 2003; 112:689–697


Event-Related Gamma Activity in Schizophrenia Patients During a Visual Backward-Masking Task

Jonathan K. Wynn, Ph.D.
Gregory A. Light, Ph.D.
Bruno Breitmeyer, Ph.D.
Keith H. Nuechterlein, Ph.D.
Michael F. Green, Ph.D.

Objective: Schizophrenia patients experience deficits in many aspects of cognition and perception. Abnormalities in gamma activity may underlie some of these deficits, including rapid processing of visual stimuli. This study examined event-related gamma range activity during a visual backward-masking task in schizophrenia patients and normal comparison subjects.

Method: Event-related gamma activity was recorded in 15 normal comparison subjects and 32 schizophrenia patients. Participants had event-related gamma activity recorded while viewing 60 unmasked visual targets and 240 trials of visual backward masking. Effects of group, accuracy (correct versus incorrect), stimulus-onset asynchrony, and regional activity (left versus right hemisphere, anterior versus posterior regions) were assessed.

Results: Schizophrenia patients had significantly reduced gamma activity in relation to comparison subjects during the backward-masking task. Normal comparison subjects showed significantly greater gamma activity in the right hemisphere, whereas schizophrenia patients did not show this pattern of lateralization. For the unmasked target, there was no group effect and no significant interactions in gamma-band responses.

Conclusions: These results extend previous findings of abnormal gamma range activity in schizophrenia patients. Patients showed overall less gamma activity and failed to show lateralization of activity to the right hemisphere during masking, but they showed comparable levels of gamma activity to unmasked stimuli. Schizophrenia patients’ poorer performance during a masking task may be partly influenced by this abnormal level and the distribution of gamma activity.

Schizophrenia patients exhibit a number of cognitive and basic perceptual abnormalities across all sensory modalities. It has been proposed that failure to maintain cortical-cortical oscillations in the gamma range (30–70 Hz, usually centered around 40 Hz) may subserve many of the perceptual and neurocognitive deficits seen in schizophrenia (1). This article examines the role of event-related gamma oscillations in the commonly reported visual backward-masking deficits in schizophrenia.

Backward masking occurs when a briefly presented visual target is followed shortly by a mask that interrupts processing of the target (2, 3). On the basis of an influential theoretical model, masking occurs when the transient channel activity elicited by the mask interrupts sustained channel processing elicited by the target (2). Schizophrenia patients consistently show deficits on visual backward-masking tasks in that they require a longer interval between the target and the mask to correctly identify the target (4–7).

Recent advances in the visual-masking literature have strongly suggested, on the basis of performance data, that sustained cells at the cortical level exhibit oscillations with activity in the gamma (40–70 Hz) range (8). Masking performance can fluctuate with these cortical oscillations because maximal masking occurs when the transient activity elicited by the mask coincides with the peak of the oscillating activity generated by the target (8–10).

Gamma abnormalities in schizophrenia have recently attracted interest. Two early studies examining gamma abnormalities in the auditory modality have demonstrated reduced gamma activity in schizophrenia patients (11, 12), indicative of a failure to maintain neural activity in the gamma range. More recent studies have shown event-related gamma deficits in schizophrenia in the visual modality as well (e.g., references 10, 13).

The purpose of the current study was to expand upon our preliminary findings of gamma abnormalities during masking in schizophrenia and examine in greater detail how gamma activity is influenced by the strength of the masking effect. The backward-masking task we employed in this study is an appropriate way to study gamma activity because participants must discriminate the form of a target from the mask and process contour information of that target. Perception of visual form presumably depends upon integrating basic visual elements (e.g., line or edge orientation) to construct higher-order conjunctions (e.g., angles or vertices), and finally, to form a coherent representation of an object. It is believed that a key role of gamma activity is to achieve this kind of binding together of visual features. In this study, we examined the effects of masking interval, un-
masked versus masked trials, correct versus incorrect trials, and brain region on gamma activity in schizophrenia patients and healthy comparison subjects.

**Method**

**Participants**

Thirty-six patients with schizophrenia and 16 normal comparison subjects participated in the study. Data from three patients were excluded from the final analysis because of equipment malfunction, whereas data from one patient and one comparison subject were excluded because of invalid behavioral performance. Hence, the final group consisted of 32 patients (two women) and 15 normal comparison subjects (one woman).

All subjects were participating in a larger study of early visual processing (Early Visual Processing in Schizophrenia, M.E.G., principal investigator). The schizophrenia patients were recruited from outpatient treatment clinics at the Veterans Affairs (VA) Greater Los Angeles Healthcare System and through presentations in the community. Twenty-four patients were receiving atypical antipsychotic medication, four patients were receiving typical antipsychotic medication, and four were not taking antipsychotic medication at the time of testing. All patients were administered the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (14) and met diagnostic criteria for schizophrenia. The patients were clinically stable and exhibited mild clinical symptoms. Table 1 shows the mean scores of the BPRS clusters (17) and SANS global scores.

All SCID interviewers were trained by the Treatment Unit of the Department of Veterans Affairs VISN 22 Mental Illness Research, Education, and Clinical Center to a minimum kappa of 0.75 for key psychotic and mood items. All participants had the capacity to give informed consent and provided written informed consent after all procedures had been fully explained in accordance with procedures approved by the institutional review boards at the University of California at Los Angeles and the VA Greater Los Angeles Healthcare System.

Table 1 lists the demographic characteristics of the patient and comparison groups, as well as symptom ratings on the Brief Psychiatric Rating Scale (BPRS) and the Scale for the Assessment of Negative Symptoms (SANS) (16) for the patients. The schizophrenia patients were significantly older than the normal comparison subjects, but the groups did not differ in the amount of education. Because most of our patient participants were recruited from VA clinics, the group was also predominantly men. The patients were clinically stable and exhibited mild clinical symptoms. Table 1 shows the mean scores of the BPRS clusters (17) and SANS global scores.

**Procedures**

**Backward-masking procedure.** The participants performed a computerized backward-masking task while simultaneously having their event-related gamma activity recorded. The participants identified targets in two separate blocks. The first block entailed identifying an unmasked target (60 trials). The second block included the masking procedure, consisting of identifying a masked target (240 trials). The target was a square with a gap in one of three locations (top, bottom, or left side) that could appear at one of four locations on the computer screen (for an example, see Figure 1). The mask was composed of squares that overlapped all four possible target locations. The subjects were required to name the direction of the gap.

Before testing, all participants were equated for unmasked performance with a psychophysical staircase method (18, 19) that adjusted the contrast of the target to obtain a critical stimulus intensity. This procedure entailed adjusting the grayscale of the target, presented for 13.3 msec (two screen sweeps at 150 Hz) until the participant was at 84% for accuracy of identification of the unmasked target. This procedure ensured that any masking differences between groups were not due to a basic visual deficit that prevented the subjects from seeing an unmasked target. The grayscale setting established during the critical stimulus intensity procedure was then used for the unmasked and masked trials.

---

**TABLE 1. Demographic Characteristics for Normal Comparison Subjects and Demographic and Clinical Characteristics for Schizophrenia Patients**

<table>
<thead>
<tr>
<th>Group and Characteristic</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal comparison subjects (14 men and one woman)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>41.8</td>
<td>7.6</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13.4</td>
<td>1.0</td>
</tr>
<tr>
<td>Schizophrenia patients (30 men and two women)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>48.6</td>
<td>8.8</td>
</tr>
<tr>
<td>Education (years)</td>
<td>12.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Brief Psychiatric Rating Scale scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42.7</td>
<td>13.0</td>
</tr>
<tr>
<td>Thinking disturbance</td>
<td>2.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Anxiety or depression</td>
<td>2.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Withdrawal or retardation</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Hostility or suspiciousness</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Scale for the Assessment of Negative Symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global ratings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective flattening</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>Alogia</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Avolition</td>
<td>2.8</td>
<td>1.2</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>2.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Inattentation</td>
<td>1.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>

* The schizophrenia patients were significantly older than the normal comparison subjects (t=2.57, df=45, p<0.05).
GAMMA ACTIVITY IN SCHIZOPHRENIA PATIENTS

FIGURE 2. Performance Data on Visual Backward-Masking Tasks for Schizophrenia Patients and Normal Comparison Subjectsa

<table>
<thead>
<tr>
<th>Stimulus-Onset Asynchrony (msec)</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>Schizophrenia patients (N=32)</td>
</tr>
<tr>
<td>53.3</td>
<td>Comparison subjects (N=15)</td>
</tr>
<tr>
<td>106.7</td>
<td>Schizophrenia patients (N=32)</td>
</tr>
<tr>
<td>160.0</td>
<td>Comparison subjects (N=15)</td>
</tr>
</tbody>
</table>

a Schizophrenia patients performed significantly worse (i.e., correctly identified fewer targets) than comparison subjects.

On the masked trials, the mask was presented for 26.67 msec (four screen sweeps). The targets subtended 0.23° of the visual angle, and each of the four mask locations was 1.03° of the visual angle away from fixation. The background luminance of the computer screen measured 85.7 lux with a handheld light meter with a diffuser held against the screen. Sixty trials were presented in a quasi-random order for four stimulus-onset asynchronies of 0, 8, 16, or 24 screen sweeps, corresponding to stimulus-onset asynchronies of 0.0, 53.3, 106.7, and 160.0 msec. Mean percent correct values were computed separately for the unmasked and masked trials in the patient and comparison groups.

Event-related gamma recording. The participants had their EEG activity recorded during the unmasked and masked tasks. Stimulus presentation and data synchronization with the EEG were accomplished with E-Prime (Psychology Software Tools, Inc., Pittsburgh). EEG activity was collected continuously throughout the session and amplified with a Neuroscan NuAmps amplifier (Compumedics USA, El Paso, Tex.). Data were sampled at 1,000 Hz, with filter settings of 0 to 100 Hz. Thirty-two cap-mounted, sintered silver-silver chloride electrodes (Falk Minow Services, Herrsching-Breitbrunn, Germany) were positioned with a modified international 10–20 system-placement scheme. Additionally, four electrodes were used to measure a horizontal electrooculogram (EOG) (placed on the outer canthus of the left and right eyes) and a vertical EOG (placed above and below the left eye). All electrodes were referenced to the nose, and a forehead ground was employed.

All data were processed with Neuroscan Scan 4.3 software (Compumedics USA, El Paso, Tex.). Data were first high-pass filtered at 1 Hz and then epoched to 200 msec before and 823 msec after the stimulus. Artifact rejection was performed for any trial that exceeded ±75 µV at electrode sites Fz, F3, F4, and Cz. For unmasked trials, an average of 14.3% of the trials were rejected for the schizophrenia patients and 7.5% for the normal comparison subjects. For the masked trials, an average of 16.7% of the trials were rejected for the schizophrenia patients and 7.2% for the normal comparison subjects.

Correct and incorrect responses to the unmasked and masked trials were examined separately at each stimulus-onset asynchrony. Electrode sites were averaged to examine regional effects.

Results

Backward-Masking Performance

The patients were significantly older than the normal comparison participants (p<0.05) (Table 1). The patients tended to need a higher grayscale level (i.e., darker target) than the comparison subjects, although this difference did not reach statistical significance (t=1.86, df=45, p=0.07). The schizophrenia patients correctly identified an average of 80.2% (SD=9.3%) of the unmasked targets, whereas the normal comparison subjects identified an average of 85.3% (SD=10.3%)—a statistically insignificant difference (t=1.71, df=45).

The results of the performance data during the masked trials are shown in Figure 2. There were significant main effects of group (F=4.15, df=1, 45, p<0.05) and stimulus-onset asynchrony (F=125.77, df=3, 135, p<0.001, ε=0.644). The interaction between group and stimulus-onset asynchrony was not significant. The significant group effect held when age was entered as a covariate. These results show that the normal comparison subjects correctly identified a greater number of targets than the schizophrenia patients and that both groups showed the expected improvement in target detection as stimulus-onset asynchrony increased. These data are similar to what we have seen in previous behavioral studies (e.g., reference 10).

Event-Related Gamma Results for Unmasked Trials

The data for event-related gamma activity during the unmasked trials were subjected to a two-by-two-by-two (group [patients versus comparison subjects] by correct versus incorrect trials by anterior versus posterior...
Event-Related Gamma Results for Masked Trials

The data for the event-related gamma activity during the masking task were subjected to a two-by-two-by-four-by-two-by-two (group [patients versus comparison subjects] by correct versus incorrect trials by stimulus-onset asynchrony by anterior versus posterior by left versus right hemisphere) repeated-measures ANOVA. The results showed a significant main effect of group (F=5.02, df=1, 45, p<0.03), a significant group-by-hemisphere interaction (F=4.29, df=1, 45, p<0.05), and a significant group-by-number correct-by-stimulus-onset asynchrony interaction (F=3.24, df=3, 135, p<0.03, ε=0.932).

The main effect of group revealed that, overall, the normal comparison subjects produced significantly greater gamma (mean=1.25, SD=0.27) than the schizophrenia patients (mean=1.05, SD=0.28). The significant group-by-hemisphere interaction revealed that the normal comparison subjects produced significantly greater gamma activity in the right hemisphere (mean=1.30, SD=0.31) than in the left (mean=1.20, SD=0.29), whereas the schizophrenia patients showed no significant differences between the right (mean=1.04, SD=0.31) and left (mean=1.06, SD=0.29) hemispheres (Figure 3). Both of these findings held when age was entered as a covariate.

The significant three-way interaction was explored with follow-up t tests. A difference score was computed between correct and incorrect gamma scores at each stimulus-onset asynchrony separately for the patients and the comparison subjects. The normal comparison subjects tended to produce greater gamma activity to incorrect versus correct trials at a stimulus-onset asynchrony of 53.3 msec (t=2.08, df=14, p<0.06), whereas the schizophrenia patients tended to produce greater activity to correct versus incorrect trials at a stimulus-onset asynchrony of 53.3 msec (t=1.69, df=31, p<0.11).

Discussion

The results of the present study provide further support for an event-related abnormality in gamma activity in schizophrenia patients. Group differences were seen both in the amount and distribution of gamma activity during the backward-masking procedure. These results, moreover, are specific to targets presented in the presence of a mask. The lack of a group difference in gamma activity to unmasked stimuli suggests comparable processing of unambiguous visual stimuli in patients and comparison subjects.

The finding of a group-by-hemisphere difference in gamma activity in masked trials is consistent with previous findings of a deficit in the right hemisphere in schizo-

---

FIGURE 3. Grand Average z Score Gamma Waveforms for Schizophrenia Patients (red) and Normal Comparison Subjects (green)

Schizophrenia patients showed significantly less gamma activity in the 50–200 msec range, especially pronounced in the right hemisphere (indicated by the black arrows), than normal comparison subjects.
phrenia for processing visual information (e.g., references 20–22). For example, Nuechterlein and colleagues (23) scanned subjects with positron emission tomography (PET) during a perceptually demanding continuous performance test and found that the schizophrenia patients did not use the right hemisphere preferentially to process ambiguous visual stimuli as normal subjects did (24). With simple dot-pattern stimuli, Heckers et al. (22) found that normal comparison subjects showed significant decreases in PET activity in the right hemisphere from the first viewing to the last viewing of the dot pattern, suggesting that habituation of activity in the right hemisphere is indicative of processing simple visual stimuli. The schizophrenia patients showed a reversed asymmetry, suggesting that their left hemisphere was overactive in processing visual stimuli. On the basis of the gamma activity findings of the current study, it appears that schizophrenia patients lack the normal right hemisphere lateralization to process ambiguous visual information and that this deficit extends to the backward-masking task.

The finding of a significant interaction between group, stimulus-onset asynchrony, and accuracy (correct versus incorrect) was due to a tendency for the comparison subjects to produce greater gamma activity to incorrect trials at a stimulus-onset asynchrony of 53 msec, whereas the schizophrenia patients showed the reverse pattern. This pattern occurred at only one stimulus-onset asynchrony and was not predicted. We do not have an interpretation at this time but will see if the pattern replicates itself.

The findings in the present study add to a growing literature demonstrating that schizophrenia patients exhibit reduced gamma activity to simple stimuli. Previous studies with auditory stimulation have shown that schizophrenia patients are less able to engage or maintain gamma activity (11, 12). For example, Kwon et al. (11) found that schizophrenia patients failed to show EEG synchronization to 40-Hz steady-state auditory trains. In a separate study, Clementz et al. (12) found that gamma-band (e.g., 30–50 Hz) abnormalities in the auditory P50 suppression response were better than the broadband P50 response (e.g., 10–50 Hz) at detecting differences between patients and comparison subjects. Although these studies used a different modality than we did, they similarly examined time-locked, or evoked, gamma responses. The results across these studies suggest that schizophrenia patients may have abnormalities in establishing or maintaining gamma activity to simple auditory and visual stimuli. These abnormalities, in turn, may account for some of the perceptual deficits in schizophrenia.

Gamma oscillations at the neuronal level are thought to be modulated by γ-aminobutyric acid (GABA)-ergic interneuronal circuits (e.g., references 25, 26), as well as N-methyl-D-aspartic acid (NMDA)-modulated pyramidal cells (e.g., references 27, 28). Schizophrenia has been associated with abnormalities in the GABA and NMDA systems, which may explain the gamma abnormalities that have been observed. It is possible that such abnormalities in gamma synchronization could lead, through a cascade of processes, to mistakes in perceptual processing and misinterpretation of ambiguous stimuli.

Our study had a few limitations. The stimuli we used were very small (0.23° of the visual angle), which probably reduced the amount of gamma activity seen. Recent research has shown that the amount of gamma to visual stimuli increases with increasing size of the stimuli (29). For the current study, we intentionally selected stimuli that were identical to those used in our behavioral studies (30). These stimuli need to be relatively small and faint to establish each subject’s perceptual threshold.

Another limitation is that the study has a limited number of stimulus-onset asynchronies. Although this was intentional so as to make the length of the testing session tolerable to the participant, it had the disadvantage of making it more difficult to detect any stimulus-onset asynchrony-dependent, event-related gamma activity. Future studies that use a wider range of stimulus-onset asynchronies may be in a better position to detect stimulus-onset asynchrony-dependent gamma activity. On the basis of the results from the unmasked trials, we expect that gamma activity would be normal in schizophrenia patients at the longest stimulus-onset asynchronies when performance approaches unmasked levels.

The patients and comparison subjects differed in age. However, as we noted, covarying age did not change any of the significant results. Age also did not correlate with any gamma measures of interest (e.g., overall gamma activity, gamma to correct versus incorrect trials, etc.), further reducing the possibility that our results are due to age differences. Age effects on performance were minimized by matching all subjects for unmasked target discrimination with a psychophysical staircase method.

Although visual-masking procedures appear to be excellent probes for aberrant gamma activity in schizophrenia, this study does not clarify whether abnormal gamma activity is the main cause of masking deficits in schizophrenia. Evidence has accumulated showing that schizophrenia patients have abnormal transient activity (7, 31, 32), and this abnormality may partially account for visual-masking deficits. However, gamma-range activity in backward-masking tasks is considered to be a function of sustained channels (8, 9). Therefore, although the differences in gamma activity in the current study are provocative, the role that abnormal gamma activity plays in masking performance deficits in schizophrenia remains unclear.

Finally, our study examined only evoked—but not induced—gamma activity. Evoked gamma activity is phase-locked to the onset of a stimulus and can be detected by averaging EEG responses. This is indicative of early sensory registration or processing of stimuli (33). Induced gamma activity is a non-phase-locked activity, is not seen with traditional averaging techniques, and is indicative of perceptual organization or feature binding (33). Because
backward masking interrupts the very early stages of target processing, we were primarily interested in the early evoked gamma response (50–200 msec). However, a close examination of induced gamma activity in future studies could provide further insight into the gamma deficits seen in schizophrenia patients.

In conclusion, the current study extends the findings of abnormal gamma activity in schizophrenia patients during a backward-masking task (10). The current study employed a masking paradigm to elicit gamma activity. Decreased gamma activity was seen in schizophrenia patients only during the masking task, consistent with theories that abnormalities of generating or maintaining gamma activity can contribute to perceptual aberrations in schizophrenia (34). We further found that these deficits are especially pronounced in the right hemisphere, an area that is preferentially responsible for processing ambiguous visuospatial information. Although the current study has several limitations, these results continue to demonstrate the usefulness of using backward-masking procedures to explore underlying neurophysiological abnormalities in schizophrenia.


References

1. Green MF, Nuechterlein KH: Cortical oscillations and schizophrenia: timing is of the essence. Arch Gen Psychiatry 1999; 56:1007–1008
12. Clementz BA, Blumenfeld LD, Cobb S: The gamma band response may account for poor P50 suppression in schizophrenia. Neuroreport 1997; 8:3898–3893
32. Green MF, Nuechterlein KH, Mintz J: Backward masking in schizophrenia and mania, II: specifying the visual channels. Arch Gen Psychiatry 1994; 51:939–944
Neurological Soft Signs in First-Episode Schizophrenia: A Follow-Up Study

Silke Bachmann, M.D.
Christina Bottmer, M.A.
Johannes Schröder, M.D.

Objective: Neurological soft signs are frequently found in schizophrenia. They are indicators of both genetic liability and psychopathological symptoms. To further differentiate “trait” and “state” relations the authors compared the 1-year course of neurological soft signs in schizophrenia patients and comparison subjects.

Method: Thirty-nine patients with first-episode schizophrenia spectrum disorders were examined after remission of acute symptoms and 14 months later. Established instruments assessed diagnoses, psychopathological symptoms, predictors of outcome, handedness, and neurological soft signs. Twenty-two age- and gender-matched comparison subjects were also examined twice.

Results: Neurological soft sign scores in patients were significantly elevated relative to comparison subjects at both measurement points. Whereas neurological soft signs remained stable in comparison subjects (time 1: mean=4.8, SD=3.3; time 2: mean=4.6, SD=3.9), they significantly decreased in patients (time 1: mean=15.7, SD=7.1; time 2: mean=10.1, SD=7.9). This effect was more pronounced in patients with a favorable versus a chronic course and was mainly accounted for by motor signs. Predictors of follow-up neurological soft sign scores were neurological soft sign levels at remission and compliance with treatment.

Conclusions: Although neurological soft signs are intrinsic to schizophrenia, their level varies with the clinical course. Thus, neurological soft signs may correspond to both genetic liability and the activity of the disease process and may be considered as potential predictors of outcome.

Neurological soft signs refer to subtle neurological abnormalities comprising deficits in sensory integration, motor coordination, and sequencing of complex motor acts (1). A considerable body of research has established that neurological soft signs are more prevalent in schizophrenia patients, including first-episode cases, than in healthy subjects (2). Studies with neuroleptic-naive first-episode patients have demonstrated that neurological soft signs are present before medication exposure, thus they are thought to be an intrinsic feature of schizophrenia (3, 4). This notion is supported by findings of neurological soft signs in high-risk subjects (i.e., relatives of schizophrenia patients and unaffected co-twins of monozygotic twin pairs discordant for schizophrenia (5–7)). These studies found that relatives take an intermediate position between healthy and schizophrenia subjects.

It is generally accepted that neurological soft signs are associated with psychopathology, especially negative symptoms and formal thought disorders (8–10). This notion is consistent with earlier studies of our group (10–12) in which we demonstrated a significant decrease of neurological soft signs that paralleled remission of symptoms under neuroleptic treatment.

The dysfunctional networks involved in the pathogenesis of neurological soft signs have not been fully identified to date. Neuroimaging studies have suggested associations of neurological soft signs with activation changes in the sensorimotor cortex and the supplementary motor area, cerebellar abnormalities, and subcortical findings involving the basal ganglia and thalamus (4, 10, 13, 14). Although these results and their interrelations have not yet been investigated in one joint study, they strengthen the hypothesis that neurological soft signs may refer to changes of frontal-thalamic-cerebellar pathways as conceptualized in the model of “cognitive dysmetria” (15).

Given these results, neurological soft signs can be interpreted as an expression of genetic liability toward the disease as have been presented in Meehl’s concept of “schizotaxia” (16) in which dysdiadochokinesia—among others—constitutes a trait-like marker of a baseline defect (“hypokrisia”). To further understand fluctuations in the level of neurological soft signs and in particular their increase with acute symptoms of schizophrenia, Huber’s hypothesis of “process activity” (17) may serve as a complementary concept. Huber differentiates reversible and irreversible symptoms, the former representing functional states of the active disease process that may remit with clinical stabilization. From his pneumencephalographic study (18) he concluded that once the active process persists, structural and psychopathological deteriorations
run a parallel course toward an irreversible deficit. Hence, process activity does not refer to a static concept but addresses the variability of the clinical course.

Drawing on the outlined concepts, we hypothesized that 1) neurological soft sign scores would increase during acute phases of the illness and return to baseline values with stabilization through treatment (possibly reflecting the activity of the disease process); 2) a decrease in neurological soft sign levels would be a favorable prognostic criterion; and 3) even in patients with a remitting course and favorable outcome, neurological soft signs would remain increased relative to healthy comparison subjects as an expression of genetic liability or schizotypia.

Method

Subjects

Thirty-nine schizophrenia patients whose first psychotic episode had remitted before discharge from an acute care ward were consecutively included in the study. Individuals with a history of or a concomitant neurological or medical disorder or severe substance abuse were excluded. The group consisted of 21 women and 18 men, all Caucasian, with a mean of 11.6 years (SD=1.6) of education; mean age at study intake was 27.0 years (SD=7.7). Patients were discharged on atypical antipsychotic regimens according to their psychiatrists’ choice (mean dose=579.5 mg in chlorpromazine equivalents [SD=267.3]). One year later, patients were invited for follow-up assessment; the exact interval amounted to a mean of 14.2 months (SD=1.6, range=10–18). Twenty-two healthy Caucasian comparison subjects were recruited from the general population through advertisements after screening for and exclusion of major psychiatric disorders. The group consisted of 10 men and 12 women, mean age=28.0 years (SD=3.8), mean education level=12.6 years (SD=0.9), and mean follow-up interval=10.0 months (SD=1.4, range=8–13). All ratings were performed by the same trained raters (S.B., C.B.) on both occasions. To maintain intra- and interrater reliability and avoid rater drift, we implemented an ongoing program of regular reliability checks. The study was approved by the ethics committee of the Medical Faculty, University of Heidelberg. After full explanation of proceedings subjects provided informed consent to participate.

Assessments

Neurological soft signs were assessed with the Heidelberg Scale (10), which consists of five subscales comprising 16 items (Appendix 1). All items except for gait, tandem gait, Ozeretzki’s test, articulation, and right/left orientation are assessed separately for both the right and the left side. Ratings are given on a 0–3-point scale (no/slight/moderate/marked abnormality, respectively). In the initial evaluation of the Heidelberg Scale (10), a sufficient internal reliability (Cronbach’s alpha=0.85/0.89 for schizophrenia patients/healthy subjects) and interrater reliability (r=0.88, p<0.005) were established.

The Edinburgh Inventory (19) was applied to establish handedness. By rating common activities, a quantitative inventory is calculated that ranges from −100 (strong left hander) to 100 (strong right hander). We used narrow definitions of left handedness (laterality quotient less than −80) and right handedness (laterality quotient more than 80); the remaining range represented mixed handedness.

Diagnoses were established with the German version of the Structured Clinical Interview for DSM-IV (20), which also served to exclude further major psychiatric disorders. Psychopathological symptoms were rated on the Positive and Negative Syndrome Scale (21), and predictors of outcome were rated on the Strauss-Carpenter Scale (22). Side effects of medication were documented with the Simpson-Angus Rating Scale (23), Barnes Rating Scale for Drug-Induced Akathisia (24), and the Abnormal Involuntary Movement Scale (AIMS) (25). At the follow-up evaluation, diagnoses, symptoms, predictors of outcome, and side effects of medication were reassessed with the indicated instruments. Moreover, information on compliance and important areas of functioning during the follow-up interval was gathered through a thorough clinical interview. Compliance was rated as present only if patients reported regular medication intake.

Data Analysis

For all variables, Kolmogorov-Smirnov analyses were calculated to test for normal distribution. Using chi-square tests, extended by Fisher’s exact tests if warranted and analyses of variance (ANOVA), possible differences between patients and healthy subjects and between the sexes with respect to clinical variables were assessed. Those variables that had been rated twice during the clinical course were analyzed with repeated measures ANOVAs. To further analyze the changes in neurological soft sign scores over time, the patient group was dichotomized by a median split on the basis of the respective change; resulting subgroups were compared with respect to the other clinical parameters. Moreover, variables that preceded neurological soft signs at the follow-up evaluation were entered in a stepwise linear regression analysis to detect possible predictors of neurological soft sign change. Extending previous studies we established the test-retest reliability of the Heidelberg Scale. This was done in the comparison subjects, since neurological soft sign stability was expected in this group. All analyses were performed with the Statistical Package for the Social Sciences.

Results

Demographic and Clinical Data

Patients and comparison subjects were comparable with respect to age, gender, and education. Although eight patients and only one healthy subject had a family history of psychiatric disease, this difference did not reach significance level.

Patients’ initial assessment revealed the following diagnoses: schizophrenia (N=20), schizoaffective disorder (N=2), schizophreniform disorder (N=16), and psychosis not otherwise specified (N=1). At the follow-up evaluation, a diagnostic shift was seen in 16 cases: from schizophreniform disorder to schizophrenia (N=14), from psychosis not otherwise specified to schizophrenia (N=1), and from schizophrenia to schizoaffective disorder (N=1). During the follow-up period, 33 individuals adhered to psychiatric treatment regularly. Thirty-one patients received continuous treatment with atypical antipsychotics (mean dose=297.7 mg/day in chlorpromazine equivalents [SD=35.5]); four of these patients were prescribed a mood stabilizer or an antidepressant in addition. Side effects of medication were low and unchanged at the follow-up evaluation relative to remission according to the AIMS (time 1: median=0, range=13; time 2: median=0, range=5), the Barnes Rating Scale for Drug-Induced Akathisia (time 1: median=0, range=4; time 2: median=0, range=3), and...
the Simpson-Angus Rating Scale (time 1: median=12, range=7; time 2: median=11, range=4).

At the follow-up evaluation, the overall Positive and Negative Syndrome Scale score (mean=52.4, SD=25.6) did not represent a significant difference from the remission score (mean=52.0, SD=12.4). The mean follow-up Strauss-Carpenter Scale score (58.4, SD=11.5) was comparable to the mean initial score (57.4, SD=11.5). During the follow-up period two patients had been continuously ill. Relapses occurred in 13 individuals, six of whom were fully recovered at the second interview. The remaining seven subjects still suffered from psychotic symptoms (N=4) or had been readmitted to the hospital (N=3). Five individuals used cannabis, four of these were also regular alcohol users. Thirty-three patients were compliant with treatment, and 31 were compliant with their medication. Ten individuals had continued their education successfully, and 20 were fully employed. Regular participation in household duties was reported by 29 subjects, regular social contacts by 24 subjects.

**Neurological Soft Signs and Handedness**

The patient group consisted of a similar number of right handers (N=20) and mixed handers (N=19). In the comparison group right handers (N=19) outnumbered mixed handers (N=3). While this difference between groups was statistically significant (F=8.5, df=1, 59, p<0.01), analyses did not yield a difference between the sexes within diagnostic groups.

Patients’ mean initial scores on the Heidelberg Scale decreased significantly during the follow-up period (Table 1), whereas the scores of the comparison subjects remained almost unchanged. Good test-retest reliability of the Heidelberg Scale (r=0.80, df=20, p<0.001) was seen in the healthy subjects. Neurological soft sign subscales were analyzed separately. Only for the motor coordination subscale were significant results detected for time, diagnosis, and the time-by-diagnosis interaction.

To further investigate the decrease of neurological soft signs in patients, this group was split according to the median change in neurological soft sign total scores. Subsequently, 21 patients with a pronounced neurological soft sign decrease were compared with 18 patients with stable or increasing neurological soft sign levels and to healthy subjects by means of a repeated measures ANOVA, which yielded a significant difference (Table 2). Post hoc Duncan tests revealed that both patient groups were comparable at first assessment whereas the group experiencing a decrease in neurological soft sign levels took an intermediate position on follow-up and did not differ statistically from healthy subjects (Figure 1).

Additional analyses of the three groups were performed for the neurological soft sign subscales (Table 2) where significant differences emerged for motor coordination, sensory integration, and complex motor tasks.

The differences between patients and healthy subjects were not related to handedness, gender, family history of psychiatric disease, or any other sociodemographic variable.

**Neurological Soft Sign Course and Clinical Measures**

Table 3 depicts sociodemographic and clinical data for patients with decreasing and stable neurological soft sign scores. No differences between groups emerged with respect to gender, social relationships, household duties, relapse, or alcohol/cannabis use. The subgroup with decreasing neurological soft sign levels showed better compliance and educational and vocational achievement. Time-by-group interactions revealed more favorable findings for patients with decreasing neurological soft sign levels as opposed to those with stable levels in terms of scores on the Strauss-Carpenter Scale and the Positive and Negative Syndrome Scale. In addition, for the Positive and Negative Syndrome Scale there were significant main effects of group for the total score (F=5.0, df=1, 37, p<0.05), positive symptom score (F=5.0, df=1, 37, p<0.05), and global psychopathology score (F=4.3, df=1, 37, p<0.05).

A regression analysis was calculated for neurological soft signs at the follow-up evaluation to identify predictors of neurological soft sign decrease. Variables that preceded the second assessment were entered. The analysis revealed neurological soft sign levels at remission and compliance with treatment during the follow-up period to be relevant influences (Table 4).
To our knowledge this is the first prospective longitudinal study to compare neurological soft signs in first-episode schizophrenia patients and healthy subjects. Whereas soft signs remained almost stable in healthy subjects, they significantly decreased in first-episode schizophrenia patients during a follow-up period of 14 months. This effect was related to better outcome. Despite the significant decrease, neurological soft signs remained elevated in patients relative to healthy subjects. Our findings support the initial hypotheses and contribute to the understanding of neurological soft signs in schizophrenia patients.

Neurological soft signs were present to a significantly greater extent in patients than in comparison subjects at both measurement points. The level of neurological soft signs at remission (time 1) is well within the range of remission scores in an earlier study by our group (10), and follow-up neurological soft sign levels of patients with decreasing scores closely corresponded to scores of unaffected co-twins of monozygotic twins discordant for schizophrenia (7). Our results clearly confirm the general finding of increased neurological soft sign scores in schizophrenia patients (1, 2, 10, 26), including first-episode cases, and accord with the view that neurological soft signs range among the most consistent neurobiological characteristics of schizophrenia (27).

FIGURE 1. Neurological Soft Sign Levels at Remission and 14-Month Follow-Up Evaluation in First-Episode Schizophrenia Patients Classified by Symptom Course and Age- and Gender-Matched Healthy Comparison Subjects

The patient group was dichotomized by a median split according to neurological soft sign score changes into those with decreasing (N=21) and those with stable neurological soft sign levels (N=18) (F=64.81, df=2, 58, p<0.001, repeated-measures analysis of variance).

Discussion

To our knowledge this is the first prospective longitudinal study to compare neurological soft signs in first-episode schizophrenia patients and healthy subjects. Whereas soft signs remained almost stable in healthy subjects, they significantly decreased in first-episode schizophrenia patients during a follow-up period of 14 months. This effect was related to better outcome. Despite the significant decrease, neurological soft signs remained elevated in patients relative to healthy subjects. Our findings support the initial hypotheses and contribute to the understanding of neurological soft signs in schizophrenia patients.

Neurological soft signs were present to a significantly greater extent in patients than in comparison subjects at both measurement points. The level of neurological soft signs at remission (time 1) is well within the range of remission scores in an earlier study by our group (10), and follow-up neurological soft sign levels of patients with decreasing scores closely corresponded to scores of unaffected co-twins of monozygotic twins discordant for schizophrenia (7). Our results clearly confirm the general finding of increased neurological soft sign scores in schizophrenia patients (1, 2, 10, 26), including first-episode cases, and accord with the view that neurological soft signs range among the most consistent neurobiological characteristics of schizophrenia (27).

During the follow-up period, neurological soft sign scores clearly decreased in patients but remained almost unchanged on a low level in healthy subjects. In particular, patients with decreasing neurological soft sign scores experienced further stabilization of symptoms and functioning, whereas clinical findings of patients with stable scores foreshadowed a chronic course. As suggested in previous studies (10, 28), this effect arose through disturbed motor and sensory integration signs rather than orientation difficulties or hard signs. Similar findings were obtained in earlier studies by our group, namely a parallel decrease of neurological soft signs and acute symptoms in patients with remitting schizophrenia as well as in first-episode patients under initial treatment with a typical neuroleptic (10–12). Further support of a positive correlation between improvement in clinical status and neurological performance stems from studies describing a significant improvement of neurological soft signs during a 6-month follow-up period (26), and more pronounced long-term deterioration in neuroleptic-free compared with medicated patients (28). Along these lines, cross-sec-

neurological soft signs with increased symptom levels, poor premorbid adjustment, and unfavorable outcome (1, 3, 10, 29) as well as with neurobiological measures such as neuropsychological deficits and structural and functional cerebral abnormalities (10–14). A correlation between neurological soft signs and negative symptoms was especially reported in drug-naive and medicated patients (2). Both neurological soft signs and negative symptoms might be a consequence of dopaminergic hypoactivity. This notion accords with the difference in neurological soft signs between medication responders and nonresponders (10, 12, 26) as well as with the finding that neurological soft signs are most prominent in chronic forms of schizophrenia (1). The aforementioned relationships were also confirmed by our study, since neurological soft signs were significantly related to the different symptom dimensions and predictors of outcome at both measurement points. Thus, our results hint at the necessity to differentiate between patient subgroups according to their symptoms and outcome.

Since neurological soft signs are present before medication exposure, it is generally accepted that they are an intrinsic feature of schizophrenia rather than a side effect of medication (2, 3). This view is supported by reports on spontaneous abnormal involuntary movements in schizophrenia patients, which had already been observed in the preneuroleptic era (30). Further support for a genetic determination stems from studies on relatives of schizophrenia patients (7, 8). Of interest in our study was that follow-up neurological soft sign scores of patients with a favorable outcome were in the range of unaffected co-twins of monozygotic twin pairs discordant for schizophrenia (7). This is in line with evidence from studies on neuroleptic-naive patients (3, 12, 27) and suggests that a remitting disease course leads to an increase in neurological soft sign levels during a limited period of time (i.e., an acute psychotic exacerbation) and then a return to the genetically

| TABLE 3. Demographic and Clinical Characteristics of First-Episode Schizophrenia Patients by Change From Baseline in Neurological Soft Sign Levela |
|-----------------------------------------------|-------------------------------|-------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Characteristic                              | Stable Neurological Soft Sign Levels (N=18) | Decreasing Neurological Soft Sign Levels (N=21) | Analysis                | Time 1 | Time 2 | Time 1 | Time 2 | Analysis |
|                                              | N | % | N | % | N | % | p       | N | % | N | % | p       |
| Relapse                                      |    |   |    |   |    |   | 0.23    |    |   |    |   | 0.07    |
| Yes/continuously ill                         | 9 | 50.0 | 7 | 33.3 |    |   |    | 0.05  |    |   |    | 0.07  |
| No                                           | 9 | 50.0 | 14 | 66.7 |    |   |    | 0.07  |    |   |    | 0.07  |
| Treatment compliance                         |    |   |    |   |    |   | 0.07    |    |   |    |   | <0.05    |
| Yes                                          | 13 | 72.2 | 20 | 95.2 |    |   |    | <0.05  |    |   |    | <0.05  |
| No                                           | 5 | 27.8 | 1 | 4.8 |    |   |    | <0.05  |    |   |    | <0.05  |
| Medication compliance                        |    |   |    |   |    |   | 0.07    |    |   |    |   | <0.05    |
| Yes                                          | 11 | 61.1 | 20 | 95.2 |    |   |    | <0.05  |    |   |    | <0.05  |
| No                                           | 7 | 38.9 | 1 | 4.8 |    |   |    | <0.05  |    |   |    | <0.05  |
| Alcohol use                                  |    |   |    |   |    |   | 0.64    |    |   |    |   | 0.14    |
| Regular/irregular                            | 16 | 88.9 | 19 | 90.5 |    |   |    | 0.14  |    |   |    | 0.14  |
| No                                           | 2 | 11.1 | 2 | 9.5 |    |   |    | 0.14  |    |   |    | 0.14  |
| Cannabis use                                 |    |   |    |   |    |   | 0.058   |    |   |    |   | <0.05    |
| Regular/irregular                            | 13 | 72.2 | 19 | 90.5 |    |   |    | <0.05  |    |   |    | <0.05  |
| No                                           | 5 | 27.8 | 2 | 9.5 |    |   |    | <0.05  |    |   |    | <0.05  |
| Educational achievement                      |    |   |    |   |    |   | 0.058   |    |   |    |   | <0.05    |
| Yes                                          | 2 | 11.1 | 8 | 38.1 |    |   |    | <0.05  |    |   |    | <0.05  |
| No                                           | 16 | 88.9 | 13 | 61.9 |    |   |    | <0.05  |    |   |    | <0.05  |
| Employment                                   |    |   |    |   |    |   | <0.05   |    |   |    |   | <0.05    |
| Full-time                                    | 6 | 33.3 | 14 | 66.7 |    |   |    | <0.05  |    |   |    | <0.05  |
| Part-time/none                               | 12 | 66.7 | 7 | 33.3 |    |   |    | <0.05  |    |   |    | <0.05  |
| Household duties                             |    |   |    |   |    |   | 0.26    |    |   |    |   | 0.61    |
| Yes                                          | 12 | 66.7 | 17 | 81.0 |    |   |    | 0.61  |    |   |    | 0.61  |
| No                                           | 6 | 33.3 | 4 | 19.0 |    |   |    | 0.61  |    |   |    | 0.61  |
| Social relationships                         |    |   |    |   |    |   | 0.61    |    |   |    |   | 0.61    |
| Yes                                          | 11 | 61.1 | 13 | 61.9 |    |   |    | 0.61  |    |   |    | 0.61  |
| No                                           | 7 | 38.9 | 8 | 38.1 |    |   |    | 0.61  |    |   |    | 0.61  |

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>F&lt;</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strauss-Carpenter Scale (points)</td>
<td>58.3</td>
<td>11.7</td>
<td>53.9</td>
<td>12.6</td>
<td>56.6</td>
<td>11.6</td>
<td>62.2</td>
<td>9.2</td>
<td>9.8</td>
</tr>
<tr>
<td>Positive and Negative Syndrome Scale scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive symptoms</td>
<td>10.3</td>
<td>4.9</td>
<td>14.8</td>
<td>10.6</td>
<td>9.5</td>
<td>2.2</td>
<td>8.7</td>
<td>2.3</td>
<td>6.4</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>14.7</td>
<td>4.8</td>
<td>16.8</td>
<td>7.5</td>
<td>14.8</td>
<td>4.3</td>
<td>12.0</td>
<td>4.7</td>
<td>5.7</td>
</tr>
<tr>
<td>Global psychopathology</td>
<td>27.6</td>
<td>8.6</td>
<td>32.3</td>
<td>16.4</td>
<td>27.1</td>
<td>5.4</td>
<td>21.8</td>
<td>5.6</td>
<td>9.0</td>
</tr>
<tr>
<td>Total</td>
<td>52.6</td>
<td>14.9</td>
<td>63.9</td>
<td>32.6</td>
<td>51.4</td>
<td>10.0</td>
<td>42.6</td>
<td>10.8</td>
<td>9.1</td>
</tr>
</tbody>
</table>

a There were eight men and 10 women in the stable group and 10 men and 11 women in the decreasing group, a nonsignificant difference; the mean follow-up period was 14.2 months (SD=1.6).
b Four-factor Fisher's exact test.
c Time-by-group interaction (repeated measures analysis of variance [df=1, 37]).
determined baseline thereafter. These results are consistent with Meehl’s model of “schizotaxia” (16) in which dysdiadochokinesia—i.e., a sign of dysfunctional motor coordination—is conceptualized as a marker of the baseline defect (“hypokrisia”). On the other hand, the persisting elevation of neurological soft sign scores in chronic cases may hint at a still enduring process activity reflecting premorbid changes of neurodevelopmental origin (31) or the presence of acquired, irreversible deficits.

Thus, neurological soft signs in schizophrenia seem to adopt characteristics of both state-like and trait-like features. During the early course of acute psychosis when symptoms fluctuate, the state-like features of an active disease process (17, 18) may be predominant. On the other hand, the trait-like features that represent the genetically determined baseline may prevail after remission of the acute illness. In conclusion, our results and the body of literature allow for the notion that neurological soft signs represent surrogate markers of the schizophrenic disease process, i.e., the process activity that is more prominent and fluctuating during the early course of the disease; in later phases of the disease the process may come to a standstill, but it may also completely resolve or deteriorate into deficit states (17, 18).

This study may be limited by a recruitment bias. However, patients initially were included in the study consecutively as they necessitated hospital treatment. Neuroleptic medication was not standardized but restricted to atypical compounds, chosen according to the patients’ individual needs, and extrapyramidal side effects were rare and not associated with neurological soft signs. Most important, compliance with medication was a positive predictor of neurological soft sign decrease. In general, the comparability among studies is limited because of the absence of a universally accepted structured instrument. However, most instruments in use comprise a set of similar subtests, e.g., tandem gait, Romberg, diadochokinesis, finger-nose tapping, finger-thumb opposition, fist-edge palm test, Ozeretski’s test, mirror movements, graphesthesia, stereognosis, right/left orientation (32). Therefore, overall data comparability among studies is relatively high. In spite of this methodological limitation, the evidence for a higher rate of neurological abnormalities in schizophrenia is consistent and compelling. The test-retest reliability of our Heidelberg Scale over a longer period of time was sufficient.

In summary, neurological soft signs are intrinsic to schizophrenia but their measured quantity and magnitude may serve as a surrogate marker for the activity of the disease process as well as a predictor of outcome. Overall, the assessment of neurological soft signs represents a hardly time-consuming, inexpensive, and meaningful tool in clinical psychiatry and has the potential to bridge the gulf between neurobiological research and clinical practice.

APPENDIX 1. Tests Comprising the Neurological Soft Sign Subscales of the Heidelberg Scale

<table>
<thead>
<tr>
<th>Heidelberg Scale Subscale and Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. Motor coordination</td>
</tr>
<tr>
<td>Ozeretski’s test</td>
</tr>
<tr>
<td>Diadochokinesis</td>
</tr>
<tr>
<td>Pronation/supination</td>
</tr>
<tr>
<td>Finger/thumb opposition</td>
</tr>
<tr>
<td>Articulation</td>
</tr>
<tr>
<td>2. Sensory integration</td>
</tr>
<tr>
<td>Gait</td>
</tr>
<tr>
<td>Tandem gait</td>
</tr>
<tr>
<td>2-point discrimination</td>
</tr>
<tr>
<td>3. Complex motor tasks</td>
</tr>
<tr>
<td>Finger-to-nose test</td>
</tr>
<tr>
<td>Fist-edge-palm test</td>
</tr>
<tr>
<td>4. Right/left and spatial orientation</td>
</tr>
<tr>
<td>Right/left orientation</td>
</tr>
<tr>
<td>Graphesthesia</td>
</tr>
<tr>
<td>Face/hand sensory test</td>
</tr>
<tr>
<td>Stereognosis</td>
</tr>
<tr>
<td>5. Hard signs</td>
</tr>
<tr>
<td>Arm-holding test</td>
</tr>
<tr>
<td>Mirror movements</td>
</tr>
</tbody>
</table>

APPENDIX 2. Model Summary

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model R²</th>
<th>Adjusted R²</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Beta Weight</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological soft sign level at remission</td>
<td>0.55</td>
<td>0.53</td>
<td>22.38</td>
<td>36</td>
<td>&lt;0.001</td>
<td>0.67</td>
<td>5.51</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Treatment compliance</td>
<td>0.67</td>
<td>0.64</td>
<td>16.28</td>
<td>36</td>
<td>&lt;0.001</td>
<td>–0.32</td>
<td>–2.89</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

a Excluded: family history of psychiatric disease, side effects at remission according to the Barnes Rating Scale for Drug-Induced Akathisia, side effects at follow-up evaluation according to the Simpson-Angus Rating Scale, and the following variables over the follow-up interval: alcohol and drug use, educational achievement, social and vocational functioning, and participation in household duties.

References

5. Lawrie, Byrne, Miller, Hodges, Clafferty, Cunningham, Owens, Johnstone: Neuropsychological indices and the development of psychotic symptoms in subjects at high risk of schizophrenia. Br J Psychiatry 2001; 178:524-530
7. Niethammer, Weisbrod, Schiesser, Grothe, Pepeu, Maier, Pietschmann, Schiesser, Schiesser: Motor dysfunction and sensorimotor cortex activation changes in schizophrenia: a study with functional magnetic resonance imaging. Neuroimage 1999; 9:81-87
22. Strauss, Carpenter: The prediction of outcome in schizophrenia, II: relationship between predictors and outcome variables. Arch Gen Psychiatry 1974; 31:37-42
30. Kraepelin: Psychiatrie. Leipzig, Germany, Barth, 1913, p 751
Caffeine Dependence in Combination With a Family History of Alcoholism as a Predictor of Continued Use of Caffeine During Pregnancy

Dace S. Svikis, Ph.D.
Nathan Berger, M.D.
Nancy A. Haug, Ph.D.
Roland R. Griffiths, Ph.D.

Objective: The purpose of the study was to examine whether caffeine dependence and a family history of alcoholism are associated with continued use of caffeine during pregnancy.

Method: Forty-four women seeking obstetrical care in an office-based practice completed questionnaires and provided saliva samples at three prenatal visits occurring 2–3, 3–4, and 7 months post-conception. On visit 1, the patients received the physician's instructions to stop using caffeine. Structured interviews were used to assign a diagnosis of caffeine dependence (lifetime) and to identify family history of alcoholism. Outcome measures included self-reported levels of caffeine use and saliva caffeine levels at the three prenatal visits.

Results: Although most women eliminated or substantially reduced their caffeine consumption between pregnancy awareness and prenatal visit 1, those with a lifetime diagnosis of caffeine dependence and a family history of alcoholism had higher levels of caffeine use and lower rates of abstinence throughout pregnancy. Saliva caffeine levels confirmed these effects. Withdrawal symptoms, functional impairment, and craving were cited as reasons they failed to eliminate or cut back on caffeine use. Fifty percent of the women with both a lifetime diagnosis of caffeine dependence and a family history of alcoholism continued to use caffeine in amounts (>300 mg/day) greater than those considered safe during pregnancy, compared to none of the women without caffeine dependence and a family history of alcoholism. Women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism also reported higher rates of past cigarette smoking and problematic alcohol use.

Conclusions: Caffeine-dependent women with a family history of alcoholism were not able to follow their physician's advice to reduce or eliminate caffeine consumption during pregnancy, despite their wanting to do so. This subgroup may require more intensive intervention to ensure caffeine abstinence and may be at greater risk for abuse of or dependence on other drugs.

Caffeine is the most widely used mood-altering drug in the world, with 80% to 90% of children and adults in North America regularly consuming caffeine-containing foods (1, 2). Mean daily caffeine consumption among adult caffeine consumers in the United States has been estimated to be 280 mg/day, which is equivalent to about three 6-oz cups of coffee or five 16-oz bottles of cola soft drink (3, 4).

Several case report studies documented that some people fulfill the DSM-IV diagnostic criteria for substance dependence applied to caffeine use (5–7). In a random-digit telephone survey of 162 caffeine users, 30% fulfilled the diagnostic criteria for caffeine dependence, with 56% of caffeine users reporting a persistent desire or unsuccessful efforts to cut down or control caffeine use and 14% reporting continued use despite knowledge of a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by caffeine (8). Although the diagnosis of caffeine dependence is not included in DSM-IV, the validity of the diagnosis is suggested by two studies that prospectively demonstrated that the incidence of withdrawal (9) and the severity of withdrawal (5) were greater in individuals who met the criteria for a caffeine dependence diagnosis, based on the generic criteria for DSM-IV substance dependence.

It has been well established that family members of persons with alcoholism are more likely to be alcohol dependent, and adoption and twin studies have suggested a role for genetic factors in the familial transmission of alcohol and drug dependence (10–14). Twin studies have also demonstrated the heritability of problematic caffeine use, including heavy caffeine use, caffeine tolerance, and caffeine withdrawal (15, 16). With regard to a link between alcoholism and caffeine, there is a high co-occurrence between alcoholism and caffeine use (5, 17, 18), and twin studies examining alcohol use, caffeine use, and cigarette smoking concluded that a common genetic factor (poly-
substance use) underlies the use of these three substances, with 28% to 41% of the heritable effects of caffeine use (or heavy use) shared with those of alcohol use and smoking (15, 19, 20).

Caffeine use during pregnancy has been associated with a variety of adverse consequences, including decreased fecundity, spontaneous abortion, and reduced fetal growth (21–24). Although methodological limitations preclude definitive scientific conclusions (25), governmental health agencies of the United States, Canada, and the United Kingdom have issued health warnings about limiting the use of caffeine during pregnancy. The U.S. Food and Drug Administration has advised pregnant women to “avoid caffeine-containing foods and drugs, if possible, or consume them only sparingly” (26), and Health Canada (27) and the Food Standards Agency of the United Kingdom (28) have advised that pregnant women consume ≤300 mg/day of caffeine. Although obstetricians frequently counsel women to reduce or eliminate caffeine use during pregnancy (29), women often continue to consume caffeine throughout pregnancy (30–31).

The present study examined caffeine use in women who received advice from their obstetrician to abstain from caffeine use during pregnancy. The study sought to identify factors associated with a woman’s ability to abstain from or reduce prenatal caffeine use. The factors were a history of caffeine dependence (i.e., a lifetime diagnosis of caffeine dependence, based on the DSM-IV diagnostic criteria for substance dependence applied to caffeine use) and a family history of alcoholism. Questionnaire assessments occurred at three prenatal visits 2–3, 3–4, and 7 months after conception, and a personal interview occurred between the second and third questionnaire assessments. Because caffeine use and alcohol use are often associated and because both are known to occur more frequently in some families, it was hypothesized that women with caffeine dependence and a family history of alcoholism would have more difficulty eliminating or reducing caffeine use during pregnancy.

Method

Recruitment

Study participants were drawn from a group of pregnant women seeking prenatal care from a private obstetrics and gynecology practice in a suburban community. Independent of their current or past caffeine use, the women, who were waiting to see their practitioner at their first clinic visit after learning they were pregnant (generally about 8–12 weeks after conception), were invited to participate in a three-session questionnaire study of caffeine use in pregnancy. Of 109 women offered study participation, only nine (8%) refused. Those providing informed consent (N=100) completed a 20-minute questionnaire at this first prenatal visit. Of these, 87 (84%) completed the first follow-up questionnaire at prenatal visit 2 (occurring about 12–16 weeks after conception) and 84 (84%) completed the second follow-up questionnaire at prenatal visit 6 (occurring about 28 weeks after conception). Reasons for dropout or study discontinuation included miscarriage (N=8), change in obstetrician (N=2), decision to discontinue study participation (N=2), and unknown reasons (N=4). Data from an additional subject were eliminated from the data set because she provided contradictory responses. Thus, 83 women provided informed consent and completed all three questionnaires. When these women were offered the opportunity to participate in a personal interview, 65 (78%) consented to the interview and 50 (77%) completed the interview. Of these women, five (10%) reported no prepregnancy caffeine use, and their data were eliminated from subsequent analyses. Data from one additional participant were subsequently eliminated because she miscarried. All interview participants received $50 as compensation for their time and effort.

Participants

The final study group consisted of 44 women who provided informed consent to a personal interview about caffeine use and other health behaviors and reported some caffeine use in the 3 months before becoming aware of pregnancy. Subjects had a mean age of 31.9 years (SD=4.0); 96% (N=42) were Caucasian; 100% (N=44) were married; and 50% (N=22) had a 4-year college degree, with an additional 23% (N=10) reporting an advanced degree. Nearly three-fourths (73%, N=32) were employed full-time (>35 hours/week), and 23% (N=10) were employed part-time (≤35 hours/week). More than one-half of the women (57%, N=25) met the criteria for a lifetime diagnosis of caffeine dependence (i.e., met the DSM-IV criteria for lifetime substance dependence applied to caffeine use). Eighteen percent (N=8) met the DSM-III-R criteria for lifetime alcohol abuse (2%, N=1) or dependence (16%, N=7); however, none had a current alcohol use diagnosis, and none had been treated for alcohol problems. Seven percent (N=3) of the women met the DSM-III-R criteria for lifetime substance abuse or dependence applied to drugs other than nicotine or caffeine. All of these cases involved past (not current) cannabis use. About one-quarter (27%, N=12) of the women reported a history of daily cigarette smoking, but only three (7%) continued to smoke during their pregnancy.

More than one-half (52%, N=23) of the women reported at least one first-degree relative who met the Family History Research Diagnostic Criteria (FH-RDC) for alcoholism (32). Specifically, 32% (N=14) of the women reported alcoholism in their biological father, 14% (N=6) in their biological mother, 11% (N=5) in both parents, and 39% (N=17) in at least one sibling. More than one-half (55%, N=24) of the subjects met the DSM-III-R criteria for a lifetime diagnosis of either a mood disorder and/or an anxiety disorder (major depression: 41% [N=18] lifetime, 2% [N=1] current; panic, obsessive-compulsive, and/or generalized anxiety disorders: 20% [N=9] lifetime, 9% [N=4] current).

Questionnaire Data and Saliva Samples

The initial questionnaire, which was completed at the first prenatal visit, assessed caffeine, tobacco, alcohol, and other drug use during the 6 months before pregnancy awareness and during the 7 days before the first prenatal visit. The two follow-up questionnaires assessed changes in caffeine and other substance use since the last visit and specifically for the 7 days before the second and sixth prenatal visits. For caffeine assessments, participants provided the number of days during the last week that they consumed caffeine-containing products as well as the number of servings, serving size, and brand of the caffeinated products they consumed. Previous studies suggested the reliability and validity of self-reports of caffeine use (33, 34), although pregnant women may underreport drug use to health care workers (35). The questionnaires were given to the subjects by nurses, sealed in envelopes after completion, and forwarded to research staff for data entry.

To validate the questionnaire data about caffeine consumption, participants provided a 5-ml saliva sample at the first, sec-
to examine differences in caffeine consumption at each timepoint in women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all other groups combined.

Another analysis was conducted to determine whether the differences between the women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all other groups combined could be attributed to other disorders. The analysis used a two-way analysis of covariance (ANCOVA) with group (women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all other groups combined) and time as factors and with the following covariates: lifetime history of daily cigarette smoking, DSM-III-R alcohol abuse or dependence, DSM-III-R drug abuse or dependence, and DSM-III-R diagnosis of any other psychiatric diagnosis assessed by the SCID. This model used a type I estimable function to allow the covariates to be entered first. Follow-up planned comparisons (t tests) with the adjusted means were used to examine differences in caffeine consumption at each timepoint in women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all other groups combined.

To assess the validity of self-reported caffeine consumption, Pearson's correlation coefficients were calculated for the relationship between reported amounts of caffeine consumed (mg/week) and caffeine saliva levels (mg/ml) at each of the three prenatal visits. At test was used to compare caffeine levels on occasions when subjects reported caffeine abstinence versus occasions when subjects reported caffeine use. Finally, in an analysis similar to that of the caffeine consumption data, caffeine saliva levels were examined for the four subgroups of women based on the presence or absence of caffeine dependence and a family history of alcoholism. The analysis used a two-way mixed ANOVA with subgroup and time (prenatal visits 1, 2, and 6) as factors. Planned comparisons (t tests) were used to examine differences in caffeine saliva levels at each timepoint in women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all other groups combined.

Categorical data (e.g., rates of abstinence, history of daily cigarette smoking, lifetime DSM-III-R alcohol abuse or dependence, DSM-III-R drug abuse or dependence) were compared by using Pearson's chi square, with Fisher's exact two-tailed probability levels used when the number of observations per cell was five or less.

For all statistical tests, effects were considered significant at p<0.05.

Results

Caffeine Use Before Pregnancy

In response to the questionnaire about caffeine use before pregnancy awareness, 41% (N=18) of the women indicated that they would find it difficult to be without caffeine. Most (93%, N=41) had experience abstaining from caffeine for a day or more, and 63% (N=26) of those who had abstained endorsed one or more of the following withdrawal symptoms: headache (49%, N=20), fatigue (39%, N=16), anxiety (10%, N=4), and nausea or vomiting (2%, N=1). Fifty-two percent (N=23) of the total study group reported using caffeine to avoid withdrawal symptoms, and 21% (N=9) had previously been told by a health care professional that they should cut back or eliminate their caffeine use because of its association with or effect on various medical conditions, including fibrocystic breast disease, headaches, pregnancy, insomnia, and stomach problems. Twenty-six percent (N=11) reported
that they believed they had physical problems that had been caused or made worse by caffeine.

**Diagnostic Interview**

More than one-half of the subjects (57%, N=25) fulfilled the DSM-IV criteria for lifetime substance dependence applied to caffeine use by endorsing three or more of the seven diagnostic criteria. The proportion of the study group that endorsed each of the seven DSM-IV criteria were: 1) tolerance as defined by a need for markedly increased amounts of the substance to achieve desired effect, or markedly diminished effect with continued use of same amount of substance (50%, N=22); 2) characteristic withdrawal syndrome or use of the substance to relieve or avoid withdrawal symptoms (77%, N=34); 3) substance taken in larger amounts or over a longer period than was intended (45%, N=20); 4) persistent desire or unsuccessful efforts to cut down or control substance use (45%, N=20); 5) a great deal of time spent in activities necessary to obtain the substance (25%, 11); 6) important social, occupational, or recreational activities given up or reduced because of substance use (0%); and 7) continued use despite knowledge of a persistent or recurrent physical or psychological problem likely to have been caused or exacerbated by the substance (43%, N=19).

**Difficulties Complying With Instructions to Stop Caffeine Use**

After receiving verbal and written instructions during the first prenatal visit to eliminate all caffeine during pregnancy, almost all of the women (98%, N=43) reported on prenatal visits 2 and 6 that they attempted to completely eliminate or cut back their caffeine use during pregnancy. More than one-half of these women (54%, N=23) reported one or more symptoms that they attributed to quitting or cutting back caffeine use, with 34% (N=15) reporting fatigue, 32% (N=14) sleepiness, 27% (N=12) headaches, 27% (N=12) craving for caffeine, 21% (N=9) yawning, 18% (N=8) nausea, 16% (N=7) less motivation for work, and 7% (N=3) vomiting. Fourteen percent (N=6) of the total study group (or 26% of those reporting withdrawal symptoms) reported that their caffeine withdrawal symptoms were severe enough to interfere with their performance at work, school, or home. Headache did not consistently accompany such functional impairment. Examples of functional impairment provided in written comments to open-ended questions included: “flu symptoms prevented going to work”; “less active at work”; “When I cut back, I am fatigued; it is difficult to concentrate.” Examples of reasons that women provided for failing to eliminate or cut back on their caffeine use included “cravings/headache/nervous,” “migraines,” “need to stay awake/headaches,” “severe withdrawal,” “couldn't concentrate at work,” and “need coffee to wake up.”

**Amount of Caffeine Use in Pregnancy**

Data on caffeine consumption in women with and without a lifetime diagnosis of caffeine dependence and a family history of alcoholism are summarized in Figure 1. The ANOVA results suggested effects on caffeine consumption of caffeine dependence (F=6.94, df=1, 40, p<0.01), family history of alcoholism (F=7.69, df=1, 40, p<0.01), time (F=28.32, df=3, 120, p<0.001), and the interaction of caffeine dependence and a family history of alcoholism (F=6.17, df=1, 40, p<0.05). Caffeine consumption (mg/week) was significantly higher in the 6 months before pregnancy awareness (mean=1671 mg/week, SD=1197) than it was at...
CAFFEINE USE DURING PREGNANCY

FIGURE 2. Caffeine Consumption Before and During Pregnancy in Women With and Without Lifetime Caffeine Dependence and a Family History of Alcoholism

| Consumption in the subgroup of women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism was higher than in the other three groups at all timepoints. |
| Significant differences between the subgroup with a lifetime diagnosis of caffeine dependence and a family history of alcoholism and all other subgroups combined were found for the period before pregnancy and at 2–3 and 7 months during pregnancy (p<0.05, planned-comparison t tests). |

Caffeine consumption decreased to 571 mg/week (SD=825) at the time of the first prenatal visit and then decreased further at the second visit, which followed the visit during which subjects received the physician’s advice to eliminate caffeine use (mean=366 mg/week, SD=682). At the sixth prenatal visit (approximately 4 months later), average caffeine consumption was 519 mg/week (SD=838).

At all timepoints caffeine-dependent women consumed more caffeine than those without caffeine dependence; these differences were significant at the first and sixth prenatal visits (Figure 1, left panel). Similarly, women with a family history of alcoholism consumed more caffeine than those without a family history of alcoholism; these differences were significant before pregnancy and at the first and sixth prenatal visits (Figure 1, right panel).

Patterns of caffeine use before pregnancy awareness through the sixth prenatal visit for the four subgroups are summarized in Figure 2. There were significant group (F=8.25, df=3, 40, p<0.001) and time (F=28.32, df=3, 120, p<0.001) effects. The planned t tests showed that women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism differed in caffeine consumption from the other three groups combined before pregnancy and at the first and sixth prenatal visits. Across the three prenatal visits, women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism consumed 3.4–5.4 times more caffeine than the other subjects. The ANCOVA confirmed these results after adjustment for lifetime history of daily cigarette smoking, DSM-III-R alcohol abuse or dependence, DSM-III-R drug abuse or dependence, and presence of any other DSM-III-R psychiatric diagnosis. The planned t tests with the adjusted means were significant for the period before pregnancy and for the first and sixth prenatal visits.

Inspection of the distribution of caffeine consumption across participants revealed that 50% of the women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism reported consuming more than 2100 mg/week of caffeine (or >300 mg/day) and 43% reported consuming more than 2800 mg/week (or >400 mg/day) at one or more of their prenatal visits. In contrast, no women from any of the other three subgroups reported consuming more than 2100 mg/week of caffeine on any occasion. The difference in the proportion of women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism versus all others with consumption of more than 2100 mg/week during 1 or more weeks was significant (χ²=17.84, df=1, Fisher’s exact p<0.001).

Caffeine Abstinence in Pregnancy

More than one-third (39%, N=17) of the women reported abstinence (i.e., no self-reported use) from caffeine during the week before the first prenatal visit. This abstinence rate increased to 50% (N=22) in the week before the second prenatal visit, with an intermediate rate of 41% (N=17) at the sixth prenatal visit. Almost one-half of the women (48%, N=21) reported consuming some caffeine during the week before all three prenatal visits, and 30% (N=13) reported caffeine abstinence at all three visits.

The rates of abstinence (i.e., no self-reported use during the assessed week) were lower for women with caffeine dependence, compared to women without caffeine dependence, at the first (32% versus 47%), second (44% versus 58%), and sixth (33% versus 46%) prenatal visits, but these differences only approached significance. The rates of abstinence in women with a family history of alcoholism were about one-half of those of women without a family history of alcoholism at the first (26% versus 52%) (χ²=3.20, df=1, p=0.07), second (35% versus 67%) (χ²=4.46, df=1, p=0.04), and sixth (24% versus 57%) (χ²=4.84, df=1, p=0.03) prenatal visits. Concordant with the analysis of amounts of caffeine, the subgroup of women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism had lower rates of abstinence than the other women at the first (14% versus 50%) (χ²=5.14, df=1, Fisher’s exact p=0.04), second (21% versus 63%) (χ²=6.71, df=1, Fisher’s exact p=0.02), and sixth (23% versus 48%)
Relationship to Histories of Other Substance Use

Women who had a caffeine dependence diagnosis were almost nine times more likely to report a period sometime before pregnancy during which they smoked cigarettes daily, compared to women without the diagnosis (44% versus 5%) ($\chi^2=8.17$, df=1, Fisher's exact $p=0.006$). Although a family history of alcoholism alone was not related to daily smoking, women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism were more than six times more likely to report a history of daily smoking than were the other women (64% versus 10%) ($\chi^2=14.18$, df=1, Fisher's exact $p=0.001$), and all three subjects who continued to smoke during pregnancy were in the group with a lifetime diagnosis of caffeine dependence and a family history of alcoholism.

Although presence of a caffeine dependence diagnosis was not related to having a DSM-III-R diagnosis of lifetime alcohol abuse or dependence, women with a family history of alcoholism were more likely than those without a family history of alcoholism to have a diagnosis of lifetime alcohol abuse or dependence (30% versus 5%) ($\chi^2=4.86$, df=1, Fisher's exact $p=0.05$). Women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism were more than six times more likely to have a DSM-III-R diagnosis of lifetime alcohol abuse or dependence, compared to the other women (43% versus 7%) ($\chi^2=8.40$, df=1, Fisher's exact $p=0.008$).

Only four subjects had a DSM-III-R diagnosis of lifetime drug abuse or dependence, all involving cannabis. Neither caffeine dependence nor family history of alcoholism was related to lifetime drug abuse or dependence. However, women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism were seven times more likely than the other women to have a DSM-III-R diagnosis of lifetime drug abuse or dependence (21% versus 3%), although the difference was not statistically significant ($\chi^2=3.78$, df=1, Fisher's exact $p=0.09$).

Caffeine Levels in Saliva

To determine the validity of the self-reported amounts of caffeine consumption, subjects provided saliva samples at each of the three clinic visits. Correlations between the amount of caffeine consumed (mg/week) and the caffeine saliva level (ng/ml) at each of the three timepoints ranged between 0.48 and 0.54 and were significant. Furthermore, the mean caffeine level on occasions when the subjects reported that they abstained from caffeine (mean=124 ng/ml, SD=192) was consistent with levels reported to be associated with abstinence in previous studies (40, 41) and was significantly lower ($p<0.001$, t test) than the mean level on occasions when the subjects reported that they were not abstinent (mean=647 ng/ml, SD=936).

Corresponding with the analysis of differences in caffeine intake shown in Figure 2, the ANOVA of data on caffeine saliva levels showed significant effects of group ($F=4.00$, df=3, 40, $p<0.05$) and time ($F=5.42$, df=2, 80, $p<0.01$). The women with a lifetime diagnosis of caffeine dependence and a family history of alcoholism had a higher mean caffeine saliva level than all other subjects at the first (mean=464 ng/ml, SD=535, versus mean=107 ng/ml, SD=165), second (mean=860 ng/ml, SD=1212, versus mean=232 ng/ml, SD=310), and sixth (mean=1255 ng/ml, SD=1539, versus mean=330 ng/ml, SD=301) prenatal visits. These differences were significant at the first ($t=-2.44$, df=42, $p<0.05$) and sixth ($t=-2.23$, df=42, $p<0.05$) prenatal visits.

Discussion

A primary finding of the present study was that women with both a lifetime diagnosis of caffeine dependence and a family history of alcoholism maintained higher levels of caffeine use throughout pregnancy and had lower rates of caffeine abstinence than women without these two characteristics. These relationships were demonstrated with data on self-reported caffeine consumption and by measurement of caffeine levels in saliva samples. The observation that the combination of a history of caffeine dependence and a family history of alcoholism is necessary for the effect suggests that a genetic vulnerability reflected in the family history of alcoholism may be necessary to express the problematic features of caffeine dependence. The suggestion of a genetic component to problematic caffeine use is consistent with previous studies showing a greater co-occurrence of heavy caffeine use, caffeine tolerance, and caffeine withdrawal in monozygotic than in dizygotic twin pairs, with heritabilities of these characteristics between 35% and 77% (15, 16).

Another intriguing finding is that women with both caffeine dependence and a family history of alcoholism were more likely to have a history of daily cigarette smoking as well as a lifetime diagnosis of alcohol or drug abuse or dependence. These observations extend previous findings that suggested a high co-occurrence between caffeine use, alcoholism, and cigarette smoking (42); between caffeine use and alcoholism (5, 17, 18); and between caffeine use and cigarette smoking (17). Furthermore, twin studies indicated that a common genetic factor (polysubstance use) underlies the use of caffeine, alcohol, and tobacco (15, 19, 20).

The present study provides valuable new information about a DSM diagnosis of substance dependence applied to caffeine. Three previous studies described a series of case reports of caffeine dependence in adults and adolescents (5–7), but they provided no meaningful information on the prevalence of the disorder. The present study showed that more than one-half (57%) of a group of employed and highly educated caffeine-using pregnant women fulfilled the DSM-IV criteria for lifetime substance dependence ap-
plied to caffeine, with 45% of the group reporting that they had persistent desire or unsuccessful efforts to cut down or control caffeine use and 43% reporting continued use despite knowledge of a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance. The only previous study of the prevalence of caffeine dependence was a random-digit telephone survey conducted in 162 caffeine users in Vermont (8). In that study, 30% of the sample fulfilled the DSM-IV substance dependence criteria applied to caffeine, 56% endorsed a persistent desire to cut down or quit, and 14% endorsed continued use despite problems.

The present study helps validate the diagnosis of caffeine dependence as a clinically significant phenomenon. Women in this study were instructed by their physician to quit caffeine use during pregnancy, and 98% reported attempting to quit or reduce use. Those who reported failing to eliminate or cut back on caffeine use (predominately women with caffeine dependence and a family history of alcoholism) cited withdrawal symptoms, functional impairment, and craving as reasons they continued to use the substance. Furthermore, 50% of the women who had both caffeine dependence and a family history of alcoholism reported continuing to use caffeine in amounts (>300 mg/day) greater than those thought to be safe during pregnancy (22, 27, 28).

The present study also contributes new information about the caffeine withdrawal syndrome. Despite the fact that more than 66 experimental and survey studies have characterized various aspects of the caffeine withdrawal syndrome (43), a diagnosis of caffeine withdrawal is not officially recognized in DSM-IV (44). In the present study, more than one-half of the women reported withdrawal symptoms when they attempted to eliminate or cut back their caffeine use during pregnancy, with 14% of the overall study group (26% of those reporting withdrawal symptoms) reporting that their caffeine withdrawal symptoms were severe enough to interfere with their performance at work, school, or home. This rate of functional impairment because of caffeine withdrawal is comparable to the 13% rate of clinically significant distress and functional impairment reported in experimental studies of caffeine abstinence in healthy volunteers (43). The importance of the current study is that it extends the demonstration of functional impairment because of caffeine withdrawal to a medical population.

The small number of subjects (N=44) and the homogeneous nature of the population are limitations of the study. Replication of the study in a larger and more heterogeneous group of subjects would be valuable. There were, however, also strengths in the homogeneity of the study population, which consisted primarily of well-educated, employed, Caucasian pregnant women who reported low current rates of nicotine use and of alcohol or drug problems. As reflected in the low rates of substance use, these women appeared to be quite medically conscientious, with 98% reporting that they attempted to eliminate or reduce caffeine use during pregnancy. The good correspondence between the biological measure of caffeine use and subjects’ self-reported use suggests that the participants were also conscientious in completing the questionnaires and interviews for the study.

The present study has clinical implications for medical management of pregnancy as well as other medical conditions for which caffeine use may be contraindicated. As for pregnancy, the present study shows that the majority of pregnant women spontaneously reduce caffeine intake at time of pregnancy awareness and maintain low levels throughout pregnancy. The study also shows that caffeine-dependent women with a family history of alcoholism had higher levels of caffeine consumption before and throughout pregnancy, compared with women without both characteristics. This subgroup appears to require intervention in addition to instructions from their physician and written materials in order to assure caffeine abstinence.

Finally, the observation that caffeine-dependent women with a family history of alcoholism had higher rates of past cigarette smoking and problematic alcohol use, and possibly other drug use, suggests that the caffeine dependence diagnosis may be a useful clinical and scientific marker for vulnerability to dependence on other drugs of abuse. Because there is little social stigma attached to caffeine use, self-reports may be more accurate for caffeine dependence than for problematic use of alcohol or drugs. Scientifically, it would be valuable if future studies of vulnerability to the more classic forms of drug abuse included measures of caffeine dependence.

Received March 26, 2004; revisions received Sept. 20, 2004, and Jan. 18, 2005; accepted Feb. 14, 2005. From the Department of Psychology, Virginia Commonwealth University (D.S.S.); the Departments of Gynecology and Obstetrics (N.B.), Psychiatry and Behavioral Sciences (R.R.G.), and Neuroscience (R.R.G.), Johns Hopkins University School of Medicine, Baltimore; and the Department of Psychiatry, University of California, San Francisco (N.A.H.). Address correspondence and reprint requests to Dr. Svikis, Virginia Commonwealth University, 1101 East Marshall St., Sanger Hall, Room 11–028, Box 980343, Richmond, VA 23298–0343; dsvikis@hsc.vcu.edu (e-mail).

Supported in part by NIH grants RO1 DA-03890, R01 DA-11476, R01 AA-11802, and P50 DA-09258.

The authors thank Paula Moylan for data entry and chart reviews and Paul Nuzzo for consultation on statistical analysis.

References

ford BB. Chevy Chase, Md, American Society of Addiction, 2003, pp 193–224
Altered Prefrontal Dopaminergic Function in Chronic Recreational Ketamine Users

Rajesh Narendran, M.D.
W. Gordon Frankle, M.D.
Richard Keefe, Ph.D.
Roberto Gil, M.D.
Diana Martinez, M.D.
Mark Slifstein, Ph.D.
Lawrence S. Kegeles, M.D., Ph.D.
Peter S. Talbot, M.D.
Yiyun Huang, Ph.D.
Dah-Ren Hwang, Ph.D.
Leyla Khenissi, B.A.
Thomas B. Cooper, M.A.
Marc Laruelle, M.D.
Anissa Abi-Dargham, M.D.

Objective: Ketamine is a noncompetitive antagonist at the glutamatergic N-methyl-D-aspartate (NMDA) receptor that is currently used in human and animal medicine as an injectable anesthetic. The illegal use of ketamine as a recreational drug is rapidly growing. Very little is currently known about the consequences of repeated ketamine exposure in the human brain. Animal studies indicate that the prefrontal dopaminergic system is particularly vulnerable to the toxic effects of repeated administration of NMDA antagonists. In this study, dopamine D1 receptor availability was assessed by using positron emission tomography and the selective D1 receptor radioligand \([1^{11}C]\)NNC 112 in a group of 14 recreational chronic ketamine users and matched healthy subjects.

Method: History of ketamine abuse was confirmed in subjects by hair analysis. \([1^{11}C]\)NNC 112 binding potential was measured with kinetic analysis using the arterial input function.

Results: Dorsolateral prefrontal cortex D1 receptor availability was significantly up-regulated in chronic ketamine users (\([1^{11}C]\)NNC 112 binding potential: mean=1.68 ml/g, SD=0.40) relative to comparison subjects (mean=1.35 ml/g, SD=0.35). No significant differences were noted in other cortical, limbic, or striatal regions. In the chronic ketamine user group, dorsolateral prefrontal cortex \([1^{11}C]\)NNC 112 binding potential up-regulation was significantly correlated with the number of vials of ketamine (with a vial representing approximately 200–300 mg of ketamine) used per week.

Conclusions: Chronic ketamine users exhibited a regionally selective up-regulation of D1 receptor availability in the dorsolateral prefrontal cortex, a phenomenon observed following chronic dopamine depletion in animal studies. These data suggest that the repeated use of ketamine for recreational purposes affects prefrontal dopaminergic transmission, a system critically involved in working memory and executive function.

Ketamine, a noncompetitive antagonist at the glutamatergic N-methyl-D-aspartate (NMDA) receptor, is currently used in human and animal medicine as an injectable anesthetic (1, 2). Ketamine is also a controlled substance, illegally used as a recreational drug (“Special K,” “Vitamin K”). The recreational use of ketamine is prevalent at dance events (3, 4). While no epidemiological study has specifically addressed the scope and growth rate of the nonmedical use of ketamine, the rates of emergency room visits nationwide for use of ketamine grew 20-fold between 1994 and 1999 (5).

At subanesthetic doses, ketamine induces a state of dissociation (including distortion of space, time, and body image) and feelings ranging from euphoria to detachment, an experience described by users as a mind or spiritual exploration (6–8). While the cognitive deficits observed during acute administration of ketamine are well documented (7, 9, 10), very little is known of the long-term effects of repeated ketamine administration in the human brain. High incidence of psychiatric symptoms, such as recurrent hallucinations and psychotic episodes, have been described in subjects abusing phencyclidine (PCP), a more potent noncompetitive NMDA antagonist (11). In chronic ketamine abusers, limited studies suggest the persistence of neurocognitive deficits up to 3 days after use, but these studies are compromised by the polysubstance use in these samples (12, 13).

The animal literature suggests that repeated exposure to noncompetitive NMDA antagonists leads to sustained impairment of performance in numerous cognitive domains, such as working memory tasks (reviewed by Jentsch and Roth [14]). These deficits induced by NMDA antagonists have been linked to reduced function of the prefrontal dopaminergic system, which plays a critical role in sustaining working memory and executive functions (15–19). Monkeys chronically treated with the noncompetitive antagonist MK-801 showed decreased performance on working memory tasks and decreased prefrontal dopamine levels measured with microdialysis (20). Sustained decrease in prefrontal dopamine leads to an up-regulation of prefrontal dopamine D1 receptors, the main dopaminergic receptor in the cortex (21). A positron emission to-
mography (PET) study showed increased binding of the selective dopamine D1 receptor radiotracer [11C]NNC 112 in the prefrontal cortex of monkeys chronically exposed to MK-801 (20). Thus, in experimental animals, chronic exposure to NMDA antagonists has led to deficits in presynaptic dopaminergic function in the prefrontal cortex, which are associated with a compensatory up-regulation of postsynaptic dopamine D1 receptors.

It is currently unknown if such a phenomenon (decreased prefrontal dopamine function and up-regulation of D1 receptors) is also present in humans chronically exposed to NMDA antagonists. Here, we studied the impact of repeated ketamine exposure on dorsolateral prefrontal cortex D1 receptor binding potential using PET and [11C]NNC 112 in a group of 14 recreational chronic ketamine users and 14 healthy comparison subjects matched for age, gender, race, socioeconomic status of the family of origin, and nicotine smoking.

In addition to examining potential toxic effect of repeated ketamine exposure, this study was also relevant to the pathophysiology of schizophrenia. In schizophrenia patients, we previously observed a regionally selective up-regulation of [11C]NNC 112 binding potential in the dorsolateral prefrontal cortex (22). A deficit in NMDA transmission has been implicated as a fundamental aspect of the pathophysiology of this illness (23–25). Therefore, we speculated that in schizophrenia, a chronic deficit in NMDA transmission might lead to a decrease in prefrontal dopamine function and D1 receptor up-regulation. To document that D1 receptor up-regulation might result from chronic NMDA antagonist exposure in humans would reinforce the biological plausibility of this model.

Method

Subjects

The study was approved by the institutional review boards of the New York State Psychiatric Institute and Columbia University Medical Center. Chronic ketamine users were recruited through flyers distributed by volunteers of Dance Safe.org (promoters of a harm reduction model within the drug abuse community); through discussions with rave/party organizers and closed web groups; and by word of mouth. Study criteria for chronic ketamine users were 1) age between 18 and 50 years; 2) history of at least 2 years of ketamine use, with an average use of one vial per week or more over the last 3 months (a vial contains 200 mg–300 mg of ketamine); 3) history of psychotic symptoms during acute ketamine intoxication; 4) ability to provide 3 cm of hair, and history of ketamine use confirmed by hair analysis (average hair ketamine concentration higher than 10 ng/ml per month in the last 3 months); 5) absence of DSM-IV axis I diagnosis other than ketamine or cannabis abuse or dependence; 6) absence of psychotropic medication for at least 30 days preceding study entry; 7) absence of concomitant or past severe medical conditions; and 8) absence of pregnancy. The healthy comparison subjects had no past or present neurological or psychiatric illnesses including substance abuse. Chronic ketamine users were admitted to an inpatient research unit 3 days before the PET scan to ensure they were drug-free at the time of the scan (the half-life of ketamine is 20 minutes). Healthy comparison subjects underwent the scan as outpatients.

Hair Analysis

Hair samples were collected, and the scalp end was taped to a card so that the laboratory could section the hair and give a chronological estimate of drug intake. The hair was then sectioned into 1–2 cm sections (sufficient to achieve >10 mg weight), weighed, and rinsed with distilled water, methanol, and methylene chloride and then allowed to air dry. Deuterated ketamine and PCP were added as internal standards, and the hair digested in 1M Na2S at 90°C for 10 minutes. Drug-free hair samples were spiked with pure compounds to form a 7-point standard curve encompassing the expected range and processed exactly as samples. The digest was then cooled in ice water and extracted with 1.5% iso-amyl alcohol in heptane and the organic phase back extracted into 0.1 M HCl. This aqueous phase was adjusted to pH 9.5 and extracted into 50 µl 15% iso-amyl alcohol in toluene. A 2-µl aliquot was injected into a GC/MS operated in the EI mode and fitted with a 15 m, 0.25 mm i.d. Trx-5-Amine capillary column (Restek Corp., Bellefonte, Pa.). Simultaneous ion monitoring of the compounds of interest and their respective isotopomers enabled quantitation using the classical isotope dilution method. Standard curves were linear throughout the range with r²=0.99+. Interassay coefficient of variation was <8% for both analyses.

Imaging Protocol

[11C]NNC 112 was prepared by N-methylation of the desmethyl precursor (+)-5-(7-benzofuranyl)-8-chloro-7-hydroxy-3-methyl-2,3,4,5-tetrahydro-1H-3-benzazepine using [11C]methyl triflate as previously described (26). PET imaging sessions were conducted with the ECAT EXACT HR+ camera as previously described (27). Following a 10-minute transmission scan, [11C]NNC 112 was injected intravenously over a 45-second period, and emission data were collected in the three-dimensional mode for 90 minutes.

Input function measurement. Following radiotracer injection, arterial samples were collected every 10 seconds with an automated sampling system for the first 2 minutes and manually thereafter at longer intervals (for a total of 30 samples per experiment). Six samples (collected at 2, 8, 16, 30, 50, and 70 minutes) were further processed by high-pressure liquid chromatography to measure the fraction of plasma activity representing unmetabolized parent compound. The measured input function values (Ci(t), µCi/ml) were analyzed as previously described and used as input to the kinetic analysis of the regional brain uptake (27). The clearance of the parent compound (liters/hour) was calculated as the ratio of the injected dose to the area under the curve of the input function. The determination of the plasma-free fraction (f) was calculated as the ratio of ultrafiltrate to total activity concentrations as previously described (27).

MRI acquisition and segmentation procedures. MRIs were acquired on a GE 1.5-T Signa Advantage system. Steps for MRI segmentation performed within MEDx (Sensor Systems, Inc., Sterling, Va.), with original subroutines implemented in MATLAB (The Math Works, Inc., Natick, Mass.), included correction for field inhomogeneities, fitting of the voxel distribution to a combination of 3 Gaussian, voxel classification, and post filtering (27).

Image analysis. Image analysis was performed blind to subject diagnosis with MEDx (Sensor Systems, Inc., Sterling, Va.). Correction for head movement and coregistration of the PET data to the MR were done with the aid of automated image registration (27).

Derivation of receptor parameters. Derivation of [11C]NNC 112 regional distribution volumes (Vr, ml of plasma/g of tissue) was performed with kinetic analysis using the arterial input function as previously described (22). A one-tissue compartment was
DOPAMINERGIC ALTERATIONS IN KETAMINE ABUSERS

TABLE 1. Demographic and Clinical Characteristics of Chronic Recreational Ketamine Users and Healthy Comparison Subjects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Healthy Subjects (N=14)</th>
<th>Chronic Ketamine Users (N=14)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N %</td>
<td>N %</td>
<td>χ²  df   p</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5 36</td>
<td>4 29</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 64</td>
<td>10 71</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2 14</td>
<td>1 7</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 7</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>11 79</td>
<td>13 93</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>6 43</td>
<td>10 71</td>
<td>2.14     1.2</td>
</tr>
<tr>
<td>Age (years)</td>
<td>26 5 25</td>
<td>26 0.38</td>
<td>26 0.71</td>
</tr>
<tr>
<td>Socioeconomic statusα</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>34 16 38 16 0.67</td>
<td>26 0.51</td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>47 14 51 11 0.67</td>
<td>26 0.51</td>
<td></td>
</tr>
</tbody>
</table>

α Measured with the Hollingshead scale.

used in the cerebellum, and a two-tissue compartment in other regions. The primary outcome measure was the binding potential (binding potential, ml/g), derived as the difference in V₃ between the region of interest and the cerebellum, a region with negligible levels of D₁ receptors. The relationship of [¹¹C]NNC 112 binding potential and D₁ receptor parameters can be calculated as binding potential=V₃/V₂KD, where V₃ (unitless) is the free fraction of [¹¹C]NNC 112 in plasma, bₐₚ₃ (nanomoles/g of tissue) is the regional concentration of D₁ receptors, and Kₐₚ₃ (nanomoles/ml of water) is the affinity of [¹¹C]NNC 112 for D₁ receptors. Another outcome measure frequently used in PET neuroreceptor imaging is the equilibrium specific-to-nonspecific partition coefficient, denoted here V₃′. The relationship of V₃′ and D₁ receptor parameters can be calculated as binding potential=V₃′/V₂KD, where V₂ (ml/g) is the nonspecific volume of distribution. V₃′ was calculated as binding potential/cerebellum V₃.

Analyses

Region of interest. Kinetic analysis was performed first on region of interest time-activity curves. Regions of interest included the dorsolateral prefrontal cortex, medial prefrontal cortex, orbitofrontal cortex, parietal cortex, temporal cortex, occipital cortex, anterior cingulate cortex, amygdala, entorhinal cortex, hippocampus, parahippocampal gyrus, associative striatum, sensorimotor striatum, and ventral striatum. Anatomical criteria and methods used to delineate the cortical regions of interest and striatal regions of interest can be found in Abi-Dargham et al. (22) and Martinez et al. (28), respectively.

Voxelwise. In a secondary analysis, kinetic analysis was performed on each voxel to derive V₃. Binding potential was computed at each voxel as V₃(Voxel)-V₃(cerebellum), where V₃(Voxel) was computed using a two-tissue compartment model and V₃(cerebellum) was computed as the mean across cerebellum voxels of V₃ computed with a one-tissue compartment model. All data were fitted with a basis function approach (29). Binding potential maps were then normalized to the MINI T₁ template in SPM 99 (30) and compared across groups.

Statistical. For the region of interest analysis, between-group comparisons were performed using two-tailed unpaired t tests. For the voxelwise analysis a one-tailed two-group t test for binding potential of chronic ketamine users versus healthy subjects was computed in SPM 99. A probability value of 0.05 was selected as significance level for both analyses. On the basis of the aforementioned animal data, the primary hypothesis of this study was that dorsolateral prefrontal cortex D₁ receptor availability would be altered in chronic ketamine users. Therefore, no correction for multiple comparison testing was applied to this region (31). Other regions were analyzed in an exploratory fashion, to investigate the regional specificity of potential findings in the dorsolateral prefrontal cortex.

Neurocognitive Assessment

A neurocognitive battery was administered the day before the scan to both groups of subjects. The primary neurocognitive test used in this study was the N-back test (32). The N-back paradigm engages white matter by requiring subjects to maintain information about previous stimuli as well as manipulate this information (i.e., to make a comparison with the current stimuli). The N-back was selected as the primary neurocognitive test because a previous study in patients with schizophrenia found a significant relationship between up-regulated D₁ receptors in the dorsolateral prefrontal cortex (measured with [¹¹C]NNC 112 using a method similar to the one used in this study) and poor performance on the N-back (22).

In addition to the N-back task, subjects were evaluated using a battery of neurocognitive tests developed for the schizophrenia trial of the National Institute of Mental Health-sponsored Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project. The CATIE battery comprises a wide variety of tasks aiming to characterize subjects across multiple domains of cognitive functioning. The battery has been described in fuller detail by Keefe et al. (33).

Results

Following a telephone screening interview, 54 potential chronic ketamine users were evaluated at the research site. Thirty-three potential participants were excluded from the study because of medical conditions (N=4), psychiatric comorbidity (N=3), abuse of other drugs (N=12), or use of ketamine below study criteria (N=14). Among the 21 subjects who met study criteria, 14 agreed to participate in the study after explanation of the nature and the risks of the study. All recruited subjects completed the study.

The two study groups are further described in Table 1. The groups were predominantly composed of young Caucasian males. Chronic ketamine users had used ketamine for a mean of 4.1 previous years (SD=2.4, range=11 months to 2 years). Age did not correlate with duration of use (r²=0.01, df=12, p=0.68). Over the preceding 3 months, the mean ketamine consumption was 2.8 vials per week (SD=1.9, range=1–7). Twelve subjects reported inhalation of ketamine as their usual mode of administration, and two subjects reported intramuscular injection. Two of the 14 chronic ketamine users also met criteria for cannabis dependence. Mean hair ketamine concentration was 32 ng/ mg of hair (SD=21, range=9–70).

Groups were matched for familial socioeconomic status. No group difference was found in the socioeconomic status of the subjects. Among the chronic ketamine users, six were employed full-time, three were employed part-time, two were students, and three were unemployed. The difference between socioeconomic status of subjects and...
their family of origin was similar in the chronic ketamine users (mean=–10 points [SD=22] on the Hollingshead scale) and healthy subjects (mean=–13 points [SD=19]). Hence, the frequent use of ketamine did not result in a significantly lower socioeconomic status in chronic ketamine users than in healthy subjects from similar socioeconomic backgrounds.

Scan Parameters

Critical PET scan parameters are listed in Table 2. [11C]NNC 112 injected dose, specific activity at time of injection, and injected mass did not differ between the groups. No significant between-group differences were observed in the clearance rate of [11C]NNC 112 from the plasma compartment, in [11C]NNC 112 plasma free fraction, or in [11C]NNC 112 cerebellum distribution volume.

Regional Volumes

No significant between-group differences were found in dorsolateral prefrontal cortex volumes (healthy subjects: mean=49.9 cm³ [SD=16.1]; chronic ketamine users: mean=47.9 cm³ [SD=10.0]) nor in volumes of the other regions. These data indicate that the use of ketamine in the subjects included in the study did not result in detectable changes in regional brain volumes.

D1 Receptor Measurement

Region of interest analysis. As seen in Figure 1, dorsolateral prefrontal cortex [11C]NNC 112 binding potential was significantly higher in chronic ketamine users compared with the healthy subjects. Similar results were obtained when V3′′ was used as the outcome measure (healthy subjects: mean=0.68 ml/g, SD=0.20; chronic ketamine users: mean=0.85 ml/g, SD=0.22) (t=2.37, df=26, p<0.03). The two chronic ketamine users with highest dorsolateral prefrontal cortex [11C]NNC 112 binding potential positively correlated with the number of vials of ketamine used per week (Figure 2). Dorsolateral prefrontal cortex [11C]NNC 112 binding potential was not associated with the duration of ketamine use (r²<0.01, p=0.77) nor with the average hair ketamine concentration (r²<0.01, p=0.87).

Statistical parametric mapping analysis. Group comparison of [11C]NNC 112 binding data on a voxel basis (results shown in Figure 3) revealed various clusters of increased [11C]NNC 112 in chronic ketamine users relative to comparison subjects. The main area was localized within the dorsolateral prefrontal cortex, confirming the finding of the region of interest analysis. No clusters of decreased [11C]NNC 112 in chronic ketamine users relative to comparison subjects were identified.

Hair Analysis

Mean hair ketamine concentration in the chronic ketamine users was 32 ng/mg of hair (SD=21, range=9–70). Mean hair concentration was not correlated with the average number of vials reported per week (r²=0.07, p=0.33). This lack of correlation was not unexpected because of possible errors in the reported frequency of use and because of several factors such as hair pigmentation and frequency of hair washing, which affect the relationship between dose exposure and hair concentration.

<table>
<thead>
<tr>
<th>Scan Parameter</th>
<th>Healthy Subjects (N=14)</th>
<th>Chronic Ketamine Users (N=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>Mean SD</td>
<td></td>
</tr>
<tr>
<td>Injected dose (mCi)</td>
<td>13.5 3.7</td>
<td>13.8 4.5</td>
</tr>
<tr>
<td>Specific activity (Ci/mmol)</td>
<td>1107 576</td>
<td>1169 543</td>
</tr>
<tr>
<td>Injected mass (µg)</td>
<td>4.5 1.7</td>
<td>4.3 1.3</td>
</tr>
<tr>
<td>Plasma free fraction (%)</td>
<td>0.90 0.35</td>
<td>1.03 0.61</td>
</tr>
<tr>
<td>Plasma clearance (liter/hour)</td>
<td>87.5 25.6</td>
<td>91.1 25.1</td>
</tr>
<tr>
<td>Cerebellum Vt (ml/g)</td>
<td>2.1 0.4</td>
<td>2.0 0.4</td>
</tr>
</tbody>
</table>

* Chronic ketamine users displayed increased D1 receptor availability in the dorsolateral prefrontal cortex ([11C]NNC 112 binding potential: mean=1.68 ml/g, SD=0.40) relative to healthy subjects (mean=1.35 ml/g, SD=0.35) (t=2.34, df=26, p<0.03).
Neurocognitive Assessment

No significant performance differences were found between the groups on tests involving working memory (semantic, visuospatial, and auditory), executive functions, attention, reaction time, verbal learning and memory, verbal fluency, motor function, and intellectual functioning. Chronic ketamine users performed better than healthy subjects on a social cognition task. In the chronic ketamine users, no significant relationships were found between dorsolateral prefrontal cortex \([11C]\)NNC 112 binding potential and performance on the N-back tests (1-back: \(r^2=0.19, p=0.15\); 2-back: \(r^2=0.05, p=0.50\); 3-back: \(r^2=0.19, p=0.14\)).

Discussion

In this study, we investigated D\(_1\) receptor availability in a group of subjects who regularly used ketamine for recreational or mind-exploring purposes. These subjects had a strong and exclusive interest in ketamine (with the exception of two subjects who also abused cannabis). The regular use of ketamine in this cohort did not result in a detectable downward shift in socioeconomic status. As per inclusion criteria, these subjects were free of medical, neurological, and psychiatric conditions while not under the influence of ketamine. In contrast to the ketamine users studied previously, who were polydrug abusers (12, 13), subjects included in this cohort did not exhibit neurocognitive alterations on a battery of tests administered after 2 days of monitored abstinence. The clinical read and volumetric analysis of the MRI was unremarkable.

In contrast to cognitive performances and regional brain volumes, which were similar in healthy subjects and chronic ketamine users, the PET scan detected an increase in \([11C]\)NNC 112 binding potential in chronic ketamine users that reached significance only in the dorsolateral prefrontal cortex in both the region-of-interest and voxel-based analyses. The quantitative method used in this study is not sensitive to possible alterations in regional cerebral blood flow that could be associated with ketamine use (34–36). In addition, no between-group differences were observed in \([11C]\)NNC 112 plasma-free fraction, in \([11C]\)NNC 112 nonspecific binding (cerebellum \(V_T\)), or in dorsolateral prefrontal cortex volumes. Under these conditions, in-
creased $^{[11C]}$NNC 112 binding potential and $V_3$” clearly indicated increased availability of $D_1$ receptors. This increased availability could be due to increased receptor density or affinity. The imaging methods used in this study do not discriminate between these mechanisms.

The significance of this finding did not survive correction for multiple comparisons, either in region-of-interest or in statistical parametric analyses. Since animal data indicated that the dorsolateral prefrontal cortex dopamine projections were especially vulnerable to repeated NMDA antagonist administration, this study was primarily designed to look at this region, and under these conditions, no multiple comparison correction is required to ascertain the statistical significance of the findings. Regarding statistical parametric analysis, it has been argued by Friston and colleagues (31) that the use of “corrected” $p$ values in statistical parametric analysis is unnecessary and inappropriately conservative when the target region of interest is predicted in advance.

It cannot be ascertained if this increased $D_1$ receptor availability in the dorsolateral prefrontal cortex of chronic ketamine users is associated with a vulnerability to develop ketamine abuse or is a consequence of repeated ketamine exposure. The positive correlation between average number of ketamine vials used per week and dorsolateral prefrontal cortex $^{[11C]}$NNC 112 binding potential does not, per se, indicate that this alteration is a result of toxic effects of ketamine. However, the fact that a similar increase in dorsolateral prefrontal cortex $^{[11C]}$NNC 112 binding potential has been observed in nonhuman primates chronically treated with the NMDA antagonist MK-801 strongly supports the hypothesis that elevated $^{[11C]}$NNC 112 binding potential is a consequence of repeated ketamine exposure. In rodents, glutamatergic projections from the prefrontal cortex to the ventral tegmental area make direct synaptic contacts onto dopaminergic cells that project back to the cortex (37). This circuit provides an anatomical substrate by which prefrontal dopamine projections might be more vulnerable than other dopamine projections to repeated alterations of glutamatergic transmission induced by ketamine.

The normality of the cognitive performances in the chronic ketamine users was an unanticipated result of this study. From the aforementioned animal studies, it is reasonable to postulate that the $D_1$ receptor up-regulation observed in the present study might be secondary to a drug-induced deficit in prefrontal dopamine function. Given the absence of detectable neurocognitive impairment observed in these subjects, it is tempting to speculate that, at this stage in the condition, the up-regulation of $D_1$ receptors might be relatively efficient at compensating for a deficit in prefrontal dopamine function. However, this up-regulation might be an early sign of system dysregulation. Because of the design of the study, subjects with potentially more severe consequences of NMDA antagonist exposure were less likely to participate, since alcoholism, polysubstance abuse, and comorbid psychopathology tend to develop with the progression of the addiction and were exclusion criteria.

It is very important to stress that the recruitment criteria (absence of psychiatric comorbidity and other substance abuse) used in this study resulted in a sample of “high-functioning” recreational ketamine users, and that this group might not be representative of the majority of ketamine abusers. Therefore, the absence of cognitive deficits in chronic ketamine users enrolled in this study does not indicate that the recreational use of ketamine is safe for cognitive functions. In fact, this study shows that, even in the absence of cognitive deficits, repeated ketamine exposure is associated with signs of disruptions of a critical component of cognition, the prefrontal dopamine system.

The findings of the present study also have implications for the pathophysiology of schizophrenia. Schizophrenia, a severe and chronic mental illness, is believed to be associated with an imbalance in dopamine transmission, characterized by a persistent deficit in prefrontal cortical dopamine function involving $D_1$ receptors (contributing to the cognitive impairment) and an intermittent excess of subcortical dopamine function involving $D_2$ receptors.

FIGURE 3. Regions of Significantly Greater $^{[11C]}$NNC 112 Binding Potential in Chronic Ketamine Users (N=14) Relative to Healthy Comparison Subjects (N=14)$^a$

\[ a \text{ One-tailed two-group test computed in SPM 99; the rendered t statistic is thresholded to uncorrected } p<0.01. \text{ The primary brain region showing at this level is the right dorsolateral prefrontal cortex.} \]
DOPAMINERGIC ALTERATIONS IN KETAMINE ABUSERS

...effects of the noncompetitive NMDA antagonist, ketamine, in humans: psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry 1994; 51:199–214
14. Jentsch JD, Roth RH: The neuropsychopharmacology of phencyclidine: from NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia. Neuropsychopharmacology 1999; 20:201–225

References

14. Jentsch JD, Roth RH: The neuropsychopharmacology of phencyclidine: from NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia. Neuropsychopharmacology 1999; 20:201–225

Received April 14, 2004; revision received Jan. 10, 2005; accepted Feb. 23, 2005. From the Departments of Psychiatry and Radiology, Columbia University College of Physicians and Surgeons, New York; and the Department of Psychiatry, Duke University, Durham, NC. Address correspondence and reprint requests to Dr. Narendran, New York State Psychiatric Institute, 1051 Riverside Dr., Box 31, New York, NY 10032; nn2012@columbia.edu (e-mail).

Supported by a grant from the Lieber Center for Schizophrenia Research at Columbia and an Independent Investigator Award from the National Alliance for Research on Schizophrenia and Depression to Dr. Abi-Dargham.

The authors thank Carol Caton, Ph.D., and John Keilp, Ph.D., from Columbia University for their assistance in conducting the study.

http://ajp.psychiatryonline.org

Am J Psychiatry 162:12, December 2005


34. Lahti AC, Holcomb HH, Medoff DR, Tamminga CA: Ketamine activates psychosis and alters limbic blood flow in schizophrenia. Neuroreport 1995; 6:869–872


Damage to the orbitofrontal cortex has been associated with disinhibited or socially inappropriate behavior and emotional irregularities (1–5). Prominent characteristics of several personality disorders, in particular DSM-IV borderline personality disorder, include impulsivity and affective instability. Despite the abundance of evidence linking impulsivity to frontal lobe dysfunction, evidence for underlying frontal lobe deficits in personality disorder patients with a history of impulsivity is limited, and there is even less evidence linking borderline personality disorder specifically to frontal lobe deficits (6). Neuroimaging studies show differences in the prefrontal cortex in people with borderline personality disorder, compared to healthy subjects, at baseline (7–9) and in response to aversive stimuli (10) and neuropharmacological probes associated with impulsivity (11, 12). Some evidence from brain imaging studies suggests orbitofrontal cortex dysfunction in borderline personality disorder (13–16), specifically hypometabolism (7, 8, 11) and smaller volume (9, 13) of the orbitofrontal cortex.

In this study, our aim was to determine if certain aspects of borderline personality disorder, in particular impulsivity, are associated with orbitofrontal cortex dysfunction. We used questionnaires on personality, emotion, and impulsivity together with a number of computer-based tasks that are sensitive to frontal lobe dysfunction to assess possible underlying factors related to impulsivity in patients with orbitofrontal cortex lesions, patients with lesions in the prefrontal cortex but not in the orbitofrontal cortex, patients with borderline personality disorder and a history of self-harm, and healthy comparison subjects. This combination of tests allows for new comparisons of cognitive and behavioral traits and for identification of patterns and dissociations between different types of dysfunctions within borderline personality disorder. A unique feature of this study is the comparison of...
the clinical population of borderline personality disorder patients to patients with lesions in the prefrontal cortex. This study design provides an excellent opportunity to increase understanding of the biological etiology of borderline personality disorder as well as of the functions of the brain.

On the basis of previous research, we hypothesized that frontolimbic circuit dysfunction, involving both the orbitofrontal cortex and amygdala regions, may be involved in borderline personality disorder and that some aspects of borderline personality disorder may be related to dysfunction of the orbitofrontal cortex. Our goal is to provide evidence of these relationships by comparing the performance of patients with borderline personality disorder and patients with orbitofrontal cortex lesions on a number of tests. In the absence of uniquely defined and identifiable neurological damage in patients with borderline personality disorder, this study is intended to provide information about the neurobiological correlates of certain behaviors in borderline personality disorder.

Method

Participants

Ethics approval was obtained from the University of Oxford, the Institute of Psychiatry (King’s College, London), and the Oxfordshire Psychiatric Research Ethics Committee. After complete description of the study to the subjects, written informed consent was obtained.

Healthy comparison subjects. The healthy comparison group consisted of 39 participants (10 male participants) ranging in age from 18 to 71 years (mean=40.3 years, SD=20.5). The healthy comparison subjects were recruited through the subject panel at the Department of Neurosurgery, King’s College Hospital, London (N=19); the International Subarachnoid Aneurysm Trial at the Radcliffe Infirmary, Oxford, England (N=16); and the Oxford Centre for Enablement, Oxford, England (N=8). The time since the patients sustained their lesion varied considerably, from 6 months to 20 years (mean=5.0 years, SD=4.8). Clinical information for each lesion patient and the basis for division into orbitofrontal cortex and dorsolateral prefrontal cortex lesion groups are summarized in Table 1 of our earlier publication (5).

Twenty patients (eight male patients) constituted the comparison patients with prefrontal cortex lesions outside the orbitofrontal cortex. The patients ranged in age from 19 to 71 years (mean=46.0 years, SD=15.1). The dorsolateral prefrontal cortex (with or without the medial prefrontal cortex) was the main site of damage in the patients included in this group.

The patients with orbitofrontal cortex lesions consisted of 23 patients (15 male patients) ranging in age from 30 to 63 years (mean=48.7 years, SD=10.0) with damage including or restricted to the orbitofrontal cortex (either bilaterally or unilaterally).

Patients with borderline personality disorder. Nineteen self-harming inpatients with DSM-IV borderline personality disorder (18 female patients) ranging in age from 19 to 49 years (mean=32.37 years, SD=8.4) were tested at the Bethlem Royal Hospital Crisis Recovery Unit, London. (About 75% of borderline personality disorder diagnoses are made in female patients.) The unit provides a rehabilitation program in which the patients are taught to seek alternatives to and gain a better understanding of self-harm and to tolerate distressing feelings. Cutting and burning were among the most common forms of self-harm in the patients who participated in the study. Self-harm provided an objective indication of impulsive behavior and was an important factor in obtaining a homogenous patient group. Descriptive and systematic data reveal that repetitive self-mutilation is typically an impulsive act (17). Potential participants were excluded if they had current substance or alcohol abuse. Because all of the patients were in a 6-month inpatient rehabilitation program, we knew that for at least the time they were in the program they were not dependent on substances or alcohol. (Although most patients with borderline personality disorder were taking medications intended to improve their borderline personality disorder symptoms, they still performed poorly on tasks related to impulsivity, emotion, and personality and did not perform poorly on other tasks such as spatial working memory tasks.)

Materials and Procedures

See our earlier publication (5) for the rationale for choice of measures.

Questionnaires. The self-report impulsivity measure was the Barratt Impulsiveness Scale, version 11 (18). This 30-item, 4-point Likert-type questionnaire was used as a trait measure of impulsivity. Long-term patterns of behavior were assessed with three subscales: nonplanning, motor, and cognitive impulsivity.

The personality questionnaire was the Big Five Inventory (19), a 44-item, 5-point Likert-type questionnaire designed to measure the five scales/domains of the Five-Factor Model of Personality (20): extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (21).

The front behavioral questionnaire is a self-report 20-item, 5-point Likert-type questionnaire (5) designed to measure types of behavioral problems believed to result from frontal brain damage (22), e.g., disinhibition, social inappropriateness, perseveration, and cooperativeness.

The borderline personality disorder questionnaire (23), which consists of 18 yes/no questions modeled on the DSM-III criteria for borderline personality disorder, was used for assessment of borderline personality disorder characteristics.

The subjective emotion questionnaire measures how often participants experience sadness, anger, fear, happiness, and disgust in their current daily life. Ratings are made on a 4-point Likert-type scale. The emotional change questionnaire was given only to the patients with lesions and the patients with borderline personality disorder. This questionnaire assesses change (increase, decrease, or no change) since brain injury or onset of personality disorder in the capacity to feel sadness, anger, fear, happiness, and disgust. Both questionnaires were adapted from a previously developed subjective emotion questionnaire (1, 24).

Tests. The Probabilistic Reversal Test includes an “acquisition” task in which participants learn to touch one of two patterns on the computer screen and to avoid touching the other and a “reversal” task in which they then learn to reverse or extinguish that choice based on monetary rewards and punishers (5). Measures of performance include 1) punishment insensitivity (the total number of consecutive touches to a stimulus after having lost a minimum of £250), which measures the extent to which participants fail to switch immediately from a stimulus on the next trial following a
large loss, and 2) reward insensitivity (the total number of times a participant touched a stimulus and won a minimum of £80 but did not touch the same stimulus again on the next trial), which measures the extent to which participants fail to select the same stimulus again after a large gain was associated with that stimulus.

The Matching Familiar Figures Test (25), a standard cognitive behavioral measure of impulsivity, measures reflection-impulsivity, a composite of two dimensions: 1) latency to first response and 2) accuracy of choice or total errors. Participants observe a standard picture and select from a set of highly similar pictures the one that is the same as the standard picture. Twelve target objects with eight variants each were administered. The mean time of the latency of participants’ first response across all trials and the number of errors made before choosing the correct item were recorded.

The Spatial Working Memory Task from the Cambridge Neuropsychological Test Automated Battery (CeNeS Ltd., Cambridge, U.K.), a standard test (5, 26), was used as a comparison condition. The measures were 1) between errors (number of times the subject revisits a box where a token has already been found), 2) within errors (number of times a subject revisits a box already found to be empty during the same search), and 3) strategy (the number of times the subject begins a new search with the same box) (26). A high score denotes poor use of the strategy, and a low score denotes effective use.

### Table 1. Correlation of Scores on the Most Salient Variables of Measures of Personality, Emotion, Impulsivity, Sensitivity to Reinforcers, and Spatial Working Memory in a Study of the Contribution of Orbitofrontal Cortex Dysfunction to Borderline Personality Disorder

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Total self-report impulsivity score (N=101)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Extraversion (N=101)</td>
<td></td>
<td>–0.32*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Agreeableness (N=101)</td>
<td></td>
<td>–0.25</td>
<td>0.28*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.02</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Conscientiousness (N=101)</td>
<td></td>
<td>–0.67</td>
<td>0.40</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>&lt;0.0005</td>
<td>&lt;0.0005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Neuroticism (N=101)</td>
<td>0.46</td>
<td>–0.57</td>
<td>–0.33*</td>
<td>–0.48</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>&lt;0.0005</td>
<td>0.001</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>6 Openness to experience (N=101)</td>
<td></td>
<td>–0.37</td>
<td>0.22</td>
<td>0.06</td>
<td>0.23</td>
<td>–0.25</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>0.03</td>
<td>0.58</td>
<td>0.021</td>
<td>0.012</td>
</tr>
<tr>
<td>7 Frontal behavior questionnaire score (N=101)</td>
<td>0.51</td>
<td>–0.20</td>
<td>–0.47</td>
<td>–0.31*</td>
<td>0.46</td>
<td>–0.24</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>0.049</td>
<td>&lt;0.0005</td>
<td>0.002</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>8 Total subjective emotion score (N=100)</td>
<td>0.43</td>
<td>–0.35</td>
<td>–0.21</td>
<td>–0.29*</td>
<td>0.59</td>
<td>–0.10</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>&lt;0.0005</td>
<td>0.003</td>
<td>&lt;0.0005</td>
<td>0.31</td>
</tr>
<tr>
<td>9 Borderline personality disorder questionnaire score (N=60)</td>
<td>0.57</td>
<td>–0.28</td>
<td>–0.32</td>
<td>–0.47</td>
<td>0.60</td>
<td>–0.01</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.0005</td>
<td>0.031</td>
<td>0.013</td>
<td>&lt;0.0005</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>10 Behavioral impulsivity (errors/second on the Matching Familiar Figures Test)</td>
<td>0.33*</td>
<td>0.06</td>
<td>–0.17</td>
<td>–0.16</td>
<td>0.26</td>
<td>–0.21</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>0.003</td>
<td>0.59</td>
<td>0.14</td>
<td>0.17</td>
<td>0.019</td>
</tr>
<tr>
<td>N</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>11 Total amount earned on the reversal task (£)</td>
<td>–0.20</td>
<td>–0.15</td>
<td>–0.14</td>
<td>–0.04</td>
<td>–0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>&lt;0.061</td>
<td>0.14</td>
<td>0.17</td>
<td>0.67</td>
<td>0.51</td>
</tr>
<tr>
<td>N</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>12 Between errors on the Spatial Working Memory Task</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
<td>–0.09</td>
<td>–0.32*</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>0.33</td>
<td>0.44</td>
<td>0.40</td>
<td>0.54</td>
<td>0.42</td>
</tr>
<tr>
<td>N</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>13 Total time estimation</td>
<td>0.19</td>
<td>–0.01</td>
<td>0.10</td>
<td>–0.05</td>
<td>0.13</td>
<td>–0.39</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>0.10</td>
<td>0.92</td>
<td>0.37</td>
<td>0.67</td>
<td>0.25</td>
</tr>
<tr>
<td>N</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>14 Total time production</td>
<td>–0.33*</td>
<td>0.06</td>
<td>0.11</td>
<td>0.23</td>
<td>–0.26</td>
<td>0.35*</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>0.003</td>
<td>0.61</td>
<td>0.34</td>
<td>0.043</td>
<td>0.022</td>
</tr>
<tr>
<td>N</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
</tbody>
</table>

*Correlations were calculated by using scores of all subjects in the study, including patients with borderline personality disorder (N=19), patients with orbitofrontal cortex lesions (N=23), patients with lesions in the prefrontal cortex but not in the orbitofrontal cortex (N=20), and healthy comparison subjects (N=39). A Bonferroni correction was applied to the correlations, resulting in a critical alpha level for any one correlation of 0.0005. Significant correlations at p<0.0005 (two-tailed) are highlighted in boldface type. Significant correlations at p<0.01 (two-tailed) are marked with an asterisk.
The time perception task (5) has two elements: time estimation and time production. For time estimation, participants were asked to estimate time intervals (10, 30, 60, and 90 seconds; each presented twice in a random sequence) during which they were distracted by being required to read aloud from a computer screen a series of randomly generated numbers from 1 to 9 that ranged in presentation time from 100 to 2900 milliseconds in order to prevent subvocal counting. The number of seconds estimated after each interval, averaged across two runs, and the total time estimated (the sum of the average times estimated at each interval divided by 190, which was the total number of seconds that actually passed) were recorded. The procedure for time production was the same as that for time estimation (reading aloud randomized numbers) except that participants were asked to press a computer key when they thought a set number of seconds had passed. For each time interval, the time produced was compared to the actual time participants were asked to produce. The number of seconds produced at each interval, averaged across two runs, and the total time produced (sum of the average times produced at each interval divided by 190, which was the total number of seconds participants were asked to produce) were recorded.

**Statistical Analyses**

A one-way analysis of variance (ANOVA) was performed for each of the variables to determine if the mean scores differed significantly by group. If an ANOVA yielded a significant F
value, a Fisher’s least significant difference post hoc test was performed to identify the specific source of the difference. An alpha level of 0.05 was used for all statistical tests. Kruskal-Wallis nonparametric tests were performed for variables for which normality was not present. If the Kruskal-Wallis test yielded nonsignificant results, no results are reported for that variable.

Because ANOVAs revealed between-group differences in terms of age (F=4.36, df=3, 97, p=0.006) and gender (F=7.30, df=3, 97, p<0.001), analyses of covariance were performed across all variables, and these analyses discounted any confounding effects of age or gender on the results. No significant difference in the number of years since their lesion was found between the patients with orbitofrontal cortex lesions (mean=4.94, SD=4.22) and those with lesions in the prefrontal cortex but not in the orbitofrontal cortex (mean=4.95, SD=5.40), and years since lesion did not correlate significantly with any of the test variables, with Bonferroni correction (see Results section).

The patients recruited from King’s College completed most questionnaires, but they did not complete the behavioral tests because of time constraints and did not complete the borderline personality disorder questionnaire because of sensitivity issues. A few other participants did not complete all of the tasks because of testing time constraints.

Results

When not otherwise noted, all stated differences for post hoc least significant difference analyses were significant at p<0.05, and many were significant at p<0.001.

Similarities Between Patients With Orbitofrontal Cortex Lesions and Borderline Personality Disorder Patients

Results are presented here (and in Figure 1, Figure 2, and Figure 3) for measures on which both the orbitofrontal cortex lesion group and the borderline personality disorder group were significantly impaired, compared to the healthy group (and, for some variables, compared to patients without orbitofrontal cortex lesions), regardless of whether they differed significantly from each other.

Self-report impulsivity. ANOVAs performed on the self-report impulsivity total score and on the three subscale variables were all significant (total self-report impulsivity score: F=25.89, df=3, 97, p<0.001; nonplanning impulsivity: F=20.82, df=3, 97, p<0.001; motor impulsivity: F=10.83, df=3, 97, p<0.001; and cognitive impulsivity: F=13.42, df=3, 97, p<0.001).

Least significant difference post hoc tests showed that on all self-report impulsivity measures, the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder were significantly more impulsive than both the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex, the patients with borderline personality disorder were more impulsive than the patients with orbitofrontal cortex lesions, and the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex.
the healthy comparison subjects, and the patients with orbitofrontal cortex lesions. There was no significant difference between the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex (for number of errors per second). An alpha level of 0.05 was used for all statistical tests.

*p<0.05. **p<0.001.

cortex did not differ significantly from each other (Figure 1).

**Behavioral impulsivity.** ANOVAs performed on the Matching Familiar Figures Test variables (Figure 2 and Figure 3) showed significant differences between groups on all variables (errors per second: F=6.49, df=3, 75, p=0.001; number of errors: F=11.49, df=3, 75, p<0.001; and time latency: F=5.28, df=3, 75, p=0.002).

Post hoc analysis revealed that patients with orbitofrontal cortex lesions and patients with borderline personality disorder were more behaviorally impulsive than the healthy comparison subjects on all three measures and that the patients with orbitofrontal cortex lesions were more impulsive than the patients with borderline personality disorder on all but the time latency measure. The healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex did not differ significantly from each other on the errors per second and time latency measures, but the patients with lesions outside the orbitofrontal cortex lesions were more impaired on the number of errors measure.

**Frontal behavior questionnaire.** There were significant differences between groups on the measure of behavioral problems believed to result from frontal damage (F=12.40, df=3, 97, p<0.001) (Figure 1). Post hoc analysis revealed that the mean frontal behavior questionnaire total scores of the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder were significantly higher than those of both the patients with lesions outside the orbitofrontal cortex and the healthy comparison subjects. Post hoc analysis showed that the mean frontal behavior questionnaire total score of the patients with orbitofrontal cortex lesions was significantly higher than that of the patients with orbitofrontal cortex lesions. There was no significant difference between the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex.

**Borderline personality disorder questionnaire.** There were significant group differences in scores on the borderline personality disorder questionnaire (F=21.40, df=3, 55, p<0.001) (Figure 1). Post hoc analysis revealed that the scores of the patients with borderline personality disorder were significantly higher than those of all other groups (p<0.001) and that the patients with orbitofrontal cortex lesions had significantly higher scores than those of the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex (p<0.05). There was no significant difference between the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex.

**Time production.** ANOVAs showed significant group differences in total time production and in time production at 60 and 90 seconds (total time production: F=2.95, df=3, 74, p<0.05; time production at 60 seconds: F=2.82, df=3, 74, p<0.05; and time production at 90 seconds: F=5.04, df=3, 74, p<0.01). Whereas all subjects overproduced time (except the patients with orbitofrontal cortex lesions at 90 seconds), least significant difference post hoc analysis revealed that the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder produced significantly less time than the healthy comparison subjects (indicating a sped-up subjective sense of time) in measures of total time production (p<0.05) and of time production at 60 seconds (p<0.05) and 90 seconds (p<0.01). Underproduction became more apparent at longer time intervals (Figure 3).

**Personality.** ANOVAs performed on each of the personality questionnaire (Big Five Inventory) variables indicated that participants’ scores on extraversion, conscientiousness, neuroticism, and openness to experience were significantly related to their group (extraversion: F=12.15, df=3, 97, p<0.001; conscientiousness: F=12.71, df=3, 97, p<0.001; neuroticism: F=23.00, df=3, 97, p<0.001; openness to experience: F=6.43, df=3, 97, p=0.001). However,
openness to experience (Figure 3) was the only variable on which both the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder were impaired. The other variables are discussed later because only the patients with borderline personality disorder were affected. The healthy comparison subjects were significantly more open to experience than all other participants ($p < 0.005$).

**Subjective emotion questionnaire.** ANOVAs performed on each of the variables (total subjective emotion score, sadness, anger, fear, disgust, and happiness) revealed that participants’ scores on all six measures were significantly related to their group (total subjective emotion: $F=17.17$, df=3, 96, $p<0.001$; sadness: $F=10.84$, df=3, 96, $p<0.001$; anger: $F=14.48$, df=3, 96, $p<0.001$; fear: $F=23.78$, df=3, 96, $p<0.001$; disgust: $F=6.62$, df=3, 96, $p<0.001$; happiness: $F=19.00$, df=3, 96, $p<0.001$).

Post hoc tests revealed that the patients with borderline personality disorder and the patients with orbitofrontal cortex lesions both had a deficit in subjective anger and subjective happiness (Figure 1). The subjective anger of the patients with borderline personality disorder was significantly higher and their subjective happiness was significantly lower, compared to all other groups. The subjective anger of the patients with orbitofrontal cortex lesions was significantly higher and their subjective happiness was significantly lower, compared to the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex. There were no significant differences between the scores of the healthy comparison subjects and the patients with lesions outside the orbitofrontal cortex on these measures. All other variables are discussed in the next section because significant differences were found only for patients with borderline personality disorder.

**Differences Between Patients With Orbitofrontal Cortex Lesions and Borderline Personality Disorder Patients**

Histograms representing the means of variables discussed in this section are shown in Figure 4 and Figure 5.

**Personality.** ANOVAs for each personality (Big Five Inventory) variable indicated that participants’ scores on extraversion, conscientiousness, neuroticism (Figure 4), and openness to experience (Figure 3) were significantly related to their group (see statistical test results reported earlier). Post hoc tests revealed that the patients with borderline personality disorder were significantly less extra-
verted and conscientious and more neurotic than the participants in all other groups (p<0.001).

**Subjective emotion.** Post hoc analysis (conducted after the ANOVAs reported earlier) revealed that the total subjective emotion score of the patients with borderline personality disorder was significantly higher than that of all other groups (p<0.001). Furthermore, the patients with borderline personality disorder rated themselves as experiencing significantly more sadness (p<0.001), fear (p<0.001), and disgust (p<0.005) than all other groups (Figure 4).

**Emotional change.** ANOVAs performed on each of the emotional change variables indicated that participants’ changes in sadness, anger, and fear (Figure 4) were significantly related to their group (change in sadness: F=5.12, df=2, 49, p<0.01; change in anger: F=4.29, df=2, 49, p<0.05; change in fear: F=3.84, df=2, 49, p<0.05). The healthy comparison subjects did not complete this questionnaire.

Post hoc analysis revealed that the sadness of the patients with borderline personality disorder increased significantly more than that of the patients with lesions outside the orbitofrontal cortex (p<0.01). There was a significant difference in the change in anger between the patients with borderline personality disorder and the patients with lesions outside the orbitofrontal cortex (p<0.01). The anger of patients with borderline personality disorder increased, and that of the patients with lesions outside the orbitofrontal cortex decreased. The patients with borderline personality disorder showed a significant difference in change in fear, compared with both the patients with lesions outside the orbitofrontal cortex (p<0.05) and those with orbitofrontal cortex lesions (p<0.01). The fear of the patients with borderline personality disorder increased, whereas that of the patients with lesions outside the orbitofrontal cortex and of the patients with orbitofrontal cortex lesions decreased. The patients with orbitofrontal cortex lesions did not differ significantly from those with lesions outside the orbitofrontal cortex on any measure.

**Probabilistic reversal.** ANOVAs indicated that the participants’ scores differed significantly by group (Figure 5) in terms of the total pounds accumulated by the 100th trial, total number of reversals achieved by 100 trials, and punishment insensitivity variables (total pounds: F=7.05, df=3, 89, p<0.001; total reversals: F=6.60, df=3, 89, p<0.001; punishment insensitivity: F=3.85, df=3, 89, p<0.05). (The King’s College neurosurgical patients were not included in the reversal analysis; their data are published elsewhere [4].)

The patients with orbitofrontal cortex lesions were markedly impaired on the reversal tasks, compared to all other groups. The patients with borderline personality disorder were not impaired on these tasks, compared to the other groups. The groups also differed in reward insensitivity (F=3.22, df=3, 89, p<0.05), and the patients with orbitofrontal cortex lesions were significantly more reward insensitive than the healthy comparison subjects (p=0.003) only when outliers were included.

**Time estimation.** ANOVAs for each time perception variable indicated that participants differed significantly by group in terms of total time estimation (F=3.45, df=3, 74, p<0.05) (Figure 5). Post hoc analysis revealed that the patients with orbitofrontal cortex lesions estimated that significantly more time had passed than did the healthy comparison subjects (p=0.002). Although the patients with orbitofrontal cortex lesions overestimated time, indicating a faster subjective sense of time, the healthy comparison subjects underestimated time, indicating a slower subjective sense of time. The patients with borderline personality disorder did not differ from the healthy comparison subjects on this variable.

**Spatial working memory**. Participants’ scores differed significantly by group on all three spatial working memory measures (Figure 5) (between errors: F=19.27, df=3, 77, p<0.001; within errors: F=3.65, df=3, 77, p<0.05; strategy errors: F=6.69, df=3, 77, p<0.001).

Both the patients with lesions outside the orbitofrontal cortex and those with orbitofrontal cortex lesions (most had dorsolateral prefrontal cortex damage) made more between errors than did the healthy comparison subjects and those with borderline personality disorder (p<0.001). The patients with lesions outside the orbitofrontal cortex made significantly more within errors than did the healthy participants and those with borderline personality disorder (p<0.01). There were no significant differences in within errors between the healthy comparison subjects, the patients with orbitofrontal cortex lesions, and the patients with borderline personality disorder. Finally, the patients with lesions outside the orbitofrontal cortex and those with orbitofrontal cortex lesions both had more strategy errors, compared with the healthy subjects (p<0.01 and p=0.001, respectively) and with patients with borderline personality disorder (p<0.05).

**Correlations**

To investigate the relationships between the different measures, Pearson's correlations (two-tailed) were performed across the data for all participants (Table 1). To compensate for the large number of comparisons, only the total score or main variable for each measure (all variables within each measure were significantly correlated with each other) was used in this analysis. A Bonferroni correction was applied, which resulted in the critical alpha level for any one correlation becoming 0.0005. Some correlations in which the alpha level was 0.01 are mentioned if they were deemed a priori to be interesting. (The emotional change questionnaire was not included in the correlation analysis because it was not administered to the healthy subjects.)
Discussion

This investigation used a relatively new approach of comparing the psychiatric population of patients with borderline personality disorder to neurological patients with orbitofrontal cortex brain lesions.

Patients with orbitofrontal cortex lesions and patients with borderline personality disorder performed similarly on some tests in that they were more impulsive and reported more inappropriate behaviors, more borderline personality disorder traits, more anger, and less happiness than subjects in both of the comparison groups (pa-
tients with prefrontal lesions outside the orbitofrontal cortex and healthy comparison subjects). They were also less open to experience and had a faster perception of time (in terms of time production) than the healthy comparison subjects. These findings suggest that orbitofrontal cortex functions may be related to these aspects of borderline personality disorder but not to the other borderline personality disorder characteristics that were measured, including levels of extraversion, conscientiousness, neuroticism, and emotion. Furthermore, patients with orbitofrontal cortex lesions but not patients with borderline personality disorder were impaired at reversing stimulus-reinforcer associations, suggesting that this orbitofrontal cortex dysfunction is not an essential part of borderline personality disorder.

Because the patients with borderline personality disorder were not impaired on the spatial working memory task and because the patients with lesions outside the orbitofrontal cortex and those with orbitofrontal cortex lesions did not perform poorly on the same tests, the neuropsychological deficits of the patients with borderline personality disorder and the patients with orbitofrontal cortex lesions could not be attributed to spatial working memory deficits or dorsolateral prefrontal cortex dysfunction.

**Similarities Between Patients With Orbitofrontal Cortex Lesions and Borderline Personality Disorder Patients**

**Impulsivity.** A major finding is that patients with orbitofrontal cortex lesions and patients with borderline personality disorder are similar in that they are both significantly
more impulsive, as assessed with both behavioral and self-report measures, than healthy subjects and patients with prefrontal lesions outside the orbitofrontal cortex. Although it is well established that patients with borderline personality disorder are impulsive (impulsivity is one of the diagnostic criteria), the direct comparison with findings in patients with orbitofrontal cortex damage shows that patients with orbitofrontal cortex lesions are as impulsive as patients with borderline personality disorder. This finding suggests that this aspect of borderline personality disorder could be related to orbitofrontal cortex function.

It is interesting to note that both the patients with borderline personality disorder and those with orbitofrontal cortex lesions had significantly lower time latencies on the behavioral impulsivity task, compared with the healthy subjects. This result may be related to a desire to complete the task fast, combined with a lack of sensitivity to punishment in the patients with orbitofrontal cortex lesions and perhaps with the desire for the reward of finishing sooner in patients with borderline personality disorder.

**Frontal behaviors and borderline personality disorder characteristics.** Other important findings are that the patients with orbitofrontal cortex lesions performed similarly to the patients with borderline personality disorder on the borderline personality disorder questionnaire and that the patients with borderline personality disorder performed like the patients with orbitofrontal cortex lesions on the frontal behavior questionnaire. Support for a relationship between frontal behaviors and borderline personality disorder characteristics comes from the highly significant positive correlation between scores on the frontal behavior questionnaire and the borderline personality disorder questionnaire both within the healthy comparison group and across all participants (p<0.0005). Analysis of responses on individual frontal behavior questionnaire items (see reference 21) for the borderline personality disorder group showed that the correlations were highest (0.22–0.53) for the questions concerned with inappropriate behavior; being uncooperative, aggressive, abusive, angry, or irritable; not worrying about oneself; being listless; stopping to help others in need; stopping to think before acting or making a decision; and gambling and taking risks when gambling.

On both the frontal behavior questionnaire and the borderline personality disorder questionnaire, the scores of the patients with borderline personality disorder were significantly higher than those of the patients with orbitofrontal cortex lesions. It is interesting to note that the patients with borderline personality disorder experienced significantly more inappropriate frontal behaviors than the patients with orbitofrontal cortex lesions. The borderline personality disorder patients may have other dysfunctions (perhaps related to amygdala function), in addition to orbitofrontal cortex dysfunction, that make their emotional deficits and socially inappropriate behaviors more severe than those of patients with orbitofrontal cortex lesions.

**Subjective emotion.** The patients with orbitofrontal cortex lesions and those with borderline personality disorder both reported being significantly less happy and more angry than the patients without orbitofrontal cortex lesions and the healthy comparison subjects. This finding suggests a common underlying cause, namely orbitofrontal cortex dysfunction. In general, the borderline personality disorder group was more emotional than all other groups, and the orbitofrontal cortex group approached the emotional scores of the borderline personality disorder group in just these two aspects of emotion. Thus, although orbitofrontal cortex dysfunction could contribute to the changes in emotion in patients with borderline personality disorder, other factors and brain regions also appear to be involved.

**Time production.** Both the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder produced less time than the healthy comparison subjects. A common cause for these two groups could be a higher level of frustration in waiting for the time to elapse. This study supports the evidence that impulsivity and time perception are related (27, 28). The frustration in waiting and/or the faster cognitive tempo that may cause patients with orbitofrontal cortex lesions and patients with borderline personality disorder to underproduce time may also be related to some of the inappropriate social and emotional behaviors they display, as demonstrated by their higher scores on the borderline personality disorder, frontal behavior, and emotion questionnaires, compared to the healthy comparison subjects and the subjects with prefrontal cortex lesions outside the orbitofrontal cortex. In support of this relationship, lower levels of time production correlated with higher levels of frontal behaviors and lower levels of openness to experience across all subjects.

**Personality.** Both the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder were less open to experience than were the healthy comparison subjects. There was no significant difference in this measure between the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder.

**Differences Between Patients With Orbitofrontal Cortex Lesions and Borderline Personality Disorder Patients**

**Personality.** Patients with borderline personality disorder were more neurotic, less extraverted, and less conscientious than all other groups. Our evidence suggests a strong relationship between certain personality traits and emotions, which may be related to the same underlying neurological correlates that are affected in patients with borderline personality disorder. This interpretation is supported by the fact that the patients with borderline personality disorder had both personality and emotional abnormalities and by the strong correlations between total
subjective emotion scores, borderline personality disorder traits, and measures of neuroticism, conscientiousness, and extraversion across all subjects. Since the patients with orbitofrontal cortex lesions did not differ from the healthy comparison subjects in measures of neuroticism, conscientiousness, extraversion, and subjective emotion, some brain region other than the orbitofrontal cortex (perhaps the amygdala) may be related to the personality and emotional abnormalities of patients with borderline personality disorder (21).

It is interesting to note that the borderline personality disorder subjects reported a higher level of introversion, relative to the healthy comparison subjects, but that the patients with orbitofrontal cortex lesions did not. In borderline personality disorder patients, impulsivity (of the type measured here), interacting with introversion and high levels of emotionality, may cause self-harm, whereas self-harm may not develop in the patients with orbitofrontal cortex lesions, who are impulsive but not introverted or extremely emotional. In addition, the patients with orbitofrontal cortex lesions were not found to have an abnormally high level of extraversion, which, combined with impulsivity, could lead patients to harm others (29).

**Subjective emotion.** The patients with borderline personality disorder were more emotional than all other groups, consistent with the fact that a major criterion for the diagnosis of borderline personality disorder is emotional instability (DSM-IV). Conversely, the patients with orbitofrontal cortex lesions did not report being very emotional. These findings suggest that the higher level of emotionality of the patients with borderline personality disorder cannot be ascribed to orbitofrontal cortex dysfunction of the type produced by a lesion.

**Reversal.** Of all the study groups, only the patients with orbitofrontal cortex lesions had deficits on the reversal task (5), and punishment insensitivity was a prominent feature in this group. It may be that patients with borderline personality disorder, who are in some ways hyper-emotional, are actually more sensitive to reward and punishment and thus perform well on the punishment sensitivity task. A lower level of punishment sensitivity in the patients with orbitofrontal cortex lesions could lead to impulsivity because subjects with this characteristic may not care about the consequences of their actions. In borderline personality disorder subjects, increased sensitivity to punishment might make subjects more emotional, and the higher level of emotionality might then contribute to their impulsive behavior (30).

**Time estimation.** The results suggest that the same underlying brain dysfunction, perhaps related to higher levels of frustration in waiting for the time to elapse, may cause both patients with borderline personality disorder and patients with orbitofrontal cortex lesions to under-produce time, although a different dysfunction, perhaps related to a greater cognitive pace (5), may cause patients with orbitofrontal cortex lesions to overestimate time.

**Conclusions**

The results support the view that impulsivity, affective dysregulation, and personality abnormalities are core aspects of DSM-IV borderline personality disorder and that borderline personality disorder can be thought of in dimensional terms, with each patient having a unique neurobiological profile. Environmental and genetic factors may lead to brain alterations that result in specific presentations, such as self-harm (31). An implication of the findings is that new light may be shed on the etiology of borderline personality disorder by considering how the syndrome can be fractionated and how different brain systems, each with different functions, contribute to the different symptoms of borderline personality disorder. The similarities and differences found between the patients with borderline personality disorder and the patients with orbitofrontal cortex lesions may lead to a better understanding of the functions of the orbitofrontal cortex, which could have implications for rehabilitation (5, 21).

In support of the main hypothesis, the patients with borderline personality disorder had deficits that were similar to those of the patients with orbitofrontal cortex lesions and dissimilar to those of the patients with lesions outside the orbitofrontal cortex. One implication of this finding is that some of the core characteristics of borderline personality disorder, in particular impulsivity, are similar to the effects of orbitofrontal cortex damage, suggesting that orbitofrontal cortex dysfunction may contribute to some of the deficits in borderline personality disorder. On the other hand, the patients with orbitofrontal cortex lesions and the patients with borderline personality disorder performed differently on some tasks, suggesting that other characteristics of borderline personality disorder, such as high levels of emotionality and the borderline personality profile (high levels of neuroticism and introversion and a low level of conscientiousness), do not appear to be related to the type of dysfunctions produced by orbitofrontal cortex damage and are perhaps related to other brain systems (e.g., the limbic system). Patients with borderline personality disorder may have a neurochemical imbalance or a hyperactive/responsive amygdala (10), which is not present in patients with orbitofrontal cortex lesions and which may exacerbate the emotional and personality disturbances of borderline personality disorder.

The orbitofrontal cortex, with its extensive reciprocal connections with the amygdala (which is implicated in emotional behavior [32, 33]), may play a role in correcting/regulating emotional and behavioral responses (1, 3, 4, 24, 34). Limbic-orbitofrontal circuit dysfunction may be involved in borderline personality disorder, at least in a subgroup of patients (35). Borderline personality disorder may conceivably involve higher levels of limbic discharge,
lower levels of orbitofrontal cortex function, and/or hypoactive frontolimbic circuitry. The findings reported here suggest that the orbitofrontal cortex is involved with the impulsivity evident in patients with borderline personality disorder. Our findings relate well to the hypothesis that the amygdala and orbitofrontal cortex act as part of an integrated neural system, as well as alone, in guiding decision making and adaptive response selection based on stimulus-reinforcement associations (2, 10, 13, 32, 36).

In summary, patients with borderline personality disorder have some deficits that can be related to the functions performed by the orbitofrontal cortex. These deficits might be related to smaller volume of the orbitofrontal cortex or to lower levels of activity in the orbitofrontal cortex. Because patients with borderline personality disorder may be hypersensitive to reinforcers, positive feedback should be emphasized in rehabilitation. More studies are needed to investigate the involvement of the orbitofrontal cortex and amygdala in patients with borderline personality disorder, as these areas, in conjunction or in isolation, may contribute to some of the behavioral and emotional disturbances observed in these patients.

References

24. Hornak J, Rolls ET, Wade D: Face and voice expression identification in patients with emotional and behavioural changes fol-
Implications of the Changing Use of Hospitalization to Treat Anorexia Nervosa

Molly G. Willer, B.A.
Paul Thuras, Ph.D.
Scott J. Crow, M.D.

Objective: This study sought to compare inpatient hospitalization practices for treating anorexia nervosa at two time periods and to assess whether differences in practices affected treatment outcome.

Method: Hospital charts were examined for all patients with a diagnosis of anorexia nervosa and a body mass index ≤17.5 kg/m² at admission who were admitted to a university hospital during 1975–1980 (early) or 1990–1995 (later).

Results: The “later cohort” had significantly higher admission and discharge body mass indexes, shorter lengths of stay, higher rates of weight gain, and more readmissions. Predictors of rehospitalization for the later cohort were increased length of stay, prior hospitalizations, and more rapid weight gain.

Conclusions: This study suggests that the use of hospitalization to treat anorexia nervosa has changed over time and is associated with increased rates of readmission.

Anorexia nervosa is associated with serious medical complications and high mortality rates, and it often requires inpatient hospitalization. Studies have suggested that inpatient treatment may not be the most effective approach (1). However, other studies have suggested that there are characteristics of hospitalizations that make them more or less effective and that some of those are changing in potentially detrimental ways. One such study examined inpatient hospitalizations for eating disorders over time and found increasing first admissions, total admissions, and readmissions and decreased lengths of stay (2). Factors linked to poor treatment outcome include the following: long duration from illness onset to hospital admission (3); inadequate inpatient weight gain, particularly during first hospitalizations (3, 4); rapid inpatient weight gain; and periods of inpatient weight loss (5, 6). Conversely, a period of weight maintenance before discharge may improve outcome (6). The goals of this study were 1) to compare hospitalization practices for treating anorexia nervosa in the 1970s to those in the 1990s and 2) to assess whether any changes in hospitalization practices affected patient outcomes.

Method

Hospital charts were examined for all inpatients with eating disorders at the University of Minnesota Hospital and Clinic during 1975–1980 (early) (N=59) and 1990–1995 (later) (N=127). The subjects had a discharge diagnosis of anorexia nervosa and, to provide more diagnostic rigor, an admission body mass index ≤17.5 kg/m² (ICD-10). Data collected included age, age at onset of eating disorder, length of admission, admission and discharge weights, height, comorbid diagnoses, and number of prior and subsequent hospitalizations at the University of Minnesota Hospital and Clinic. Admission and discharge body mass indexes and average rate of weight gain (lb/day) were calculated. The two cohorts were compared with independent-group t tests (when necessary, equal variance was not assumed) and chi-square tests. The variables associated with rehospitalization were determined with forward stepwise conditional logistic regression. Institutional review board approval was obtained before the study was initiated.

Results

Demographic variables and comparisons between the cohorts are presented in Table 1. No demographic variables differed significantly between the cohorts, except for a longer duration from illness onset to treatment (t=–2.28, df=163.52, p<0.03) and higher rates of comorbid diagnoses, both in the “later cohort.” Over time, the body mass index at admission increased, and the discharge body mass index decreased; both cohorts were discharged while still meeting the weight criterion for anorexia nervosa. Finally, length of stay decreased significantly, whereas the rate of weight gain and subsequent hospitalizations increased significantly.

To develop a model of variables associated with readmission, regression analyses were performed separately for each cohort with the following variables included as potential predictors of outcome: length of stay, rate of weight gain, body mass index at admission, body mass index at discharge, number of comorbid diagnoses, and presence of prior hospitalizations. In the “early cohort,” no variables were significantly associated with readmission. In the later cohort, three variables were significantly associated with rehospitalization: increased length of stay (B=0.022, Wald χ²=5.91, odds ratio=1.02, 95% confidence interval [CI]=1.00–1.04, p<0.03), more rapid rate of weight gain (B=1.716, Wald χ²=4.76, odds ratio=5.56, 95% CI=1.19–26.03, p<0.03), and having prior hospitalizations (B=1.505, Wald χ²=13.15, odds ratio=4.51, 95% CI=2.00–10.16, p<0.001).

Discussion

This study suggests that hospitalization for anorexia nervosa changed between the 1970s and 1990s in ways
that affect outcome. Patients in the 1970s stayed in the hospital longer and gained weight more slowly, and although they left at a lower discharge body mass index, they were less likely to be rehospitalized. By the 1990s, a more rapid rate of weight gain, having even one prior hospitalization, and longer lengths of stay were all associated with rehospitalization.

The association between longer hospital stays and rehospitalization in the later cohort was unexpected. Wiseeman et al. (2) found that a decreasing length of stay was concomitant with an increase in managed care; this is likely the case for our group, too. Thus, our findings may be attributable in part to changes in payment, which require that patients have a more chronic condition before they are allowed coverage for longer inpatient stays. This hypothesis is supported by further analyses showing that longer length of stay is correlated with more prior hospitalizations (r=0.24, p=0.007) and a lower body mass index at admission (r=–0.36, p<0.001) but not the rate of weight gain or the number of comorbid diagnoses. Given the very high likelihood of readmission with prior hospitalizations, requiring more chronic anorexia nervosa before lengthier inpatient treatment, rather than intensive treatment when anorexia nervosa is first problematic, may not be the most cost-effective approach (7).

In contrast to these results, one study found that slower weight gain is a negative outcome predictor (8). The authors suggested that the effect of the rate of weight gain depends on the patient’s attitude toward treatment: if the patient is willing to gain weight, then rapid weight gain is a positive prognostic indicator; if the patient is resistant to treatment, slower weight gain is a negative indicator (8). Therefore, for patients willing to gain weight, more rapid weight gain may be adequate; for those unwilling to gain weight, the high caloric intake required to achieve normal weight (9) may require more gradual weight gain to address related psychological variables.

This study has several limitations. First, chart diagnoses were used rather than structured interviews. Therefore, a body mass index ≤17.5 kg/m² at admission was added as an inclusion criterion. Other limitations include the differing lengths of follow-up and treatment philosophy changing over time (from behavioral to more cognitive-behavior oriented). However, these limitations should bias the results toward more readmissions in the early cohort (the opposite of these results).

These findings suggest that it is critical to maximize recovery in the first hospitalization for anorexia nervosa to prevent readmission. It appears that for those who are amenable to weight gain, a shorter inpatient stay with more rapid weight gain may be adequate. For those who are less psychologically amenable to treatment, a longer length of stay, a less rapid rate of weight gain, and a period of weight maintenance before discharge may be a more effective approach.

Presented in part at the 2004 International Conference on Eating Disorders, Orlando, Fla., April 29 through May 2, 2004. Received July 28, 2004; revision received Nov. 4, 2004; accepted Feb. 1, 2005. From the Departments of Psychiatry and Psychology, University of Minnesota. Address correspondence and reprint requests to Dr. Crow, Department of Psychiatry, University of Minnesota, F290/2A West, 2450 Riverside Ave., Minneapolis, MN 55454; crowx002@umn.edu (e-mail).

References

<table>
<thead>
<tr>
<th>TABLE 1. Demographic and Clinical Characteristics of Patients With Anorexia Nervosa and a Body Mass Index ≤17.5 kg/m² at Admission to the Hospital During 1975–1980 or 1990–1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age at onset (years)</td>
</tr>
<tr>
<td>Age at first hospitalization at University of Minnesota Hospital and Clinic (years)</td>
</tr>
<tr>
<td>Admission body mass index (kg/m²)</td>
</tr>
<tr>
<td>Discharge body mass index (kg/m²)</td>
</tr>
<tr>
<td>Number of prior hospitalizations</td>
</tr>
<tr>
<td>Length of stay (days)</td>
</tr>
<tr>
<td>Rate of weight gain (lb/day)</td>
</tr>
<tr>
<td>Number of subsequent hospitalizations</td>
</tr>
</tbody>
</table>

| Comorbid diagnoses | Early Cohort | Later Cohort | χ² | df | p |
|                   | Mean | % | Mean | % |       |
| Mood disorder | 24 | 43 | 78 | 61 | –2.33 | 184 | <0.03 |
| Anxiety disorder | 15 | 25 | 61 | 48 | 8.52 | 1 | 0.004 |
| Substance use disorder | 6 | 10 | 6 | 5 | 1.98 | 1 | 0.14 |
| Other disorder | 4 | 7 | 15 | 12 | 1.11 | 1 | 0.22 |

a Had a skewed distribution.
b Presence of one or more comorbid diagnoses.
The Acute Antipanic Activity of Aerobic Exercise

Andreas Ströhle, M.D.
Christian Feller
Marlies Onken, M.D.
Frank Godemann, M.D.
Andreas Heinz, M.D.
Fernando Dimeo, M.D.

Objective: Regular physical activity is anxiolytic for both healthy subjects and patients with panic disorder. However, the acute antipanic activity of exercise has not yet been studied systematically.

Method: The effects of quiet rest or aerobic treadmill exercise (30 minutes at 70% of maximum oxygen consumption) on cholecystokinin tetrapeptide (CCK-4)-induced panic attacks were studied in a crossover design in 15 healthy subjects. The effects were measured with the Acute Panic Inventory.

Results: Panic attacks occurred in 12 subjects after rest but in only six subjects after exercise. In both conditions, CCK-4 administration was followed by a significant increase in Acute Panic Inventory scores; however, prior exercise resulted in significantly lower scores than quiet rest.

Conclusions: Aerobic exercise has an acute antipanic activity in healthy subjects. If the authors’ results are confirmed in patients, the optimum intensity and duration of acute exercise for achieving antipanic effects will have to be characterized.


The anxiolytic effects of training in aerobic exercise have been described in numerous studies in healthy volunteers (1). In line with this, case reports (2) and two clinical studies have suggested that exercise training may be used therapeutically in patients with anxiety neurosis (3) and panic disorder (4). In addition, studies in healthy subjects and two case reports (5, 6) have suggested that an acute bout of exercise is anxiolytic as well. In contrast, a single bout of exercise may induce acute panic attacks (4, 7) or increase subjective anxiety in patients with panic disorder more than in other people (8). To our knowledge, the potential acute antipanic activity of exercise has not been studied systematically.

Acute bouts of exercise have been shown to reduce symptoms induced by CO₂ and caffeine (9, 10). However, CO₂ and caffeine primarily induce somatic symptoms of anxiety in healthy subjects but rarely panic attacks. In contrast, cholecystokinin tetrapeptide (CCK-4) dose dependently and reliably induces panic attacks in healthy comparison subjects (11). Therefore, we studied the potential acute antipanic effects of aerobic exercise in healthy volunteers after CCK-4 administration and compared CCK-4-induced panic attacks with and without previous exercise.

Method

Fifteen healthy subjects (six women and nine men; mean age=26.4 years, SD=3.8) with no personal (12) or family history of a psychiatric disorder were recruited for the study. The subjects did not take any medication and had undergone a thorough medical examination to rule out other illnesses, drug intake, and lifestyles (sleep deprivation, changes in the sleep-wake cycle) that could interfere with the study. The protocol was approved by the local ethics committee for human experiments. After a complete description of the study to the subjects, written informed consent was obtained. The subjects were paid for their participation. Before the experiment, a treadmill spiroergometry was performed to determine the subjects’ maximum oxygen consumption. Subjects were excluded if they were too physically fit (maximum oxy-
The Acute Antipanic Activity of Aerobic Exercise

Andreas Ströhle, M.D.
Christian Feller
Marlies Onken, M.D.
Frank Godemann, M.D.
Andreas Heinz, M.D.
Fernando Dimeo, M.D.

Objective: Regular physical activity is anxiolytic for both healthy subjects and patients with panic disorder. However, the acute antipanic activity of exercise has not yet been studied systematically.

Method: The effects of quiet rest or aerobic treadmill exercise (30 minutes at 70% of maximum oxygen consumption) on cholecystokinin tetrapeptide (CCK-4)-induced panic attacks were studied in a crossover design in 15 healthy subjects. The effects were measured with the Acute Panic Inventory.

Results: Panic attacks occurred in 12 subjects after rest but in only six subjects after exercise. In both conditions, CCK-4 administration was followed by a significant increase in Acute Panic Inventory scores; however, prior exercise resulted in significantly lower scores than quiet rest.

Conclusions: Aerobic exercise has an acute antipanic activity in healthy subjects. If the authors’ results are confirmed in patients, the optimum intensity and duration of acute exercise for achieving antipanic effects will have to be characterized.

The anxiolytic effects of training in aerobic exercise have been described in numerous studies in healthy volunteers (1). In line with this, case reports (2) and two clinical studies have suggested that exercise training may be used therapeutically in patients with anxiety neurosis (3) and panic disorder (4). In addition, studies in healthy subjects and two case reports (5, 6) have suggested that an acute bout of exercise is anxiolytic as well. In contrast, exercise may induce acute panic attacks (4, 7) or increase subjective anxiety in patients with panic disorder more than in other people (8). To our knowledge, the potential acute antipanic activity of exercise has not been studied systematically.

Acute bouts of exercise have been shown to reduce symptoms induced by CO₂ and caffeine (9, 10). However, CO₂ and caffeine primarily induce somatic symptoms of anxiety in healthy subjects but rarely panic attacks. In contrast, cholecystokinin tetrapeptide (CCK-4) dose dependently and reliably induces panic attacks in healthy comparison subjects (11). Therefore, we studied the potential acute antipanic effects of aerobic exercise in healthy volunteers after CCK-4 administration and compared CCK-4-induced panic attacks with and without previous exercise.

Method

Fifteen healthy subjects (six women and nine men; mean age=26.4 years, SD=3.8) with no personal (12) or family history of a psychiatric disorder were recruited for the study. The subjects did not take any medication and had undergone a thorough medical examination to rule out other illnesses, drug intake, and lifestyles (sleep deprivation, changes in the sleep-wake cycle) that could interfere with the study. The protocol was approved by the local ethics committee for human experiments. After a complete description of the study to the subjects, written informed consent was obtained. The subjects were paid for their participation. Before the experiment, a treadmill spiroergometry was performed to determine the subjects’ maximum oxygen consumption. Subjects were excluded if they were too physically fit (maximum oxy-
gen consumption >55 ml/kg/min) or had signs of cardiovascular abnormalities during the ergometry. The subjects were told that the aim of study was to characterize the behavioral effects of exercise on CCK-4-induced symptoms.

All subjects were studied twice on different days separated by 1 week. From 9:30 a.m. to 12:00 noon, the subjects were studied in a supine position in a soundproof room with a single bed. On the first day, the subjects were randomly assigned to one of two conditions: quiet rest or exercise (treadmill walking for 30 minutes at 70% of maximum oxygen consumption from 10:20 to 10:50 a.m.); the other condition was used on the second study day. Each subject received a bolus injection of 50 µg of CCK-4 (Clinalfa, Laufelfingen, Switzerland) dissolved in 2 ml of 0.9% saline solution at 11:00 a.m. The Acute Panic Inventory (13) was administered at 10:00 a.m. (baseline), 10:55 a.m. (after rest or exercise), and 11:05 a.m. (after CCK-4 administration) by a rater blind to the procedure (A.S. or M.O.); the maximum intensity during the observation period was noted. To further characterize the possible effects of exercise, we used a previously described anxiety subscore (afraid of dying, general fear) (14, 15) and a somatic subscore (palpitations, breathing, and nausea) of the Acute Panic Inventory. A panic attack was noted when the Acute Panic Inventory total score exceeded 20 and an increment of at least 14 points over the -100 score was observed (13, 14, 16). The occurrence of a panic attack was confirmed by a trained psychiatrist. On the second study day, the subjects were additionally asked whether exercise reduced or increased CCK-4-induced symptoms or whether it had no influence at all.

The frequencies of panic attacks in the two treatment conditions were compared by using the McNemar paired chi-square test. The subjective experiences of whether exercise decreased, increased, or had no effect on CCK-4-induced symptoms were compared by using the chi-square test. For statistical comparison of mean Acute Panic Inventory scores and the subscores between the rest and the exercise conditions and among the three time points (10:00 a.m., 10:55 a.m., and 11:05 a.m.), a two-factor multivariate analysis of variance (MANOVA) with a repeated-measures design was performed. Treatment and time were the two within-subjects factors, with two or three levels. Hypothesis tests for the main and interaction effects of these factors were based on multivariate criteria, such as Wilk’s lambda and its approximated F value. When significant main or interaction effects were found, univariate F tests and/or tests with contrasts followed by MANOVA for investigating simple effects were performed. As a nominal level of significance, alpha=0.05 was accepted. To keep the type I error at ≤0.05, all post hoc tests (univariate F tests and tests with contrasts) were performed at a reduced level of significance (an adjusted alpha according to the Bonferroni procedure).

Results

The rates of panic attacks were significantly higher in the quiet rest than in the exercise conditions (12 of 15 versus six of 15) (p=0.03, McNemar’s test) (Table 1). Ten subjects estimated that exercise reduced CCK-4-induced symptoms, three did not experience a difference, and two subjects estimated that exercise increased CCK-4-induced symptoms (χ²=7.60, df=2, p=0.02).

MANOVAs of the Acute Panic Inventory score and the subscores revealed significant main effects of time (F=15.60, df=6, 52, p=0.001) and a significant interaction effect of time by exercise (F=3.11, df=6, 52, p=0.01). Univariate F tests indicated that the interaction of time by exercise was significant for the overall Acute Panic Inventory score (F=6.73, df=2, 28, p=0.01) and the anxiety subscale score (F=8.67, df=2, 28, p=0.007) and nearly significant for the somatic symptoms subscore (F=3.04, df=2, 28, p=0.09). When analyzing the simple effects of time on the Acute Panic Inventory score and the anxiety subscales, i.e., the differences of the last variables between the three time points, we found that CCK-4 increased the Acute Panic Inventory score and the anxiety subscore score compared to baseline (tests with contrasts in MANOVA, p<0.05). Compared to quiet rest, exercise significantly reduced the CCK-4-induced increase in the total Acute Panic Inventory score and the anxiety subscore (tests with contrasts in MANOVA, p<0.05).

Discussion

The main finding of this exploratory study was that 30 minutes of aerobic exercise had an acute antipanic activity. In healthy volunteers, the frequency of panic attacks and CCK-4-induced symptoms was significantly decreased after exercise. Subscore analysis of the Acute Panic Inventory scores indicated that exercise decreased CCK-4-induced anxiety symptoms rather than somatic symptoms. Additionally, the subjective experience of CCK-4-induced symptoms was decreased after exercise in most subjects.

Exercise has been shown to reduce somatic symptoms of anxiety induced by CO₂ (9) and caffeine (10) in healthy volunteers. Our results for the first time suggest that exercise has an acute antipanic activity. Although it has already been demonstrated that exercise training for 10 weeks is anxiolytic in patients with panic disorder (4), our results suggest that acute exercise may have antipanic effects. If these effects are confirmed in patients with panic disorder, acute exercise may be used in the treatment of panic disorder as well. The optimum intensity and duration of acute exercise in achieving immediate antipanic effects will then have to be characterized in patients with panic disorder after prior exercise testing.
Numerous psychological and physiological mechanisms have been suggested to explain the beneficial effects of exercise on anxiety and depression (for a review, see reference 17). Experimental evidence suggests that the effects of exercise are on the synthesis and metabolism of monoamines. In addition, activation of β-endorphin secretion has been linked to the behavioral effects of exercise. The hypothalamic pituitary adrenocortical system (18) and its modulators—for example, atrial natriuretic peptide (16)—may be further involved in the antipanic effects of exercise.

In summary, our findings give evidence for an acute antipanic activity of exercise. If confirmed in patients, our results suggest that acute exercise may be used in the treatment of panic disorder as well. However, details about the optimum dosage (duration and intensity) remain to be characterized in patients.

Received Aug. 29, 2004; revision received Nov. 10, 2004; accepted Jan. 10, 2005. From the Department of Psychiatry and Psychotherapy, Campus Charité Mitte, and the Section of Sports Medicine, Department of Internal Medicine III, Campus Benjamin Franklin, Charité-University Medicine Berlin. Address correspondence and reprint requests to Dr. Ströhle, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité-University Medicine Berlin, Schumannstr. 20/21, 10117 Berlin, Germany; andreas.stroehle@charite.de (e-mail).

The authors thank Monika Hantschke-Brüggemann and Nancy Bock for technical assistance and Alexander Yassouridis, Ph.D., for statistical advice.

References

Open-Label Trial of Riluzole in Generalized Anxiety Disorder

Sanjay J. Mathew, M.D.
Jonathan M. Amiel, B.S.
Jeremy D. Coplan, M.D.
Heidi A. Fitterling, B.A.
Harold A. Sackeim, Ph.D.
Jack M. Gorman, M.D.

Objective: There is a need to identify novel pharmacotherapies for anxiety disorders. The authors examined the safety and efficacy of riluzole, an antiglutamatergic agent, in adult outpatients with generalized anxiety disorder.

Method: In an 8-week, open-label, fixed-dose study, 18 medically healthy patients with DSM-IV generalized anxiety disorder received treatment with riluzole (100 mg/day) following a 2-week drug-free period. The primary efficacy measure was the Hamilton Anxiety Rating Scale (HAM-A) score at endpoint.

Results: Twelve of the 15 patients who completed the trial responded positively to riluzole. At 8 weeks, eight of the 15 patients had HAM-A score indicating remission of their anxiety. The median time to response was 2.5 weeks.

Conclusions: Riluzole appears to be an effective, well-tolerated, and rapidly acting anxiolytic medication for some patients with generalized anxiety disorder. Larger, placebo-controlled studies are indicated.


There is a need to identify novel therapeutic targets for the anxiety disorders (1). Recent studies in generalized anxiety disorder have investigated nonbenzodiazepine anxiolytic medications that primarily modulate the γ-aminobutyric acid (GABA) system (2, 3), but the role of the excitatory amino acid glutamate has been less scrutinized. Preclinical and clinical studies suggest abnormalities in glutamate regulation in anxiety disorders and the anxiolytic potential for antiglutamatergic agents. Nonhuman primate studies of early-life stress have shown increased glutamate activity in brain regions implicated in anxiety or fear responses (4). It has also been shown that antiglutamatergic agents suppress stress hormonal responses in nonhuman primates (5) and facilitate the extinction of conditioned fear in rodents (6).

Riluzole, a presynaptic glutamate release inhibitor approved by the U.S. Food and Drug Administration for the treatment of amyotrophic lateral sclerosis, was recently shown to have antidepressant and anxiolytic properties in treatment-resistant major depression (7, 8). To test the idea that an antiglutamatergic medication would be effective for primary anxiety disorders, we conducted an open-label, proof-of-concept investigation of riluzole in patients with generalized anxiety disorder.

Method

Eighteen adult outpatients (six men, 12 women; mean age=33.6 years, SD=9.4) were recruited through media advertisement or flyers. All patients met DSM-IV criteria for a current diagnosis of generalized anxiety disorder, confirmed by an independent evaluator using the Structured Clinical Interview for DSM-IV. Mean duration of illness was 14.5 years (SD=9.6). Comorbid conditions included panic disorder (N=7), dysthymia (N=6), social anxiety disorder (N=4), and specific phobia (N=2). Exclusion criteria consisted of major depressive episode or substance abuse/dependence (other than nicotine) within 6 months of study entry; lifetime history of psychotic, bipolar, obsessive-compulsive, post-traumatic stress, or eating disorder; mental retardation or learning disability; autism; or substantial medical or neurological conditions requiring medication.

Five patients had never received psychotropic medication, and all patients had been free of psychotropic medication for at least 2 weeks before beginning treatment with riluzole. All participants had unremarkable screening laboratory evaluations, including urine toxicology. At screening, the patients’ mean score on the Hamilton Anxiety Rating Scale (HAM-A) was 21.2 (SD=3.1).

Before riluzole treatment was initiated at the baseline visit (approximately 1–2 weeks after screening), psychometric evaluations and urine toxicology were repeated. The first dose of riluzole was administered at 50 mg/day, and subsequent doses were fixed at 50 mg b.i.d. for the remainder of the study. Patients were seen each week for the first month and biweekly for the second month for rating scales, medication compliance check, and adverse events monitoring. No psychotherapy or additional psychotropic medications were allowed during the study period.

At endpoint (week 8), patients underwent full psychiatric and medical assessment. A trained independent evaluator who was aware of patients’ medication status but masked to visit week administered rating scales with the study psychiatrist (S.J.M.). Response was defined as a 50% decrease in HAM-A score from baseline, and remission was defined as a HAM-A score of ≤7. Secondary outcome measures were the 24-item Hamilton Depression Rating Scale (HAM-D), the self-report Anxiety Sensitivity Index (9), and, for patients with comorbid panic, the Panic Disorder Severity Scale (10).

All statistical tests were two-tailed, and the significance level was set at p≤0.05 throughout. After complete description of the study to the subjects, written informed consent was obtained. The study was approved by the New York State Psychiatric Institute Institutional Review Board.

Results

A total of 87 subjects were screened for this study; the final group included 18 eligible patients. Of the 18 patients who took at least one dose of riluzole, 15 (83%) completed the 8-week trial. Two patients dropped out in the first week.
due to adverse events (dizziness and nausea in one patient, cognitive slowing in the second), and a third patient was taken out of the study because of a protocol violation. Because all three dropouts occurred before the first post-baseline assessment, no analyses of their efficacy data were conducted.

Twelve of the patients had responded positively to the drug by week 8; these patients represented 67% of all patients (intent-to-treat group) and 80% of the patients who completed the trial. Eight patients were in remission at the end of 8 weeks, representing 44% of all patients and 53% of those who completed the trial. The mean HAM-A score of the 15 patients who completed the trial decreased from 20.0 (SD=3.4) at baseline to 7.5 (SD=5.3) at week 8 (Figure 1). Mean scores on the psychic subscale of the HAM-A improved from 13.0 (SD=2.5) to 5.0 (SD=4.0) (paired t=7.42, df=14, p<0.001), and mean scores on the somatic subscale improved from 5.0 (SD=4.0) to 2.5 (SD=1.6) (paired t=5.60, df=14, p<0.001). Median time to response was 2.5 weeks (range=1–6). Analysis of variance for patients who completed the trial indicated significant improvement in HAM-A score across the duration of the trial (F=17.34, df=1, 7, p<0.0001) (Figure 1). Dunnett’s post hoc analysis showed that improvement occurred after 1 week of treatment and persisted throughout the study (p<0.05 compared with baseline at each time point).

Mean Anxiety Sensitivity Index scores decreased from 29.8 (SD=10.4) to 19.4 (SD=8.9) (paired t=4.40, df=14, p<0.001) (Figure 1), and HAM-D scores decreased from a mean of 15.3 (SD=4.4) to a mean of 9.5 (SD=7.3) at endpoint (paired t=3.63, df=14, p=0.003). Age, gender, psychiatric comorbidity (including presence of panic attacks or disorder), or previous pharmacotherapy did not differen-tiate patients whose anxiety did or did not remit. Of the seven patients with comorbid panic disorder, two discontinued because of side effects, three were remitters according to HAM-A criteria, and two were nonremitters by HAM-A criteria. All five completers experienced marked reductions in panic symptoms, from a baseline Panic Disorder Severity Scale mean of 10.4 (SD=4.92) to an endpoint mean of 1.40 (SD=1.14) (paired t=4.44, df=4, p=0.01).

The most common adverse events during the trial were insomnia/sleep disturbance (three patients [22%]), nausea/abdominal discomfort (three patients [22%]), sedation (two patients [11%]), and dry mouth (two patients [11%]). No serious adverse events were noted. One patient who completed the trial required a dose reduction of riluzole to 50 mg/day for the final 4 weeks due to sedation. Three patients exhibited a transient increase in alanine aminotransferase ranging from 1.1 to 1.8 times the upper normal limit in week 4 (N=1) or week 8 (N=2). The aminotransferase values of the patient with the earliest detected elevation normalized within 2 weeks while receiving riluzole and remained normal at endpoint; for the remaining two patients, aminotransferase values normalized after discontinuation of riluzole. No patient exhibited symptoms of hepatic toxicity.

Discussion

In this open-label pilot study, riluzole at a fixed dose of 100 mg/day was associated with rapid and sustained anxiolytic effects in patients with generalized anxiety disorder, with favorable tolerability. We did not observe substantial elevations of liver function test values in this patient group, but data from other clinical trials (11) suggest that liver function monitoring is warranted. Clinician-rated psychic and somatic aspects of anxiety as well as patient-reported measures of anxiety sensitivity were significantly improved. Remission and response rates exceeded those achieved by other nonbenzodiazepine medications (2, 3) and may have been even greater at the higher doses used by Zarate et al. (7) for treatment-resistant depression. Important limitations of the current study include the small number of subjects and lack of placebo control; the outcome might differ in a controlled design.

Although our results must be considered preliminary because of the open-label design, they offer further evidence in support of the role of the glutamatergic system in anxiety disorders and the anxiolytic properties of anti-glutamatergic agents. Riluzole was recently observed to have beneficial effects in a patient with obsessive-compulsive disorder (12), and additional studies in other anxiety and mood disorders are ongoing. It is notable that Panic Disorder Severity Scale scores uniformly and significantly decreased not only in patients whose anxiety remitted but also in those whose anxiety did not remit. This suggests potential efficacy of riluzole in panic disorder. Larger controlled trials are warranted in generalized anxi-
ety disorder and related anxiety disorders such as panic disorder and social anxiety disorder.

Received Aug. 12, 2004; revision received Nov. 23, 2004; accepted Dec. 10, 2004. From the Department of Biological Psychiatry, New York State Psychiatric Institute; the Department of Psychiatry and Department of Radiology, Columbia University College of Physicians and Surgeons, New York; the Department of Psychiatry, Downstate Medical Center, Brooklyn, N.Y.; and the Department of Psychiatry and the Department of Neuroscience, Mount Sinai School of Medicine, New York. Address reprint requests to Dr. Mathew, One Gustave Levy Place, Box 1218, New York, NY 10029; sanjay.mathew@mssm.edu (e-mail).

Supported in part by the National Alliance for Research in Schizophrenia and Depression Young Investigator’s Award and by the Frontier Fund of Columbia University. Dr. Gorman has received consulting fees unrelated to riluzole from Aventis, the manufacturer of riluzole. Aventis was not involved in any way with this study.

The authors thank Eric Smith, Ph.D., Ezekiel Young, M.A., Lili A. Bernstein, B.A., Kathryn A. Keegan, B.A., Manju Chacko, B.A., Beena Alex, B.A., and the clinical staff at the New York State Psychiatric Institute for their valuable contributions.

References
Do Patients With Schizophrenia Wish to Be Involved in Decisions About Their Medical Treatment?

Johannes Hamann, M.D.
Rudolf Cohen, Ph.D.
Stefan Leucht, M.D.
Raymonde Busch, Dipl.-Math.
Werner Kissling, M.D.

Objective: Little is known about the desire of patients with schizophrenia to be involved in medical decisions affecting their treatment.

Method: The authors administered the Autonomy Preference Index to 122 inpatients with schizophrenia. In addition, the patients filled out the Drug Attitude Inventory. Sociodemographic data and Positive and Negative Syndrome Scale ratings were available for all patients.

Results: The patients expressed a desire for shared decision making that was slightly greater than that of primary care patients. Negative attitudes toward medical treatment and younger age were associated with a higher desire for participation.

Conclusions: It is important to meet the participation needs of patients who are dissatisfied with their psychiatric treatment.

(Translaser 2005; 162:2382–2384)

The participation of patients in medical decisions affecting their treatment is increasingly being advocated in the field of mental health. Treatment guidelines for schizophrenia recommend basing decisions concerning patients' medication on their preferences and on informed consent; e.g., the National Institute for Clinical Excellence (1) states, “The choice of antipsychotic drug should be made jointly by the individual and the clinician responsible for treatment based on an informed discussion of the relative benefits of the drugs and their side-effect profiles.” However, little is known as to whether and to what extent patients with schizophrenia want to be involved in making medical decisions and which variables influence this desire. In the present study, the patients’ desire to be involved in medical decisions about their treatment and potential determinants that help to identify patients with a high desire for participation were studied.

Method

This survey reports on baseline data from a current intervention study on shared decision making for inpatients with schizophrenia. This study examines the influence of a decision-aid program for choosing antipsychotic drugs based on patients’ satisfaction, compliance, and rates of relapse. Recruitment was done in 2003–2004 in two German state hospitals and one university hospital. Broad inclusion criteria were used. All patients between 18 and 65 years of age with a diagnosis of schizophrenia (ICD-10, F20/F23) who spoke German fluently and who had no diagnosis of mental retardation were eligible for participation in the study. The patients were recruited for the study as soon as their physicians rated them on the conceptual disorganization item of the Positive and Negative Syndrome Scale for schizophrenia (2) as <5.

To measure the patients’ desire to participate, we used the decision-making preference subscale of the Autonomy Preference Index (3), a six-item self-report instrument devised to measure the patients’ general wish to participate in making medical decisions (e.g., “The important medical decisions should be made by your doctor, not by you”). Autonomy Preference Index scores range from 0 to 100, with the lowest score (0) corresponding to a desire that the doctor take complete control over medical decisions (paternalistic role of the doctor), whereas a score of 100 means that the patient wants to make all decisions on his or her own (informed choice). Intermediate scores reflect a desire for decision making that is shared equally between doctor and patient (3). The Autonomy Preference Index was validated and has been used in U.S. primary care patient populations (3, 4). Internal consistency was reported as alpha=0.82 and test-retest reliability as r=0.84 (3).

We analyzed the following variables of interest that might determine patients’ interest in participating in decision making regarding their treatment:

1. Factors that are known from the literature to influence patients’ desire to participate (age, gender, education) (3, 4)
2. Indicators for the chronicity of illness (duration of the illness, number of hospitalizations)
3. The patients’ experiences (knowledge about the disease, according to a seven-item questionnaire, experiences with present or previous involuntary treatment)
4. The patients’ attitudes toward treatment (Drug Attitude Inventory, short version [5]). This scale includes 10 questions on the subjective response to neuroleptic medication, including both positive drug effects (e.g., prevention of relapses) and negative ones (e.g., dysphoric reactions). Higher scores (in a range of 0 to 10) reflect more positive attitudes.
5. Psychopathology (the Positive and Negative Syndrome Scale for schizophrenia [2] total score)

The Autonomy Preference Index score was taken as the dependent variable. First, explorative univariate analyses were carried out for Autonomy Preference Index scores and variables of interest (t tests and Pearson’s correlations [r]). All variables were then entered into a linear regression. A two-sided p<0.05 was considered significant.

Results

Data from 122 patients were included in the analysis. There were 60 women and 62 men (age: mean=37.6 years, SD=12.1; duration of illness: mean=9.4 years, SD=9.0).
Thirty-one patients (26%) were admitted for the first time to a psychiatric hospital; 34 patients (28%) were hospitalized involuntarily. The patients were recruited for the study about 2 weeks after admission to the hospital (mean=16.3 days, SD=20.2, median=9.0). The patients’ desire for participation (Autonomy Preference Index scores) varied between 0 and 91 points (mean=46.6, SD=18.3).

Exploratory tests found no significant correlations between Autonomy Preference Index scores and the variables studied (r=−0.15, p=0.16), with the exception of the patients’ attitudes toward treatment (Drug Attitude Inventory) (r=−0.32, p=0.002). Since intercorrelations between the variables had to be assumed, a linear regression analysis was performed, showing that Drug Attitude Inventory scores and the patients’ ages significantly (R²=0.42) predicted Autonomy Preference Index scores (Table 1) when all other variables were accounted for. There tended to be more interest in participation for patients with the experience of involuntary treatment.

Discussion

The patients with the experience of involuntary treatment, those who reported negative attitudes toward medication (according to the Drug Attitude Inventory), and younger patients expressed a higher desire to participate in decision making about their care.

The patients treated involuntarily may be those who deny their illness and want to make autonomous decisions about whether psychiatric treatment as a whole should take place or not. Meeting these patients’ desire to participate is difficult because of legal constraints. However, including these patients in treatment decisions as much as possible may be especially worthwhile because such involvement may improve their attitude toward treatment and thus enhance compliance.

Patients who express their dissatisfaction with their drug treatment, according to Drug Attitude Inventory, might often have legitimate complaints (e.g., dysphoric reactions, side effects) that had up until then been ignored by their physicians for one reason or another. The desire of these patients to participate in treatment decisions should be respected because patients who are not satisfied with treatment are likely to discontinue it. Incorporation of these patients’ views and experiences (e.g., side effects) to a greater extent presumably improves treatment adherence by emphasizing tolerability and consideration of other treatment options (e.g., a drug with a different side effect profile).

One way of including patients in their treatment decisions might be the model of shared decision making (6), which aims to decrease the informational and power asymmetry between doctors and patients by increasing the patients’ information and control over treatment decisions that affect their well-being (7).

<table>
<thead>
<tr>
<th>Variable</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>−2.19</td>
<td>0.04</td>
</tr>
<tr>
<td>Gender</td>
<td>1.00</td>
<td>0.32</td>
</tr>
<tr>
<td>Education</td>
<td>1.47</td>
<td>0.15</td>
</tr>
<tr>
<td>Duration of illness</td>
<td>−0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Number of previous hospitalizations</td>
<td>0.75</td>
<td>0.46</td>
</tr>
<tr>
<td>Knowledge about schizophrenia</td>
<td>−1.73</td>
<td>0.09</td>
</tr>
<tr>
<td>Involuntary admission to present hospitalization</td>
<td>−0.85</td>
<td>0.40</td>
</tr>
<tr>
<td>Previous involuntary treatment</td>
<td>1.96</td>
<td>0.06</td>
</tr>
<tr>
<td>Severity of illness (Positive and Negative Syndrome Scale)</td>
<td>−1.13</td>
<td>0.27</td>
</tr>
<tr>
<td>Attitudes toward medication</td>
<td>−2.50</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Overall, the Autonomy Preference Index scores of inpatients with schizophrenia (mean=46, SD=18) were slightly higher than those reported for primary care patients (mean=33–42) (3, 4). One reason might be the relative youth of the patients in our group because younger age was associated with a higher interest in participation. The fact that schizophrenia is a severe chronic disease and has great influence on the patients’ life is another possible reason why patients are particularly keen on participation.

The mean Autonomy Preference Index score of 46 for inpatients reflects that there is no interest on the patients’ part to take over decisional control completely but rather that patients strongly wish to participate in medical decisions on an equal footing with their doctors.

One limitation of our study was the low internal consistency of the Autonomy Preference Index compared to the validation study (alpha=0.57 versus 0.82, respectively). The reason for this low reliability might be that the Autonomy Preference Index score was determined in nonpsychiatric settings. When we removed the two items that were reversely coded, internal consistency for the remaining four-item scale was satisfactory (alpha=0.71), and results for linear regression analysis were similar (Drug Attitude Inventory scores and patients’ age significantly predicted Autonomy Preference Index scores).

Finally, it is important to note that in the preceding studies (3, 4), the patients’ desire for information was markedly higher than their desire for participation. Although this was not measured in our study, it might be true of patients with schizophrenia as well and imply that we should adequately inform all patients of their treatment and the various treatment options even when they show no interest in influencing the decisions.

This first approach to studying the desire for participation of inpatients with schizophrenia shows that a wish for active involvement exists and is partly explained by negative attitudes toward neuroleptic medication. Further research on patient information and participation needs, as well as on strategies to meet these needs, is urgently re-
Evidence for Sensory Prediction Deficits in Schizophrenia

Sukhwinder S. Shergill, M.D., Ph.D.
Gabrielle Samson, B.Sc.
Paul M. Bays, M.A.
Chris D. Frith, Ph.D.
Daniel M. Wolpert, M.D., Ph.D.

Objective: Patients with schizophrenia experiencing delusions and hallucinations can misattribute their own actions to an external source. The authors test the hypothesis that patients with schizophrenia have defects in their ability to predict the sensory consequences of their actions.

Method: The authors measured sensory attenuation of self-produced stimuli by patients with schizophrenia and by healthy subjects.

Results: Patients with schizophrenia demonstrated significantly less sensory attenuation than healthy subjects.

Conclusions: Patients with a diagnosis of schizophrenia have a dysfunction in their predictive mechanisms.

Self-monitoring is fundamental to normal cognitive functions in planning, controlling, and anticipating the consequences of motor acts (1). Prediction plays a key role in such monitoring, allowing a comparison between expected and actual outcomes of an action to be computed (2, 3). An efference copy of the outgoing motor command in conjunction with a predictive model can be used to generate such a prediction (4). One role of this predictive process is to permit the sensory consequences of a movement to be anticipated and used to attenuate the perceptions related to these sensations, thereby enhancing the salience of sensations that have an external cause (5–7). An additional role may be to label movements as generated by oneself or by an external source. If predicted sensory consequences match the actual sensory consequences, the movement is labeled as one's own. However, if the predicted and actual sensory consequences are discordant, as when one's arm is passively moved by someone else, the movement is labeled as externally generated.

A dysfunctional predictive mechanism would lead to incorrect predictions, causing the misattribution of self-generated actions as externally generated. Patients with schizophrenia can demonstrate just such difficulties, when self-generated actions are experienced as being of alien origin—delusions of control or the misperception of self-generated speech as an auditory hallucination (8). Both psychophysical and neuroimaging studies have suggested that self-monitoring may be dysfunctional in patients with schizophrenia (9, 10). Here we directly test the hypothesis that patients with schizophrenia have a defect in their ability to predict the sensory consequences of their actions.
Evidence for Sensory Prediction Deficits in Schizophrenia

Sukhwinder S. Shergill, M.D., Ph.D.
Gabrielle Samson, B.Sc.
Paul M. Bays, M.A.
Chris D. Frith, Ph.D.
Daniel M. Wolpert, M.D., Ph.D.

Objective: Patients with schizophrenia experiencing delusions and hallucinations can misattribute their own actions to an external source. The authors test the hypothesis that patients with schizophrenia have defects in their ability to predict the sensory consequences of their actions. 

Method: The authors measured sensory attenuation of self-produced stimuli by patients with schizophrenia and by healthy subjects.

Results: Patients with schizophrenia demonstrated significantly less sensory attenuation than healthy subjects.

Conclusions: Patients with a diagnosis of schizophrenia have a dysfunction in their predictive mechanisms.

Self-monitoring is fundamental to normal cognitive functions in planning, controlling, and anticipating the consequences of motor acts (1). Prediction plays a key role in such monitoring, allowing a comparison between expected and actual outcomes of an action to be computed (2, 3). An efference copy of the outgoing motor command in conjunction with a predictive model can be used to generate such a prediction (4). One role of this predictive process is to permit the sensory consequences of a movement to be anticipated and used to attenuate the perceptions related to these sensations, thereby enhancing the salience of sensations that have an external cause (5–7). An additional role may be to label movements as generated by oneself or by an external source. If predicted sensory consequences match the actual sensory consequences, the movement is labeled as one's own. However, if the predicted and actual sensory consequences are discordant, as when one's arm is passively moved by someone else, the movement is labeled as externally generated.

A dysfunctional predictive mechanism would lead to incorrect predictions, causing the misattribution of self-generated actions as externally generated. Patients with schizophrenia can demonstrate just such difficulties, when self-generated actions are experienced as being of alien origin—delusions of control or the misperception of self-generated speech as an auditory hallucination (8). Both psychophysical and neuroimaging studies have suggested that self-monitoring may be dysfunctional in patients with schizophrenia (9, 10). Here we directly test the hypothesis that patients with schizophrenia have a defect in their ability to predict the sensory consequences of their actions.
We have previously shown (11) that although healthy subjects are accurate at reproducing a target force when using a joystick, they significantly overestimate the force required when it is directly self-generated. In other words, they exert a greater force than the target when they are using their index finger to make the match. This occurs because self-generated forces are perceived as weaker than externally generated forces of the same magnitude, which arises from a predictive process in which the sensory consequences of the movement are anticipated and partially removed from the perception. Our hypothesis in the current study was that patients with a diagnosis of schizophrenia would be more accurate than healthy subjects in reproducing the external force. This increased accuracy would reflect the failure of the normal sensory attenuation mechanism due to a dysfunctional sensory predictive process.

During the joystick condition, subjects reproduce the external force by using their right index finger to move a joystick that controls the force output of a torque motor. In this situation the active hand is not generating the force directly, but the movement of the hand is translated into a force through the torque motor. It has been shown that in this unusual situation, predictive mechanisms are not employed (12); we anticipated that both patients and healthy subjects should be accurate at this task.

**Method**

We used a recently developed force-matching task that allows us to quantify the sensory attenuation of self-produced stimuli (11) (Figure 1). A target force is applied to the subject’s left index finger by a torque motor. Subjects are then required to reproduce the force they just experienced, either directly by pressing with their right index finger or by using a joystick that controlled the torque motor. Each subject participated in both conditions in a counterbalanced order. The applied forces were measured by using a force transducer mounted in the lever of the torque motor.

**Results**

The 20 patients and 20 healthy volunteer subjects were well matched for age (patients’ mean age=36.4 years, SD=13.4; volunteers’ mean age=35.9, SD=14), gender (12 in each group were men), handedness, and premorbid IQ (13) (patients’ mean=110, SD=8; volunteers’ mean=114, SD=6). All patients except one were treated with antipsychotic medication, the majority with atypical antipsychotic medication, but all were still moderately symptom-
atic; their mean score on the Brief Psychiatric Rating Scale (14) was 36. Most patients had prominent positive symptoms: 16 subjects had a score of 4 or more on items related to suspiciousness or hallucinatory behavior. Two participants (one patient and one healthy volunteer) were found to have produced matching forces that did not significantly correlate with the corresponding target forces in either task ($r^2<0.04$, $p>0.10$). We concluded that they had not properly understood the instructions and removed their data from further analysis.

All participants consistently applied a greater force when using their right index finger to directly match the externally applied target force; they consistently underestimated the force they were applying because their perception of the force was likely to be attenuated. To quantify attenuation in the patients and healthy subjects, we calculated the percentage of the matched force by which the matched force exceeded the target force. The patients were more accurate at the task, showing 27.5% attenuation (Figure 1, red circles), compared with 43.5% in the comparison group (Figure 1, blue circles). When subjects matched the target force using the joystick, both patients and healthy volunteers reproduced the original force much more accurately (Figure 1, diamonds). A repeated-measures two-way analysis of variance (patients versus healthy subjects and direct versus joystick) of the mean matching force normalized by the mean target force showed a significant interaction ($F=4.88$, df=1, 36, $p<0.04$). This interaction arose from the patients having significantly less attenuation in the direct force-generation task (simple main effect $F=4.71$, df=1, 36, $p<0.04$) but no significant difference from the healthy subjects in the joystick task (simple main effect $F=0.16$, df=1, 36, $p=0.69$).

Discussion

The results show that self-generated forces were attenuated less in the patient group, suggesting a dysfunction in their ability to predict the sensory consequences of their actions. This would be in accord with previous imaging data in the verbal domain (10). Although most of the patients were treated with antipsychotic medication, the absence of any difference between patients and healthy volunteers in the joystick matching task suggests that there was no systematic effect of medication on motor performance; indeed, there was no significant correlation between the chlorpromazine equivalents of medication and the degree of attenuation demonstrated by patients. The present study provides support for the presence of a dysfunctional sensory predictive mechanism in schizophrenia. Future work planned to follow up this interesting finding will concentrate on clarifying the state versus trait nature of this deficit by examining patients longitudinally over time, with changes in symptom profile, and the specificity of this deficit for schizophrenia by examining patients with a diagnosis of bipolar disorder also treated with antipsychotic medication.

Received Sept. 22, 2004; revision received Dec. 17, 2004; accepted Jan. 10, 2005. From the Institute of Neurology, University College London; and the Institute of Psychiatry Kings College London. Address correspondence and reprint requests to Dr. Shergill, Institute of Psychiatry Kings College London, De Crespigny Park, London, SES BA6; s.shergill@iop.kcl.ac.uk (e-mail).

Supported in part by the Wellcome Trust, the McDonnell Foundation, and the Human Frontiers Science Programme.

References

The Prevalence of Teachers Who Bully Students in Schools With Differing Levels of Behavioral Problems

Stuart W. Twemlow, M.D.
Peter Fonagy, Ph.D., F.B.A.

Objective: This study looked for a relationship between the prevalence of teachers who bully students and school behavioral problems reflected in suspensions from school.

Method: A convenience sample of 214 teachers answered an anonymous questionnaire about their perceptions of teachers who bully students and their own practices. Teachers were grouped into whether they taught at schools with low, medium, or high rates of suspensions. Analyses of variance were used to analyze continuous variables, and chi-square statistics were used to study categorical variables.

Results: Teachers from schools with high rates of suspensions reported that they themselves bullied more students, had experienced more bullying when they were students, had worked with more bullying teachers over the past 3 years, and had seen more bullying teachers over the past year.

Conclusions: These findings suggest that teachers who bully students may have some role in the etiology of behavioral problems in schoolchildren.

(Clinical practice implications discussed.)

In our research efforts to reduce bullying in elementary schools (1, 2), we felt it was important to explore whether manifest staff attitudes conducive to bullying may contribute to behavioral difficulties in children. We predicted that teachers who work in schools with high levels of behavioral problems 1) will more commonly endorse attitudes accepting bullying and perceive fewer differences between a hypothetical bullying teacher and a hypothetical nonbullying teacher in terms of behavior and motivation and 2) that more teachers would admit to bullying students and more often report a history of bullying in the course of their own education.

Method

We defined a bullying teacher as one who uses his or her power to punish, manipulate, or disparage a student beyond what would be a reasonable disciplinary procedure. Teachers from a representative sample of relatively demographically homogeneous schools in a Midwestern school district were approached for participation in this study. Anonymous questionnaires were placed in each teacher’s mailbox and were to be delivered to an anonymous drop box. Seventy-five percent of all teachers in eight elementary schools, four middle schools, and three high schools participated (214 teachers from a total student enrollment of 4,034). The schools were grouped into low (two elementary, one middle, and one high school), medium (two elementary, one middle, and one high school), or high (four elementary, one middle, and two high schools) levels of student behavioral problems according to the rates of suspensions from school, as reported for the schools in the sample. The teachers were grouped according to whether they taught at schools with low, medium, or high rates of suspensions. The demographic characteristics of the teachers showed no significant differences in age, gender, or the experience of the teachers, nor did the schools significantly differ in the percentage of minority students, the percentage of students in special education, or class size. Some schools with high rates of suspensions had a somewhat higher percentage of students registered in free lunch programs. We combined these variables into a simple risk indicator and used it as a covariate in univariate analyses of variance (ANOVA).

The questionnaire, available from the first author, covered teachers’ experiences with bullying teachers, their personal experience of bullying students, and the characteristics of bullying and nonbullying teachers rated on a 4-point Likert scale ranging from never to always. Our previous work (3) has shown that bullying teachers can be classified into a sadistic type who spitefully humiliate students and hurt their feelings and a bully-victim type who fail to set limits and leave others to solve their problems, i.e., they bully reactively. Test-retest reliability was assessed over 3 weeks with 30 subjects and was in excess of 0.8 across all scales. The two ratings of the characteristics of bullying and nonbullying teachers were subtracted from each other to produce difference scores. The average squared discrepancy across subjects was considered to provide an indication of the extent to which the teachers perceived differences between bullying and nonbullying colleagues. Nine items directly addressed attitudes toward bullying in teachers—e.g., bullying teachers have quiet classrooms, use needless force to discipline students, and put students down to attain order in the classroom—and these were aggregated to yield a single score indicating a favorable attitude toward bullying (Cronbach’s alpha=0.65). ANOVAs were used to contrast teachers’ ratings on continuous variables, and chi-square statistics were applied to categorical variables. All means and percentages are reported in Table 1.

Results

Our prediction that attitudes favoring bullying would be more characteristic of schools with high or medium, rather than low, rates of suspensions was not confirmed. That is, most teachers did not favor bullying attitudes. Although teachers from schools with high and low rates of suspensions rated bullying and nonbullying teachers similarly, there was a significant difference between the teachers from schools with low and medium rates of suspensions (t=2.3, df=90, p<0.03), with more differences between teachers from schools with medium rates of sus-
TABLE 1. Characteristics of Teachers Who Bully Students in Schools With Low, Medium, and High Rates of Suspensions From School

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of teacher (years)</td>
<td></td>
<td>42.0</td>
<td>10.1</td>
<td>40.2</td>
<td>9.7</td>
<td>41.4</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>Number of years of experience</td>
<td></td>
<td>14.7</td>
<td>9.3</td>
<td>13.8</td>
<td>9.5</td>
<td>15.1</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td>Class size</td>
<td></td>
<td>20.1</td>
<td>8.7</td>
<td>22.8</td>
<td>5.7</td>
<td>20.7</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Discrimination between bullies and</td>
<td></td>
<td>6.8</td>
<td>2.8</td>
<td>6.1</td>
<td>2.8</td>
<td>7.4</td>
<td>2.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>nonbullying teachers</td>
<td></td>
<td>2.40</td>
<td>0.46</td>
<td>2.30</td>
<td>0.45</td>
<td>2.40</td>
<td>0.45</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Attitudes favoring bullying</td>
<td></td>
<td>1.67</td>
<td>0.57</td>
<td>1.47</td>
<td>0.50</td>
<td>1.70</td>
<td>0.63</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Experience having been bullied</td>
<td></td>
<td>1.80</td>
<td>0.96</td>
<td>1.50</td>
<td>0.93</td>
<td>1.70</td>
<td>0.93</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>How many bullying teachers have you</td>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>0.5</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>seen bullying students?</td>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>0.5</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>How many bullying teachers have you</td>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>0.5</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>worked with in the last 3 years?</td>
<td></td>
<td>1.3</td>
<td>2.3</td>
<td>0.5</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>χ² (df=2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex in teachers</td>
<td></td>
<td>169</td>
<td>79.0</td>
<td>38</td>
<td>79.0</td>
<td>61</td>
<td>85.0</td>
<td>70</td>
<td>74.0</td>
</tr>
<tr>
<td>Nonwhite race in teachers</td>
<td></td>
<td>76</td>
<td>35.6</td>
<td>13</td>
<td>26.7</td>
<td>24</td>
<td>34.5</td>
<td>39</td>
<td>41.1</td>
</tr>
<tr>
<td>Students with subsidized lunches</td>
<td></td>
<td>90</td>
<td>41.4</td>
<td>16</td>
<td>32.0</td>
<td>28</td>
<td>30.0</td>
<td>47</td>
<td>49.0</td>
</tr>
<tr>
<td>Students in special education</td>
<td></td>
<td>55</td>
<td>25.7</td>
<td>9</td>
<td>18.7</td>
<td>17</td>
<td>23.6</td>
<td>30</td>
<td>30.8</td>
</tr>
</tbody>
</table>

Pensions and teachers from schools with both high and low rates of suspensions. ANOVAs showed significant differences between schools with low, medium, and high rates of suspensions on four variables, confirming the remaining predictions. The teachers who reported that they bullied students were more often seen in schools with high rates of suspensions (p<0.04). Teachers who reported having experienced being bullied themselves as students were more often working in schools with high rates of suspensions (p<0.001). Teachers from schools with high rates of suspensions also reported that they had seen other bully teachers more often over the past year (p<0.001) and had worked with teachers over the past 3 years who had bullied students (p<0.0001) (Table 1).

Discussion

The study (by matching schools and using covariate techniques) controlled for factors that are often associated with increasing behavioral problems in schools, such as large percentages of minority students and special education students, large class sizes, and fewer years of teacher experience. None of these factors was found to significantly influence the findings. Teachers from schools with high and low rates of suspensions tended to see fewer differences between bullying and nonbullying teachers than teachers in schools with medium rates of suspensions. These findings are consistent with the possibility that teachers in schools with low rates of suspensions have less experience with bullying teachers and that teachers in schools with high rates of suspensions, where bullying teachers are more pervasive, have an eroded sensitivity toward bullying. The higher rates of teacher bullying in schools with more problems suggest either that teachers assimilate to the culture of violence that develops in such schools or that individuals with such predispositions drift toward or are more likely to remain in such institutions because of either preference or lack of opportunities to move to less dysfunctional locations. Because transgenerational transmission of abuse is frequently reported in the literature, it is no surprise that teachers who experienced bullying as children grow up to bully others and are more aware of teachers who bully students. Some teachers may drift toward—or even contribute to—the violent culture of problem schools rather than simply being made more violent by them.

There are obvious methodological limitations to this study. The sample was one of convenience, raising a problem of generalization, but the rates of response of the teachers were gratifyingly high. Although causal inferences cannot be made from these correlational findings and the questionnaire lacks validation, it has good reliability. Nonetheless, the findings represent an initial contribution in an area that is difficult to study.

What can the clinician do about the problem? We know that overly negative, critical, and bullying parental behavior contributes to conduct problems in children and, if challenged therapeutically, can do much to reverse the problems (4). We have been able to intervene successfully in schools using a model based on changing the responses of adults and children. The rates of suspension from school decreased significantly when these patterns were addressed (5, 6).

Received Sept. 9, 2004; revisions received Nov. 17 and Dec. 21, 2004; accepted Feb. 18, 2005. From the Department of Psychiatry, the Menninger Clinic, Baylor College of Medicine; the Psychoanalysis Project, the Peaceful Schools and Communities Project, and HOPE Unit, the Psychoanalysis Project.
Clinical research with cognitively impaired adults is important for improving the treatment of several conditions, including Alzheimer’s disease and schizophrenia. Some propose to allow cognitively impaired adults to be enrolled in research only when they complete a formal research advance directive while competent (1–4). To assess the impact of these recommendations on clinical research, we examined the rate at which adults completed a research advance directive in the clinical setting. We also assessed the preferences individuals indicated on their completed forms.

Method

We assessed all adults admitted as inpatients to the National Institutes of Health (NIH) Clinical Center, a 275-bed hospital, from March 14 to Sept. 13, 2000. All individuals admitted to the NIH Clinical Center are participating in or being considered for participation in clinical research. At each inpatient admission, adults receive a pamphlet titled “Advance Directives at the NIH.” This pamphlet describes the NIH advance directive, which allows individuals to appoint a proxy and to indicate their preferences regarding future research participation. All adult inpatients are then asked by an admitting nurse whether they have an advance directive and whether they would like to complete an NIH advance directive. Adult inpatients are encouraged to complete an NIH advance directive because it—unlike most state advance directives—explicitly addresses individuals’ preferences regarding future research participation.

Results: Overall, 11% of adult inpatients completed a research advance directive. Of those who specified preferences, 13% were not willing to participate in future research should they become unable to consent, 76% were willing to participate in research that might help them, 49% were willing to participate in research that would not help them and posed minimal risk, and 9% were willing to participate in research that would not help them and posed greater than minimal risk.

Conclusions: Proposals to allow cognitively impaired adults to participate in research only with a formal advance directive could block important research. More flexible approaches should be considered to protect these individuals.
Clinical research with cognitively impaired adults is important for improving the treatment of several conditions, including Alzheimer’s disease and schizophrenia. Some propose to allow cognitively impaired adults to be enrolled in research only when they complete a formal research advance directive while competent (1–4). To assess the impact of these recommendations on clinical research, we examined the rate at which adults completed a research advance directive in the clinical setting. We also assessed the preferences individuals indicated on their completed forms.

Method

We assessed all adults admitted as inpatients to the National Institutes of Health (NIH) Clinical Center, a 275-bed hospital, from March 14 to Sept. 13, 2000. All individuals admitted to the NIH Clinical Center are participating in or being considered for participation in clinical research. At each inpatient admission, adults receive a pamphlet titled “Advance Directives at the NIH.” This pamphlet describes the NIH advance directive, which allows individuals to appoint a proxy and to indicate their preferences regarding future research participation. All adult inpatients are then asked by an admitting nurse whether they have an advance directive and whether they would like to complete an NIH advance directive. Adult inpatients are encouraged to complete an NIH advance directive because it—unlike most state advance directives—explicitly addresses individuals’ preferences regarding future research participation.

Whether individuals have a prior state advance directive and whether they execute an NIH research advance directive are recorded in their chart and in the electronic medical information system. A copy of each individual’s NIH advance directive is stored in the NIH Clinical Center’s Department of Clinical Bioethics. We examined all three sources to determine how many adults completed an NIH advance directive during the study period. Completed NIH advance directives were examined for content.

Results: Overall, 11% of adult inpatients completed a research advance directive. Of those who specified preferences, 13% were not willing to participate in future research should they become unable to consent, 76% were willing to participate in research that might help them, 49% were willing to participate in research that would not help them and posed minimal risk, and 9% were willing to participate in research that would not help them and posed greater than minimal risk.

Conclusions: Proposals to allow cognitively impaired adults to participate in research only with a formal advance directive could block important research. More flexible approaches should be considered to protect these individuals.

References


Research Advance Directives: Protection or Obstacle?

Palaniappan Muthappan, B.S.
Heidi Forster, J.D.
David Wendler, Ph.D.

Objective: This study assessed how many adults completed a research advance directive and the preferences indicated on the completed forms.

Method: The authors analyzed all 2,371 adults admitted as inpatients to the NIH Clinical Center from March 14 to Sept. 13, 2000.

Results: Overall, 11% of adult inpatients completed a research advance directive. Of those who specified preferences, 13% were not willing to participate in future research should they become unable to consent, 76% were willing to participate in research that might help them, 49% were willing to participate in research that would not help them and posed minimal risk, and 9% were willing to participate in research that would not help them and posed greater than minimal risk.

Conclusions: Proposals to allow cognitively impaired adults to participate in research only with a formal advance directive could block important research. More flexible approaches should be considered to protect these individuals.

BRIEF REPORTS
Results

Overall, 2,371 adults were admitted to the NIH Clinical Center as inpatients during the study period: 82% were white, 10% were African American, 5% were Hispanic, and 3% were Asian/Pacific Islander; 45% were male, and 55% were female. Their mean age was 46 years (SD=15), and 75% were between 30 and 65 years of age. Some individuals were admitted multiple times during the study period, yielding a total of 3,151 inpatient admissions.

Just 11% of the subjects (261 of 2,371) completed an NIH research advance directive during the study period. In addition, 19% (N=450) had a state advance directive; 9% of these subjects (N=41) completed an NIH advance directive as well. Of the 261 inpatients who completed an NIH advance directive, 99% (N=258) appointed a substitute decision maker: 58% (N=150) appointed their spouse, 13% (N=34) appointed a parent, and 5% (N=13) appointed an adult child.

The section of the form that provides standardized check boxes was completed by 216 of the 261 individuals who completed an NIH advance directive: 13% (N=28) checked the box stating they were not willing to participate in research should they lose the ability to consent, 49% (N=106) checked the box stating they were willing to participate in research that "will not help me but might help others and involves no more than minimal risk to me," 9% (N=19) checked the box stating they were willing to participate in research that "will not help me but might help others and involves greater than minimal risk to me" (Table 1).

Discussion

Adults who are unable to provide informed consent should be enrolled in clinical research only with sufficient evidence that such enrollment is consistent with their competent preferences (5). Several proposed guidelines would require that individuals' preferences be documented in a formal advance directive that they had completed while they were competent (1–4). The present study provides the first empirical assessment of the impact that this requirement might have on clinical research. Major findings are that few adults complete a research advance directive, even when given the opportunity to do so in a clinical research setting. Of those who completed a research advance directive, almost half were willing to participate in research that offers them no chance for personal benefit.

The low rate of completion is noteworthy, given that all individuals admitted to the NIH Clinical Center are participating in or being considered for participation in clinical research. In addition, all adult inpatients receive written information describing the importance of research advance directives and are asked whether they would like to complete an NIH research advance directive. These factors suggest that the completion rate in this study may represent the upper range at which adults in general will complete a formal research advance directive.

Although a low percentage of adults completed a formal research advance directive, the absolute number of individuals who were willing to participate in research should they lose the ability to consent was relatively high. Thus, proposals to require a formal research advance directive could exclude many impaired adults whose competent preferences supported research participation. To avoid this possibility, more flexible approaches should be considered to protect cognitively impaired adults while allowing appropriate research.

One possibility would be to develop advance directives that address both research and clinical care. Combined advance directives would allow individuals to document their clinical and research preferences and provide individuals an opportunity to discuss their research preferences with surrogates. In the absence of a formal advance directive, a first-degree relative might be allowed to make research decisions for an impaired adult (6).

To increase the number of impaired research subjects who have an advance directive, investigators and institutional review boards might consider requiring research advance directives for subjects at high risk for losing the ability to consent, such as individuals with mild Alzheimer's disease enrolling in longitudinal studies. At the same time, many individuals are likely to be seen after they have lost the ability to consent. To protect these individuals, an overall approach should encourage individuals to complete a research advance directive and also provide guidance for research with impaired adults who did not complete an advance directive while they were competent.

The fact that so few adults completed a formal advance directive implies that surrogates often will have to make research decisions based on informal evidence of their charge's preferences. The need to rely on informal evidence highlights the importance of clear guidelines, including laws, to clarify who can make research decisions in the absence of a formal advance directive and the standards these individuals should follow when making research decisions (7).

Finally, the need for a mechanism to allow individuals to reject future research participation is highlighted by the finding that 13% of those who completed a research advance directive indicated that they are not willing to participate in research should they lose the ability to make their own decisions. Individuals also should be aware that formally documenting a preference to decline future research enrollment may preclude them from all research, including research with the potential to benefit them.

TABLE 1. Preferences in the Advance Directives of the 216 of 2,371 NIH Clinical Center Inpatients Who Completed the Standardized Check Boxes

<table>
<thead>
<tr>
<th>Preferences If I Become Impaired</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unwilling to participate in future research</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>Willing to participate in personally beneficial research</td>
<td>164</td>
<td>76</td>
</tr>
<tr>
<td>Willing to participate in research with no potential for personal benefit and minimal risk</td>
<td>106</td>
<td>49</td>
</tr>
<tr>
<td>Willing to participate in research with no potential for personal benefit and greater than minimal risk</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>

Willing to participate in personally beneficial research | 164 | 76 |

Unwilling to participate in future research | 28 | 13 |

Willing to participate in personally beneficial research | 164 | 76 |

Willing to participate in research with no potential for personal benefit and minimal risk | 106 | 49 |

Willing to participate in research with no potential for personal benefit and greater than minimal risk | 19 | 9 |

Toward the decreased of 2,371 NIH Clinical Center Inpatients Who Completed the Standardized Check Boxes
The present study has several limitations. First, we assessed completion rates at one institution only, and our results may not generalize to other institutions. Second, although NIH takes several steps to encourage individuals to complete a research advance directive, even more intensive approaches might yield higher completion rates. Third, some individuals may have declined to complete a research advance directive as a way of expressing their unwillingness to participate in future research. However, the NIH advance directive includes a check box that allows individuals to explicitly refuse future research enrollment. This box was selected by 13% of those who completed the form.

Received Jan. 6, 2004; revisions received July 13 and Sept. 13, 2004; accepted Oct. 1, 2004. From the Department of Clinical Bioethics, NIH. Address correspondence and reprint requests to David Wendler, Department of Clinical Bioethics, NIH, Bldg. 10, Rm. 1C118, Bethesda, MD 20892; dwendler@nih.gov (e-mail).

The opinions expressed are the authors’ own. They do not reflect any position or policy of NIH, the Public Health Service, or the Department of Health and Human Services.

References

7. 2003 California Assembly Bill 1371. California 2003-04 Regular Session
Hypertension and Aripiprazole

TO THE EDITOR: Aripiprazole is a second-generation antipsychotic approved for the treatment of schizophrenia and increasingly used in bipolar patients. It acts as a partial agonist at dopamine receptors and displays high affinity for serotonin 5-HT\textsubscript{1A} and 5-HT\textsubscript{2A} receptors on 5-HT neurons, producing partial agonism and antagonism, respectively. Its side effects include insomnia, anxiety, headaches, nausea, vomiting, and somnolence (1). Aripiprazole’s affinity for α\textsubscript{1}-adrenergic and muscarinic receptors is low; consequently, cardiovascular adverse effects are rare. Only postural hypotension (without clinical symptoms) has been described in elderly patients (2). However, we report a case of hypertensive crisis triggered by aripiprazole.

Ms. A, a 56-year-old postmenopausal patient with a 30-year history of paranoid schizophrenia, was admitted to the hospital for an acute exacerbation of her mental disorder. Upon admission, bizarre paranoid delusions and auditory hallucinations were present. A medical history and a physical examination revealed no somatic disorder. Aripiprazole was then initiated at 15 mg/day. Twenty-six hours after her first dose—30 mg—she complained of palpitations. The findings of a physical examination were tachycardia (120 bpm, regular) and arterial hypertension (220/110 mm Hg). Her neurological state was normal. Aripiprazole treatment was interrupted, and Ms. A’s blood pressure returned progressively to its usual range (130/90 mm Hg) over 72 hours without treatment. Aripiprazole was then started again at 15 mg/day. Symptoms (palpitations) and signs (a heart rate of 120 bpm and blood pressure of 220/110 mm Hg) recurred over a few hours. Ms. A was successfully treated with propanolol, and aripiprazole was stopped. A first-generation antipsychotic was introduced, and both somatic and psychiatric outcomes were favorable. Common causes of a hypertensive crisis (an interruption of antihypertensive medication, endocrine and renal diseases) were ruled out.

A number of case reports have demonstrated that the use of new antipsychotics is associated with arrhythmias, prolonged QTc intervals, and orthostatic hypotension in people lacking cardiovascular disorders (3). In this case, we reported that hypertension abated with discontinuation of aripiprazole and reemerged upon rechallenge. Furthermore, sustained hypertension was blocked by the coadministration of propanolol, a nonselective beta-blocker. This points toward adrenergic hyperactivity as the cause of hypertension. Obviously, individual sensitivity is an important factor, as demonstrated by the lack of cardiovascular effects in a case of aripiprazole over-dose (4). Further studies are needed to understand the precise mechanisms involved. This is particularly important because atypical antipsychotics are now commonly prescribed over conventional neuroleptics.

References

Parkinsonism With Intramuscular Ziprasidone

TO THE EDITOR: Intramuscular ziprasidone was introduced for use in the emergency management of acute psychosis (1). Ziprasidone is thought to have a low incidence of extrapyramidal side effects. We report on the induction of parkinsonism in a patient 3 days after its initiation.

Mr. A, a 63-year-old man with history of chronic paranoid schizophrenia, was admitted to the inpatient unit to help manage his uncontrollable violent behavior. He was hypervigilant, paranoid, and actively conversing with inner voices. He refused to take oral medications, including his preadmission regimen of olanzapine, 30 mg/day, and valproic acid, 2000 mg/day. He then assaulted staff members twice, necessitating use of restraints and intramuscular haloperidol, 10 mg/day, for 3 consecutive days. A day after his last haloperidol injection, we gave him 20 mg b.i.d. of intramuscular ziprasidone. Mr. A received 20 mg on day 1, 20 mg b.i.d. on day 2, and 20 mg b.i.d., with an additional third dose of 20 mg at noon on day 3, for emergent aggression. Drooling, a flat expression, bradykinesia, a shuffling gait, and pill-rolling tremor appeared later that day. His symptom profile fit drug-induced parkinsonism. We then restricted his dose to 40 mg/day, and Mr. A’s parkinsonism symptoms subsequently resolved. However, at that dose, his aggression continued; therefore, we discontinued ziprasidone and started intramuscular olanzapine, which reduced the agitation and allowed resumption of oral medication administration.

To our knowledge, there are no reports in the literature associating injectable ziprasidone with parkinsonism. Dystonia (2), an oculogyric crisis (3), and even tardive dyskinesia (4) have been linked to ziprasidone. In a recent study, higher doses of intramuscular ziprasidone (up to 20 mg/day) were associated with extrapyramidal side effects, but parkinsonism was not specifically mentioned (5). Our patient was susceptible to extrapyramidal symptoms because he had a history of brain injury, was older, and had some medical comorbidities. Although those confounds may diminish the role of ziprasidone inducing the complication, it should alert clinicians to the possibilities of its occurrence at higher doses and in the presence of predisposing factors.

References
LETTERS TO THE EDITOR

TO THE EDITOR: Currently, nearly one in every 13 adults abuses alcohol or is alcohol dependent. In addition, several million adults also engage in risky drinking behavior that could lead to alcohol-related problems. However, only a small proportion of the people with alcohol problems ever seek and engage in treatment (1). A national survey in the United States (2) found that only 16% of those with alcohol use disorder had received treatment in 2001. In the Netherlands (3), only 10% of the problem drinkers ever got professional help. Women, higher-educated people, employees, and elderly people are harder to reach for face-to-face care. Several things have improved the accessibility of alcohol interventions (4, 5).

To fill the gap in accessibility, an e-therapy program was developed in the Netherlands. The content and elements of the e-therapy program are comparable to the ones of face-to-face treatment as usual. This e-therapy program distinguishes from other Internet interventions by forming an ongoing personal relationship between the therapist and client that takes place solely via Internet communications. It involves more than self-help or answering a question online. It is a structured and complete treatment program in which clients remain anonymous. The aim of this study was to determine if e-therapy indeed reaches another population with alcohol problems.

We compared the baseline characteristics of 172 e-therapy clients with a consecutive series of 172 face-to-face clients admitted for treatment as usual. All e-therapy clients gave their informed consent to participate in the research project. For the face-to-face clients, we used anonymous data files. The results showed that the baseline characteristics of the two groups differed by gender, age, education, and work situation. Chi-square tests were used to compare proportions and t tests to compare means. The e-therapy group involved significantly more women than the face-to-face group ($\chi^2=46.56$, df=2, $p<0.001$) and more often employed ($\chi^2=69.13$, df=1, $p<0.001$) than the people in the face-to-face group. E-therapy clients were also significantly older than face-to-face clients ($t=3.24$, df=342, $p=0.001$).

Our conclusion is that e-therapy serves a new group of people with alcohol problems. We reach more women, higher-educated people, employed people, and elderly people—the groups that are difficult to reach in regular face-to-face therapy. The Internet offers an opportunity for improving access to therapy for problem drinkers. Our next step is to compare the efficacy and effectiveness of e-therapy and face-to-face therapy.

References


MARLOES G. POSTEL, M.A.
COR A.J. DE JONG, M.D., Ph.D.
HEIN A. DE HAAN, M.D.
Enschede, the Netherlands

cognitively impaired, socially isolated, unemployed, and marginalized. Because all of these impairments are a direct consequence of the disease and its correlates, it would be unfair to say the illness is in remission. Indeed, both the individual and the family might be puzzled as well as frustrated in their daily struggle with the impact of the illness to hear that the medical profession has declared the same patient to be in remission. Remission could also be misinterpreted by insurance companies and payers working on the premise that symptomatic and functional improvement go hand in hand.

We have been presented with clear criteria that define remission in certain important symptom domains; perhaps a label of "symptomatic remission" or some analog thereof may be more appropriate. Labels are powerful symbols; all we ask for is some deliberation on this point before the field adopts them.

Reference

GARY REMINGTON, M.D., Ph.D.
SHITIJ KAPUR, M.D., Ph.D.
Toronto, Ont., Canada

Dr. Andreasen Replies

Thank you, Drs. Remington and Kapur, for your thoughtful discussion about our choice of words. You are perhaps correct that the term "remission" may be misinterpreted and that "symptomatic remission" might have been a better choice.

I certainly concur that a definition that includes measures of quality of life and psychosocial function is preferable. As we stated in the article, we were somewhat handicapped by the lack of widespread consensus on appropriate rating methods for these aspects of schizophrenia. The field will progress, however, and a new and more complete definition of remission will be presented eventually. In fact, here at Iowa, we have been working on a "local" definition that we are already using that incorporates measures of psychosocial function.

I would disagree with you about only one small issue. There is tremendous pressure from patients and family members to have psychiatrists think and speak about schizophrenia in a less pessimistic and more upbeat manner. As you are no doubt aware, there is considerable emphasis on the concept of "recovery." Many families wish we would discuss that possibility more often. In this context, use of the term "remission" is definitely a more modest approach.

NANCY C. ANDREASEN, M.D., Ph.D.
Iowa City, Iowa

Why the Hamilton Depression Rating Scale Endures

TO THE EDITOR: As stated by R. Michael Bagby, Ph.D., et al. (1), the Hamilton Depression Rating Scale was designed to measure depression severity and clinical changes in depressed patients during treatment with antidepressants (2). The validity of the Hamilton depression scale was demonstrated in this patient population (3, 4).

The review by Dr. Bagby and colleagues relevantly underlined the extended use of this scale: only 13 (18.5%) of the 70 studies published since 1979 that examined the psychometric properties of the Hamilton depression scale were carried out in depressed patients.

We suggest that this use, as shown by Dr. Bagby et al., extended the original aim of the scale and that the relevance of the Hamilton depression scale should be discussed in terms of experimental design and specific objectives (3, 4). Concerning experimental design, nondepressed patients should not be taken into account when we study the validity of the Hamilton depression scale because it has been shown that the scale is not valid in nondepressed patients (3, 4). Concerning its specific objectives, the scale should not be compared to DSM-IV criteria because the two measures have different objectives; i.e., the Hamilton depression scale assesses depression severity in depressed patients, and the DSM-IV defines a diagnosis of major depression.

Finally, we do not agree with the conclusion of Dr. Bagby et al. about the lack of validity of the Hamilton depression scale, and we suggest that the scale is a victim of its success and of inappropriate extended use. Unless significant improvement of depression assessment emerges from objective biological and morphological techniques, we do not believe it is possible to create a new instrument that would be able to assess depression from a diagnostic point of view (such as DSM-IV) and a severity point of view (such as the Hamilton depression scale) in all circumstances and in all subjects.

References

EMMANUELLE CORRUBLE, M.D., Ph.D.
PATRICK HARDY, M.D., Ph.D.
Paris, France

TO THE EDITOR: The review of the psychometric properties of the Hamilton depression scale by Dr. Bagby et al. included the most relevant studies published from January 1980 to May 2003 that examine both interrater reliability and validity. It clearly demonstrated that the 17-item version, which in this period has been the gold standard as the outcome measure in trials with antidepressive therapies, is a multidimensional scale.

One potential evolutionary solution for a one-dimensional gold standard, as suggested by the authors, would be to use the six core items of the dimension of depression comprising depressed mood, guilt, work and interests, psychomotor retardation, psychic anxiety, and general somatic symptoms—the 6-item Hamilton depression scale—when we measure the outcome of antidepressive interventions because this subscale has been proven to be more effective than the 17-item...
version in detecting differences between active drug therapy and placebo in trials on the acute therapy of depression. Also, this subscale fulfills criteria for unidimensionality. However, the authors rejected this solution, arguing that the truncated set of six items seems limited in that these items do not "permit capture of the full depressive syndrome" (p. 2174). Exactly this argument was behind the development of the Bech-Rafaelsen Melancholia Scale (1), which contains 11 items covering the various aspects of the full depressive syndrome and is the opposite of what is seen in the full manic syndrome, as measured by the Bech-Rafaelsen Mania Scale (2), that, together with the Bech-Rafaelsen Melancholia Scale, is still the only scale for the bipolar spectrum to fulfill the item-response theory model for unidimensionality (1, 2). In the light of this, we wonder why the authors did not mention this scale in their discussion of a replacement for the Hamilton depression scale.

In a recent study (3), we confirmed in a large sample of depressed patients that the Bech-Rafaelsen Melancholia Scale and the 6-item—but not the 17-item—Hamilton depression scale have been accepted by item-response theory models to be unidimensional depression scales.

Both the 6-item Hamilton depression scale and the Bech-Rafaelsen Melancholia Scale can serve as gold standards for measuring pure antidepressive activity, whereas the other 17-item Hamilton depression scale dimensions (e.g., anxiety and sleep) might serve to identify other aspects of the treatment being examined.

References

RASMUS W. LICHT, M.D. Risskov, Denmark PER BECH, M.D. Frederiksborg, Denmark

To the Editor: The article by Dr. Bagby et al. presented a thorough review and argued persuasively for the rejection of the Hamilton depression scale as the gold standard for the measurement of depression. The results are particularly useful for those who might consider using the scale in a clinical trial.

However, we would like to raise a few concerns regarding the psychometric terms and the statistical indices used in the study. First, the authors used “predictive validity” to determine the ability of the Hamilton depression scale to detect change in depression after treatment. However, predictive validity is commonly used to predict future health status or use of health services. For example, Lahey et al. (1) examined the predictive validity of the DSM-IV diagnostic criteria for attention deficit hyperactivity disorder to predict 3-year symptoms and associated impairment. To describe the extent of a scale’s ability to detect change, “responsiveness” is often used in the literature (2).

Second, Pearson’s correlation coefficient (r) is not appropriate to be used to summarize the item-level agreement (i.e., interrater reliability and retest reliability) (3) of the Hamilton depression scale. Pearson’s r examines the level of linear association—but not agreement—between two (continuous) measurements whose distributions are assumed to follow the normal curve. However, the measurement level of each item of the scale is ordinal. Instead, the weighted kappa examines the agreement between ordinal measurements and adjusts for chance agreement and level of agreement and is the appropriate index to be used in this instance (3). The drawbacks of using Pearson’s r in examining item-level reliability should have been noted.

Third, the purposes of the criteria used for appraising reliability (e.g., Cronbach’s alpha ≥0.70 reflecting adequate reliability or Pearson’s r ≥0.7 indicating acceptable reliability) of the Hamilton depression scale were not clearly specified. The criteria used for appraising reliability in their study are acceptable for research purposes (i.e., for group comparisons) but not for clinical application (i.e., for individual comparisons) (3). For example, if the retest reliability coefficient of a scale is 0.7 (e.g., r=0.7), it means that only 49% of the variance in the data is accounted for (or up to 51% of measurement errors) between test and retest measurement. A higher benchmark (e.g., alpha ≥0.90) for appraising the reliability of a measure is suggested for monitoring an individual’s score (2).

The concerns we raise do not affect the main conclusion of this article. However, they should be clarified for readers.

References

CHING-LIN HSIEH, Ph.D. Taipei, Taiwan
CHENG-HIS HSIEH, M.D. Taoyuan, Taiwan

To the Editor: The article by Dr. Bagby et al. presented a thorough review and argued persuasively for the rejection of the Hamilton depression scale as the gold standard for the measurement of depression. The results are particularly useful for those who might consider using the scale in a clinical trial.

However, we would like to raise a few concerns regarding the psychometric terms and the statistical indices used in the study. First, the authors used “predictive validity” to determine the ability of the Hamilton depression scale to detect change in depression after treatment. However, predictive validity is commonly used to predict future health status or use of health services. For example, Lahey et al. (1) examined the predictive validity of the DSM-IV diagnostic criteria for attention deficit hyperactivity disorder to predict 3-year symptoms and associated impairment. To describe the extent of a scale’s ability to detect change, “responsiveness” is often used in the literature (2).

Second, Pearson’s correlation coefficient (r) is not appropriate to be used to summarize the item-level agreement (i.e., interrater reliability and retest reliability) (3) of the Hamilton depression scale. Pearson’s r examines the level of linear association—but not agreement—between two (continuous) measurements whose distributions are assumed to follow the normal curve. However, the measurement level of each item of the scale is ordinal. Instead, the weighted kappa examines the agreement between ordinal measurements and adjusts for chance agreement and level of agreement and is the appropriate index to be used in this instance (3). The drawbacks of using Pearson’s r in examining item-level reliability should have been noted.

Third, the purposes of the criteria used for appraising reliability (e.g., Cronbach’s alpha ≥0.70 reflecting adequate reliability or Pearson’s r ≥0.7 indicating acceptable reliability) of the Hamilton depression scale were not clearly specified. The criteria used for appraising reliability in their study are acceptable for research purposes (i.e., for group comparisons) but not for clinical application (i.e., for individual comparisons) (3). For example, if the retest reliability coefficient of a scale is 0.7 (e.g., r=0.7), it means that only 49% of the variance in the data is accounted for (or up to 51% of measurement errors) between test and retest measurement. A higher benchmark (e.g., alpha ≥0.90) for appraising the reliability of a measure is suggested for monitoring an individual’s score (2).

The concerns we raise do not affect the main conclusion of this article. However, they should be clarified for readers.

References

CHING-LIN HSIEH, Ph.D. Taipei, Taiwan
CHENG-HIS HSIEH, M.D. Taoyuan, Taiwan
clinicians who actually know the patients, the Hamilton depression scale captures an impressive range of clinical phenomena from mild to extreme illness. In this light, complaints about nonalignment of the Hamilton depression scale with DSM-IV criteria are irrelevant. Likewise, demands for the ultimate in psychometric properties are misplaced. Abridged versions that aim for essentialist purity over undisturbed clinical reality have not gained acceptance. To echo the quip about democracy, the Hamilton depression scale may be the worst depression scale ever developed, except for all the others.

Second, the call for a new scale based on contemporary concepts of major depression is unrealistic. Which proposed concepts should we use? Current definitions of major depression, instantiated in DSM-IV, for instance, are deliberately atheoretical nominalist conventions that lack unifying constructs, predictive validity, and explanatory power. That is one reason why populations diagnosed with DSM-IV major depression are so heterogeneous. In the future, we might add biomarkers or endophenotypes to clinical symptoms in assessing depression, but that day is not here.

Third, as a practical matter, the Hamilton depression scale is not surpassed on performance by any other scale. The view that the Hamilton depression scale is insensitive to change in severity of depression is simply wrong. This charge is often joined with the claim that the Montgomery-Åsberg Depression Rating Scale (2) is more sensitive and therefore preferable as an outcome measure. That claim rests on slim evidence, in a sample of only 35 patients. In a large meta-analysis, the Hamilton depression scale actually was somewhat better than the Montgomery-Åsberg Depression Rating Scale in sensitivity to change and in detecting early change with treatment while having the advantage of far more comprehensive symptom coverage (3). There is no foundation for the assertion of Dr. Bagby et al. that patients might be denied valuable new antidepressant drugs because the Hamilton depression scale lacks sensitivity to register their efficacy.

The endurance of the Hamilton depression scale is remarkable, considering how many unauthorized, nonvalidated, mutant versions now circulate (Hamilton's original 17 items have expanded to 28 at my last count). This is not progress, however, because the text versions and procedural use in many contemporary treatment trials are corrupted.

References

BERNARD J. CARROLL, M.B.B.S., Ph.D., F.R.C.PSYCH.
Carmel, Calif.

To the Editor: It is widely accepted that the Hamilton depression scale is less than ideal as a measure of outpatient depression severity. However, while we await the development and validation of other scales, the Hamilton depression scale will almost certainly continue to be used in regulatory and academic clinical trials for at least a few more years. As Dr. Bagby et al. noted, the Depression Rating Scale Standardization Team developed the GRID-HAMD to fill this gap. The Depression Rating Scale Standardization Team was formed in 1999 by individuals in academia, clinical practice, the pharmaceutical industry, and government to develop a standard approach to administering and scoring the Hamilton depression scale that would remain acceptable to the Food and Drug Administration and be used by pharmaceutical, academic, and clinical researchers (1). The authors described the GRID-HAMD as “virtually unchanged from the original” (p. 2174), but this is not the case. The Depression Rating Scale Standardization Team standardized the administration and scoring of the Hamilton depression scale to improve item reliability by clarifying and operationalizing ambiguous anchor descriptions and providing interview probes and conventions within the instrument (2). Thus, the original intent of the items and the scaling remain the same. Given the many versions of the scale in use, the Depression Rating Scale Standardization Team concluded that standardization would improve the current scale and lay the groundwork for development of a new scale. This effort is now underway by the Depression Inventory Development Project, and item development and field testing are being conducted. (The Depression Rating Scale Standardization Team and the Depression Inventory Development Project are funded by the International Society for CNS Drug Development. The GRID-HAMD can be downloaded, free of charge, at http://ISCDD.org.)

We recognize that not only does it take years to develop a new scale but also that its acceptance requires a thoughtful consideration of diverse theoretical viewpoints, acknowledgment of past efforts, and innovation. The Hamilton depression scale glass of Dr. Bagby et al. is clearly half-empty. We believe the glass should be viewed as half-full and that future efforts should take advantage of all that has been learned from the many years of use of this scale.

References

PER BECH, M.D., Ph.D.
Frederiksberg, Denmark
NINA ENGELHARDT, Ph.D.
Indianapolis, Ind.
KENNETH R. EVANS, Ph.D.
Toronto, Ont., Canada
MICHAEL GIBERTINI, Ph.D.
West Orange, N.J.
AMIR H. KALALI, M.D.
San Diego, Calif.
KENNETH A. KOBAK, Ph.D.
Madison, Wis.
JOSHUA D. LIPSTIZ, Ph.D.
JANET B.W. WILLIAMS, D.S.W.
New York, N.Y.
JAY D. PEARSON, Ph.D.
Rahway, N.J.
MARGARET ROTHMAN, Ph.D.
Raritan, N.J.

JANET B.W. WILLIAMS, D.S.W.
KENNETH A. KOBAK, Ph.D.
MICHAEL GIBERTINI, Ph.D.
KENNETH R. EVANS, Ph.D.
AMIR H. KALALI, M.D.
SAN DIEGO, Calif.
KENNETH A. KOBAK, Ph.D.
MADISON, Wis.
JOSHUA D. LIPSTIZ, Ph.D.
JANET B.W. WILLIAMS, D.S.W.
NEW YORK, N.Y.
JAY D. PEARSON, Ph.D.
RAHWAY, N.J.
MARGARET ROTHMAN, Ph.D.
RARITAN, N.J.
Dr. Bagby and Colleagues Reply

TO THE EDITOR: Drs. Corruble and Hardy correctly note that the 17-item Hamilton depression scale was designed to assess the severity of depression in patients known to be depressed but contend that the preponderance of the studies we reviewed did not use the Hamilton depression scale in this manner, thus negating our claim that the instrument is invalid. We never claimed that the Hamilton depression scale is an invalid instrument. Instead, we stated that “established criteria are met for convergent, discriminant, and predictive validity” (p. 2174). We argued that the Hamilton depression scale lacks factorial and content validity. Many of the studies we reviewed used general psychiatric samples, which likely included patients who arguably would be expected to have a negative affect, even if they did not all meet the criteria for major depressive disorder. Nonetheless, if we accept Drs. Corruble and Hardy’s more restrictive list and examine factorial validity in the studies that used only depressed samples, we still do not have any evidence for this type of validity.

As for content validity, it remains the case that the Hamilton depression scale was based on an understanding of depression that is now more than 40 years old. The proliferation of long-form versions of the Hamilton depression scale attests to the perceived need for additional items to capture the full domain of depression. A modern depression severity instrument does not have to mesh perfectly with DSM-IV. Nonetheless, the DSM-IV symptoms are part of our current conceptualization of depression and should at least be evaluated for their potential contribution to the measurement of depression severity. Drs. Corruble and Hardy conclude, without supporting evidence, that the development of a better scale is unlikely. We pointed out that a better scale can already be found within the Hamilton depression scale items themselves. The various short forms all outperform the 17-item version, a consequence of the apparent multidimensionality of the full scale. Our review also noted the problems with simply adopting one of these short forms, and so we look forward to a new instrument that incorporates contemporary psychometric methods and current definitions of depression.

In their letter, Drs. Licht and Bech note that we failed to mention the Bech-Rafaelsen Melancholia Scale as a possible candidate to replace the Hamilton depression scale. The development of this instrument used item-response theory as well as a more comprehensive list of core symptoms. Unfortunately, space limitations prohibited a full discussion of all potential replacement instruments. We agree that the Bech-Rafaelsen Melancholia Scale is an excellent candidate for a “new gold standard,” and we look forward to research comparing this instrument with the other alternatives mentioned in our review—the Inventory of Depressive Symptomatology, the Montgomery-Åsberg Depression Rating Scale, and the measure currently being developed by the Depression Inventory Development Initiative.

Drs. Hsieh and Hsieh suggest that some of the psychometric terms and statistical indices used in our evaluation of the Hamilton depression scale may not be appropriate; for example, that the term “responsiveness” should be used instead of “predictive validity” to describe the capacity of the Hamilton depression scale to detect change in severity of depression. We agree that “responsiveness” is a more precise word, but we deliberately chose to use a more common and conceptually broader psychometric term. In the case of the predictive validity of the Hamilton depression scale, we focused on whether change scores predict change in depressive severity.

We agree that Pearson’s correlations are less than ideal for assessing the reliability of individual item-to-item comparisons, especially when the scaling is different. That said, all of the studies reviewed used this coefficient, causing us by necessity to rely on it. Note that the use of Pearson’s r likely produces inflated estimates of association relative to weighted kappa, which “corrects” for chance association, with the result that many of the individual Hamilton depression scale items are likely more problematic than we concluded. We would, however, argue that Pearson's r is, in fact, appropriate for examining item-to-total correlations as composite scores, such as Hamilton depression scale total scores, approach interval-level measurement. Pearson’s r is widely used to compare individual items with total scores.

Finally, Drs. Hsieh and Hsieh suggest that a higher benchmark of internal reliability (e.g., Cronbach's alpha ≥0.90) should be employed when examining instruments that will be used for the assessment of individuals. The consequence of doing so would be to demonstrate that only two of 13 studies reported adequate internal reliability. We employed a more liberal benchmark (i.e., Cronbach's alpha ≥0.70) primarily because we did not want to be accused of applying an overly strict criterion (1, 2). The Drs. Hsieh, however, raise a potentially important distinction between group- and individual-level comparisons. The Hamilton depression scale may, in fact, be even weaker than suggested by our article, with insufficient reliability for the assessment of depressive severity in individual patients.

Dr. Carroll argues that the Hamilton depression scale was developed to record the severity of clinical depressive illness, not to “quantitate the metaphysical construct called ‘major depression.’” We agree that the focus of the instrument is quantification of severity, not the fixing of a diagnosis, but we wonder how one establishes the severity of an illness without carefully considering the diagnostic and associated features of that illness. Quantifying severity does not require a perfect correspondence between the instrument and DSM-IV, but the instrument should be informed by changes in the diagnostic system. It seems unhelpful to retain an item such as “loss of insight,” which makes neither a conceptual nor an empirical contribution. Evaluating the potential contribution of more recently noted symptoms would better serve the measurement of depression. For example, “loss of concentration” is a widely known symptom that is included in DSM-IV but is not included in the Hamilton depression scale. Concern with the outdated item content of the Hamilton depression scale surely drives the proliferation of long-form versions.

Dr. Carroll also argues that the Hamilton depression scale is a “clinimetric index...focused on the patient’s burden of illness” and that the wide range of symptoms covered is “consistent with the pleomorphic presentations of clinical depression.” We disagree. Hamilton and colleagues stated clearly that the Hamilton depression scale is a structured rating scale designed to assess depression severity. Patients do appear with a wide range of symptoms, but a rating scale for depression should be limited to the symptoms that contribute to its measurement. For this reason, we do not agree that develop-
ing “a new scale based on contemporary concepts of major depression is unrealistic.” DSM-IV is far from perfect, but it does represent the official definition of the construct whose severity we are purporting to measure, while it also identifies several potentially important symptoms not included on the Hamilton depression scale. When so much else has changed in our knowledge both of depressive symptoms and of psychometrics, it makes little sense to argue that our best effort occurred in the late 1950s. “Biomarkers” and “endophenotypes” may be a desirable long-term goal, but it is not necessary to use outdated instruments in the interim.

Finally, Dr. Carroll argues that the “Hamilton depression scale is not surpassed on performance by any other scale.” On the contrary, the Hamilton depression scale is, in fact, surpassed by subscales composed of Hamilton depression scale items. If a 6-item subscale outperforms a 17-item full scale, it would appear that a majority of the items are actually compromising the use of the total score. One study (3) found that the use of a Hamilton depression scale subscale would allow sample sizes to be cut by one-third without compromising power. We never claimed that the Hamilton depression scale was insensitive to change, and the predictive validity section of our article (pp. 2172–2173) reviews several studies that demonstrate the capacity of the Hamilton depression scale in this regard. What we did argue was that the multidimensional structure of the Hamilton depression scale makes difficult the evaluation of specific treatment effects.

Dr. Bech and associates from the Depression Rating Scale Standardization Team are concerned with our view that the GRID-HAMD was “virtually unchanged from the original,” arguing that their group, in fact, implemented many changes. We felt that this quotation may have been taken out of context. We wrote that “the GRID-HAMD content is virtually unchanged from the original” (p. 2174). In fact, we acknowledge that the GRID-HAMD offers much that we believe is necessary for construction of a better measure of depression severity, and we were concerned that the effort would be hampered by retention of the original item contents. That said, we have learned that recent efforts of the Depression Rating Scale Standardization Team have been directed toward carefully developing new items as well, a development that we applaud.

References

R. MICHAEL BAGBY, Ph.D.
DEBORAH R. SCHULLER, M.D.
Toronto, Ont., Canada
ANDREW G. RYDER, Ph.D.
MARGARITA B. MARSHALL, B.Sc.
Montreal, Que., Canada

Predictors of Premature Termination of Anorexia Nervosa Treatment

To the Editor: As clinicians and researchers, we found the article by D. Blake Woodside, M.D., F.R.C.P.C., et al. (1) of considerable interest because it is a naturalistic study focused on a problematic issue. Indeed, studying predictive factors of the failure of therapeutics for the most severely ill anorexia nervosa patients has been rare (2–4).

We identified three criteria that influence the dropout rate from inpatient programs. The first criterion has previously been described: the treatment method (2, 5). The second criterion is the definition of “dropout” itself. Indeed, none of these studies (1–4) considered “dropout” identically. As a result, the dropout rates varied from 20% if one considers only patient-initiated discharges (4) to 50% if one considers patient-initiated discharges and staff-initiated discharges (1). The third criterion is the clinical characteristics of the subjects studied. The previous studies (1–4) focused on adults with anorexia nervosa ages 20.5 years (SD=4.8) to 27.1 years (SD=9.0), on average. To our knowledge, there has been no study about this issue in adolescents.

In our adolescent inpatient unit in France, 268 girls (mean age=16.7 years, SD=2.0) were hospitalized between 1996 and 2004. The mean duration of illness was 20.4 months (SD=17.1, median=13.9). When we considered dropouts as all subjects who did not achieve their therapeutic contract (did not reach the target weight for discharge) (6), we found a dropout rate of 24.6% (N=66). One-half were staff-initiated, and one-half were patient-initiated. Only six of 66 (9%) were discharged at the first part of their therapeutic contract (early dropouts). As usual, with the youngest patients, the purging subtype was rare (N=51, 19%). To examine dropout predictors, we performed a step-by-step backward logistic regression analysis. We considered 13 predictive variables reported by the literature (1–4): body mass index (kg/m²) at admission, maximum and minimum previous body mass indexes, age at admission, length of hospitalization, time since first treatment, duration of illness, anorexia nervosa subtype, number of previous hospitalizations, body mass index at discharge, age at onset, educational status, and socioeconomic status.

Four variables were significantly related to dropout: higher body mass index at admission (odds ratio=1.5, p<0.03), lower body mass index at discharge (odds ratio=0.2, p=0.0001), longer length of hospitalization (odds ratio=1.0, p=0.0001), and later age at onset (odds ratio=1.4, p=0.005).

In contrast to the report by Dr. Woodside and colleagues, our dropouts had a longer length of hospitalization. This could be explained by the setting of our therapeutic contract. Indeed, the staff never discharge a patient because of lack of progress (e.g., lack of weight gain) in the first weeks.

The few patient-initiated discharges (N=27, 10%) were probably due to the subjects’ age: they can leave the hospital only with their parents’ permission. One may hypothesize that the therapeutic alliance between the parents and staff helps the subjects remain hospitalized.

Although most of the studies (1, 3, 4) indicated that the purging subtype of anorexia nervosa was a predictor of dropout, we did not find this result (approximately 19% had the purging subtype in both groups). Once more, the setting of our inpatient program and the adolescents’ age could explain this result. These results stress the need for further research on this issue.

MARGARITA B. MARSHALL, B.Sc.
R. MICHAEL BAGBY, Ph.D.
ANDREW G. RYDER, Ph.D.
Montreal, Que., Canada

Predictors of Premature Termination of Anorexia Nervosa Treatment

To the Editor: As clinicians and researchers, we found the article by D. Blake Woodside, M.D., F.R.C.P.C., et al. (1) of consid-
TO THE EDITOR: We read with interest the letter by Dr. Godart et al. concerning the rates of premature termination of inpatient treatment for anorexia nervosa. This issue, studied minimally in adults, has been neglected in adolescents. I expect that some of the differences in the findings of the study by Dr. Godart et al. and ours can be explained by the nature of the patient population. I would also be interested in knowing how Dr. Godart et al. established a failure in a patient’s therapeutic contract, given that the patients were all adolescents and receiving treatment at least partly at their parents’ behest.

D. BLAKE WOODSIDE, M.D., F.R.C.P.C.
Toronto, Ont., Canada

Brain-Derived Neurotrophic Factor in Patients With Remitted Depression

TO THE EDITOR: We read with interest the article by Alexander Neumeister, M.D., and colleagues (1). The neurobiology of brain-derived neurotrophic factor (BDNF) is complex and influenced by a number of different hormonal systems, including the hypothalamic-pituitary-adrenal (HPA) axis, which is known to be dysfunctional in patients with severe mood disorders. Stress-responsive corticosteroids, which are the end products of the HPA axis, have been shown to have important effects on the expression of BDNF in preclinical studies (2). We have also recently shown an interaction between cortisol and serum levels of BDNF in patients with bipolar depression and schizophrenia (3). Furthermore, tryptophan depletion has been shown to lower cortisol levels in patients with mood disorders (4) as well as in healthy comparison subjects (5). Of interest, sham tryptophan depletion has also been reported to cause a significant decrease of plasma cortisol (4). Changes in cortisol levels may, therefore, account for the increases in BDNF following sham depletion observed by Dr. Neumeister and colleagues.

The data presented by Dr. Neumeister et al. may indeed suggest an intimate link between the serotonergic and neurotrophic systems, but in the absence of any data regarding HPA axis function in these patients (and healthy subjects), it remains a possibility that the observed changes of expression of BDNF are secondary to differences in cortisol production. We advocate that further studies of BDNF in mood disorders also investigate HPA axis function.

REFERENCES


PAUL MACKIN, M.B., B.S., Ph.D., M.R.C.PSYCH.
PETER GALLAGHER, B.Sc., M.Phil.
Newcastle Upon Tyne, U.K.

Drs. Neumeister and Manji Reply

TO THE EDITOR: We read with interest the letter by Dr. Mackin and Mr. Gallagher in which they suggest interpreting the interactive effects of BDNF and serotonin in major depressive disorder in the context of HPA axis function. Preclinical studies (1) have suggested major functional interactions between knockout mice with genetically induced alterations in serotonin (5-HT) transporters and heterozygous BDNF knockout mice and have also shown that this leads to enhanced stress responses with altered HPA axis function. Notably, a decrease in BDNF concentrations does not appear sufficient to lower extracellular 5-HT; similarly, constitutional changes in extracellular 5-HT because of differences in 5-HT reuptake by 5-HT transporters do not affect BDNF protein levels. This adds to the importance of identifying additional parameters that may contribute to the interactive effects of 5-HT and BDNF in major depressive disorder. There is a wide range of evidence supporting the idea that glucocorticoids play a key role in acute and chronic stress responses. For example, stress and glucocorticoids impair hippocampal neurogenesis; furthermore, in addition to directly causing neuronal atrophy, stress and glucocorticoids also impair cellular resilience that together may lead to the well-established morphological alterations in major depressive disorder. Notably, BDNF and other...
neurotrophic factors are believed to counteract these effects (2). It has been previously demonstrated that in addition to glucocorticoids, BDNF is involved in the early response to acute stress (3). In our study, tryptophan depletion was used as a model to study the effects of acute stress in the context of reduced 5-HT function in major depressive disorder and healthy comparison subjects. Additional work is clearly necessary to delineate the causal relationships between altered 5-HT function, BDNF, and HPA axis function and the pathogenesis of major depressive disorder. Dysregulation of these cascades may be a key mechanism by which stress induces impairments of cellular plasticity. This highlights the interactive effects of different neurobiological systems in the pathogenesis of major depressive disorder, and all three of the referenced major neurobiological systems appear to be involved.

References

ALEXANDER NEUMEISTER, M.D.
West Haven, Conn.

HUSSEINI K. MANJII, M.D.
Bethesda, Md.

Validity of the One-Criterion Threshold for an Alcohol Abuse Diagnosis

TO THE EDITOR: We were surprised to read the results of the article by Marc A. Schuckit, M.D., et al. (1). The authors concluded that all four DSM-IV alcohol abuse criteria perform equally well and that their results favored the threshold of one criterion for the diagnosis of alcohol abuse. This is remarkable because the validity of the abuse category has been one of the main controversies of the DSM-IV classification for alcohol use disorders (e.g., reference 2). We have serious reservations regarding the validity of the data of Dr. Schuckit et al. and their subsequent justification of the one-criterion threshold for alcohol abuse.

First, the study group in the article by Dr. Schuckit et al. was rather unusual, with more than 70% of the subjects being relatives of treatment-seeking alcoholics. This limited the generalizability of their findings. For example, in our recent study of a large general population sample (N=7,076) (3), subjects with a DSM-IV diagnosis of alcohol abuse could not be differentiated from subjects without a DSM-IV alcohol use disorder diagnosis with a broad range of external validators (e.g., psychiatric comorbidity, functional status, familial alcohol problems, treatment seeking). Subjects with two or more criteria, however, were significantly different from subjects without a DSM-IV alcohol use disorder, indicating better validity for a threshold of at least two criteria (3).

Second, most of the validators for the DSM-IV diagnosis of alcohol abuse in the article by Dr. Schuckit et al. are rather weak. The fact that subjects with abuse had a higher intake of alcohol than the subjects without an alcohol use disorder is hardly surprising and almost tautological. When the authors looked at drug-related history to compare subjects with one-criterion abuse to subjects without an alcohol use disorder, they did mention drug use (significant difference for cocaine use only) but failed to mention the more relevant prior history of drug abuse or dependence.

Third, when evaluating the 5-year outcome, the authors seemed to ignore the fact that over 70% of the subjects with a DSM-IV diagnosis of abuse at baseline did not endorse any abuse criterion at follow-up. In our general population study, even higher rates of spontaneous remission of DSM-IV alcohol use were observed: 81% and 85% at the 1- and 3-year follow-ups, respectively (unpublished report). In a prospective evaluation of the validity of current DSM-IV abuse criteria, these findings should at least be discussed.

In summary, we feel that the limitations of the study by Dr. Schuckit et al. call for a more cautious interpretation and that their findings cannot simply be used as support for the validity of the one-criterion threshold for the diagnosis of alcohol abuse.

References

CARLA DE BRUIJN, M.D., PH.D.
WIM VAN DEN BRINK, M.D., PH.D.
Utrecht, the Netherlands

Dr. Schuckit Replies

TO THE EDITOR: Clinical issues are complex, and study results reflect the methods and subject groups used. Therefore, the “true” answer to a question is likely to be found only when one pays attention to differences as well as similarities across research reports. So if the unpublished study cited by Drs. de Bruijn and van den Brink generates some different answers than the current report, such disagreements may offer important insights into the questions raised. However, Drs. de Bruijn and van den Brink incorrectly asserted that in our study, 70% of the subjects with abuse at baseline endorsed no abuse criterion at follow-up. In Table 2 of our study, 54% of the subjects who had three or more experiences with any alcohol abuse item at study entry reported at least one problem at follow-up, 42% endorsed a dependence item, and 25% experienced at least one abuse item at least three times. Therefore, between 54% and as many as two-thirds may have had such experiences.

The letter also incorrectly states that we reported no differences between those with one versus two abuse items at baseline. The tables reveal that those with two or more items had

2400 http://ajp.psychiatryonline.org

Am J Psychiatry 162:12, December 2005
more problems at follow-up, as well as more legal difficulties, and more often developed dependence. The question is whether requiring two abuse items justifies the loss of the ability to diagnose individuals with one item when the latter also predicted adverse outcomes. Similarly, Drs. de Bruijn and van den Brink contend that our results suggested that all abuse criteria performed equally well, but our tables note differences in the proportion endorsing specific items, their demography, and some differences in outcome (e.g., 1.7 future problems for those with hazardous use versus 3.05 for those with baseline social problems). However, with the possible exception of legal problems, there were enough similarities across criteria to consider continuation of the use of those items. Contrary to what Drs. de Bruijn and van den Brink state, a wide range of cross-sectional validators and outcome measures were incorporated into our study. These include drug use, alcohol problems, and demography, as well as quantities and frequencies of intake. This range of items appears to adequately measure aspects of abuse criteria. Finally, Drs. de Bruijn and van den Brink question the validity of our results. Apparently, rather than validity, they are referring to reservations about the generalizability of the sample used by the Collaborative Study on the Genetics of Alcoholism. As we highlighted in paragraphs five, nine, and 10 of the Discussion section, of course our study group had liabilities as well as assets. We agree that the interpretation of any results must consider both the methods and the populations used. The final answer regarding the reliability and predictive validity of the abuse criteria requires careful evaluation of a range of studies using different methods in different subjects. Giving careful consideration to the different results across studies without rejecting the findings from careful investigations that disagree with one's own data is how science steps forward.

MARC A. SCHUCKIT, M.D.
San Diego, Calif.

Explanatory Pluralism and Patchy Reductionism

To the Editor: A major hurdle for clinicians regularly engaging in the critical reductionist approach advocated by Kenneth S. Kendler, M.D. (1), is our own method of diagnosing psychiatric problems, embodied in the DSM-IV-TR. This approach consists of grouping symptoms and applying a label. Purely phenomenalistic, it makes no mention of the underlying causes of those symptoms—whether they are biological, psychological, or social. We must remember that in psychiatry, unlike most other fields of medicine, a diagnosis represents the beginning of the assessment of a patient, not the end. Successful treatment is based on an explanation of the biopsychosocial factors that make up a diagnosis, not only on the diagnosis itself.

Reference

CHRISTOPHER G. IVANY, M.D.
Washington, D.C.

To the Editor: The concept of mental first-person experiences is not as irrevocably grounded as Dr. Kendler suggested. He maintained the solipsism of the individual, ignoring Wittgenstein's concept of language as a tool that individuals use to interact with the environment (1). Wittgenstein noted that language and action produced by thought are a means of producing an empathic relationship between the first and second person. Psychiatrically, this empathic interaction results in a phenomenological psychopathology, a process that is vital to the practice of psychiatry.

Dr. Kendler avoided the basic problem psychiatry faces, which is, what is a mental state? In so doing, he leaves any potential framework hanging in limbo, maintaining the gulf between the mind and the brain. Dr. Kendler did not mention intentionality, which Brentano noted in 1874 (2) as characterizing "the mental." The concept of intentional causality associated with meaning and belief and the nonintentional associated with chemical and physical law-like relationships can provide an acceptable explanation of mental function. The mind can then be explained in a framework of dynamic intentional and nonintentional causal processes in which top-down and bottom-up causality can explain all mental functions and dysfunctions, from molecular interactions to the higher-level intentional processes that produce consciousness.

References

WILLIAM G. CAMPBELL, C.C.F.P., M.D.
KEITH I. PEARCE, F.R.C.P., M.D.
Calgary, Alta., Canada

Dr. Kendler Replies

To the Editor: I appreciate the interest expressed in my recent article by the authors of these letters. I agree with the critical point made by Dr. Ivany that our current approach to psychiatric disorders is descriptively and not etiologically based. This poses important problems that I could not explore in this article. Peter Zachar and I (1) have tried to bring a bit of light to this important question in a review to be published in the Journal.

My ability to respond to the letter of Drs. Campbell and Pearce is limited by my lack of understanding of Wittgenstein's writings as well as parts of their letter. I am less certain than they are that his work, combined with an emphasis on the long-central concept of intentionality, can solve the mind-body problem. I agree with Fodor's critique of the problems of applying the work of Wittgenstein to questions in psychology (and psychiatry) (2). However, I see an ideal theory as so far beyond our current reach that I am happy to settle for explanatory pluralism and patchy reductionism.

References

KENNETH S. KENDLER, M.D.
Richmond, Va.

Reprints are not available; however, Letters to the Editor can be downloaded at http://ajp.psychiatryonline.org.

It is now 60 years since the development of the atomic bomb. The cold war worry of a nuclear confrontation between the United States and the Soviet Union has given way to fear of rogue states or terrorists unleashing havoc. Today’s concerns were foreseen by many who worked on developing the bomb, including J. Robert Oppenheimer, the scientist known as the father of the atomic bomb.

This compelling and comprehensive biography of Oppenheimer rewards the reader with a story rich in detail and excitement as it examines the life, relationships, and work of an American genius. Oppenheimer was a particularly complex man who could be generous and charming but could also be arrogant and insensitive. The authors make use of material gathered over a period of 25 years, including interviews with key characters in Oppenheimer’s life, many now deceased, previously secret FBI files, and correspondence heretofore unavailable. The book is populated by a richly presented, fascinating cast of characters. From the secret culture of Los Alamos to the post-World War II period in which right-wing America believed in massive nuclear retaliation as defense policy, the reader is swept into the McCarthyite anti-Communist craze, which ultimately led to the destruction of Oppenheimer’s reputation as well as those of many other scientists and intellectuals.

Oppenheimer was the oldest son of a New York City German-Jewish family that accumulated sufficient wealth in one generation to provide an environment of privilege and intellectual indulgence. Secular in outlook, the family was involved in the Ethical Culture movement, a humanist outgrowth of Reform Judaism emphasizing social justice that had considerable impact on Robert. His early life is a portrait of a brilliant boy who was interpersonally immature. He would eventually describe himself as discontented, socially awkward, and arrogant in his younger years. In graduate school he suffered a substantial depression fed by existential struggles and characterized by dramatic and, at times, bizarre behavior. He eventually emerged from this period on his own, established himself as a brilliant and charismatic theoretical physicist, and built a renowned theoretical physics department at Berkeley.

During the economic depression of the 1930s Oppenheimer became a supporter of leftist social causes and could count as colleagues and friends a number of Communists, including his younger brother and one of the great romantic loves of his life. That he never became a Communist himself is well substantiated in FBI files, but his affiliations in those years would repeatedly come back to haunt him. Oppenheimer shed Communist connections once the war began because he did not want them to interfere with his “usefulness to the nation.”

The story of Oppenheimer’s brilliant work as director of the central laboratory for the Manhattan Project is in part the story of how General Leslie Groves, who was in charge of the project, chose him. Recognizing his need for an ambitious genius like Oppenheimer for the success of the project, Groves was able to set aside the vast cultural and political divides between them. The elation over the success of developing the bomb was followed by the recognition of the awful consequences of its use. The question of whether dropping the bomb on Japan was necessary to force surrender is amply discussed along with the efforts of scientists to play a subsequent role in the use of the technology they had released.

Prometheus not only gave fire to humans, he also challenged the other gods with his arrogance. Following the war, Oppenheimer was celebrated as a hero. He became director of the Institute of Advanced Study in Princeton. Here he had ample opportunity to offend Lewis Strauss, a right-wing member of the Institute’s board of directors who was also chairman of the U.S. Atomic Energy Commission. Oppenheimer opposed the policy of massive nuclear retaliation supported by Strauss in favor of a policy of openness and international control. Oppenheimer was a persuasive speaker and had low regard for Strauss, which he did not hide. Strauss was all too eager to bear grudges over trivial narcissistic wounds. He eventually found a way to get even with Oppenheimer and neutralize his influence. In 1954, he established a hearing board to review Oppenheimer’s security clearance as an Atomic Energy Commission consultant. In a setting that violated constitutional rights as a matter of course, Oppenheimer lost his clearance and was publicly humiliated. The story of Oppenheimer’s destruction, his unwillingness to avoid confrontation, and the collusive forces behind his fall is high drama and reminds us of the horrors of this period in U.S. history.

Oppenheimer’s reputation was eventually resurrected in the 1960s, but the damage of the 1950s left its mark. A lifelong chain smoker, he died of throat cancer in 1967 at age 62. The authors have succeeded in producing a very well-written biography that is excellent history and rivals many novels in character development, plot, and excitement.

ARNOLD WERNER, M.D.
East Lansing, Mich.

There is a long and distinguished tradition of brief biography, from Plutarch’s Lives, through Aubrey’s Brief Lives and Johnson’s Lives of the Poets, to Strachey’s Eminent Victorians. To this list must be added the Penguin Lives series, which now numbers more than 30 volumes and has a famous set of authors writing about interesting people. The subjects range from St. Augustine (Gary Wills), through Mozart (Peter Gay), James Joyce (Edna O’Brien), Woodrow Wilson (Louis Auchincloss), to people as varied as Pope John XXIII (Thomas Cahill) and Elvis Presley (Bobbie Ann Mason). Ada Louise Huxtable’s volume on Frank Lloyd Wright is a worthy addition to this list.

Wright may indeed have been the genius that he presented to the world. However, he did not spring directly from Zeus’s brow as he suggested; he served an apprenticeship with distinguished predecessors, knew and understood both the classical language of architecture and new developments in the field, and learned his trade. His skills as both draftsman and architect were recognized quickly, and he rose as he moved from one firm to another before going out into independent practice. When he did so, Wright began to redefine “what architecture can do and how it should look” (p. vii). His “tools were the power of his imagination and his aesthetic sensibilities” (p. viii).

Huxtable’s lucid prose permits us to see Wright’s buildings whole and in historical context. His early work, the Prairie houses, were built with strong horizontals and open spatial planning “that broke the barrier between indoors and out” (p. 74); they merged into the surrounding Midwestern plains. Later he designed the Imperial Hotel in Tokyo, which withstood the worst earthquake in Japan’s history. Wright drew the plans knowing that he was “building against doomsday” (p. 150). In the 1930s he designed and built Fallingwater—a cascade of balconies hanging over a stream in Pennsylvania—and the headquarters for the Johnson Wax Company. (Fallingwater is now being restored.) Also in the 1930s he designed a simple single-family home that he called “Usonian.” Often L-shaped and with a carport (possibly a Wright invention), the Usonian house remains as a degraded derivative: the ranch house.

In the 1940s, Wright designed the Guggenheim Museum, a snail-like spiral that was built in the 1950s. The critic Neil Levine wrote,

Wright offered a highly figurative naturalism, to the Europeans’ mechanistic functionalism, he countered with a romantic expressionism, to their…standardization, he proposed instead a material- and site-specific ad hocism, and to their collective vision of a regularized urban order, he opposed a pragmatic individualism. (p. 185)

Wright was often criticized for leaking roofs and windows as well as crumbling concrete. However, he was inventing forms and using new, untested materials. His contractors were often working without standardized procedures and were unfamiliar with the materials. If he was on site, however, the defects were often corrected.

If not from Zeus’s brow, where did Wright’s genius come from? While not a psychobiography, Huxtable’s book provides us with tantalizing clues about Wright’s psychological development. As one expects, they help us to understand the person but not the genius. His parents married late. His father was a gifted musician, orator, and sometime Baptist, later Unitarian, preacher. His mother came from a large family of small farmers. She extruded his father, first from her bed and then from the household, by the time Wright was 10. Father and son had little later contact. Nonetheless, Wright probably owed much of his interest in art and music to his father.

Wright grew up adored and idolized, cared for and coddled. Used to the presence of a strong woman arranging the background of his life, Wright considered his mother an important presence until her death, maintaining a relationship of mutual dependence with her. She always “understood,” accepted wives and mistresses, and cared for him when he was ill. One of his biographers, Brendan Gill (1), thought she was slightly mad.

Shortly after he was 40, Wright’s life spun out of control. He became deeply depressed and, abandoning wife and children, deserted his practice, leaving behind debts and unfinished projects. He soon reappeared, however, with another woman. He built Taliesin, a studio/home in Wisconsin. This new mistress and their children were brutally murdered, and Taliesin burned to the ground. Wright soon took up with another woman, and then another, and another. His last wife, Olgivanna, picked up the role his mother had played; she ran the day-to-day operations and dealt with the household and studio.

Wright started his autobiography in the late 1920s. It was originally published in 1932, and he later revised it twice. Huxtable notes that it “was a creative and cathartic exercise in selective memory...that reveals as it conceals...truth and lies are woven together” (p. 181). Wright lied about his age, his education, and anything else that served his purpose. The historian Hines stated that Wright “had no conception of ‘truth’ as most people define it....His unique creative nature demanded and conceived for himself a persona, a mythic personality surrounded by a partially mythic world” (p. 35). Wright felt that “his needs were urgent, a society that condemned him held false standards, his personal values put him beyond censure in his own mind” (p. 72). He survived scandal, murder, fires, divorces, bankruptcy, social ostracism, and pursuit by the FBI for violation of the Mann Act and accusations of sedition (p. xv).

Brendan Gill (1) described him as a consummate con man. Wright appears to have believed in himself as Richard Wagner did when the latter said, “I am not made like other people. I must have brilliance and beauty and light. The world owes me what I need” (quoted by Caroline Offit [2]). Huxtable notes, correctly, that character and creativity are not correlated, that terrible people can do wonderful things. Despite Wright’s wit and charisma, I am happy to know this first “star” architect only as filtered through the wonderful story told by Huxtable.

References

WILLIAM A. FROSCH, M.D.
New York, N.Y.
In My Brother’s Shadow: A Life and Death in the SS, by Uwe Timm; translated from the German by Anthea Bell. New York, Farrar, Straus & Giroux, 2005, 160 pp., $18.00.

Uwe Timm is a leading German writer and novelist. In My Brother’s Shadow (more literally translated as “My Brother’s Example”) was published in Germany in 2003. It is a literary collage, reading like a transcript of free associations, comprising historical accounts, memories, letters, excerpts from his brother’s notebook, biographical notes, anecdotes, fantasies, dreams, quotations, and psychological and sociological insights. These are unified by Timm’s attempt to comprehend the paradoxes of his brother as older sibling, hero, and Nazi Schutzstaffel (SS) trooper, his family grappling and failing to grasp with the trauma of his brother’s death, his nation struggling to deal with the same issues on a larger scale, and how each of us constructs a personal identity out of contradictory and at times incomprehensible fragments that are, nonetheless, all that we have.

Uwe Timm’s “facts” are relatively simple. He was born in Hamburg in 1940, a third and last child (a “latecomer,” an “afterthought”). His father, a volunteer in the field artillery in World War I, was a successful taxidermist and served in the Luftwaffe in World War II. The postwar period was difficult. Timm writes, “Overnight the big people, the grown-ups, had shrunk…. There is probably a connection between this impression and the antiauthoritarian student revolution against our parents’ generation” (p. 60). Timm’s grandfather had abandoned his family for another woman. Timm’s father failed as a furrier after World War II and became alcoholic before he died. Timm’s mother said that her husband “was the only man in the world for her” (p. 38), and that “being married was something final, something dependable, a bond that was indissoluble once you had entered into it” (p. 39). Loyalty is an important and sometimes tragic theme in this story.

Timm had two older siblings. His sister, 18 years his senior, “was not the son they wanted.” His father needed “sons to make up for what had gone wrong in his own life” (p. 43). His sister never recovered from the rejection and spent her life searching in vain for a man who would treasure her. Timm’s brother, 16 years his senior, the center of the story, is the enigmatic hero. Six feet tall, fair-haired, blue-eyed, he volunteered for the elite SS Death’s Head Division in December 1942 when Timm was 2, was severely wounded and lost both of his legs in the Ukraine in September 1943, and died the next month. Timm has only a fragmentary memory of being lifted high in the air by him when he visited home from the war, along with family stories, letters home, and a diary his brother kept during the war.

The central theme of the book is Timm’s struggle to forge his own identity by constructing inner representations of his brother and his father. “To get close to them in writing is an attempt to resolve what I had barely held on to in my memory, to find myself again” (p. 14). Were his brother, and his father before him, evil, heroes, strangely blind to the humanity of others, tragically obedient to an insane and violent culture, or all of the above? “How did my brother see himself? What were his feelings? Did he acknowledge anything like personal responsibility, guilt, injustice?” (p. 82). His account invites the reader to speculate further—how do Oedipal themes and sibling rivalry color his thoughts—his father and brother were more masculine, while he was his mother’s favorite. He describes his father taking in and lovingly caring for a veteran who was a double amputee—a proxy for the lost beloved son whom Timm could never replace. His father reworks his memory of the brother; the brother’s notebook included a sketch of a lion:

I suspect that my father improved the sketch later…added some lines and shadings when the notebook was sent to him. He probably wanted to make this little sketch by his son come up to his own expectations and wishes, with some other potential reader in mind. I am that reader. (p. 133)

I compared Timm’s style to that of free association. The reader is left with the kind of impression that an analyst has after a good session, not so much of new facts or new answers but, rather, with a deeper and more nuanced experience of a set of particulars and an enhanced understanding of the meaning of the universal questions that underlie them.

ROBERT MICHELS, M.D.
New York, N.Y.


When I met Irving Sandler briefly in 2004 at a Christmas party at the home of psychiatrist and collector Scott Schwartz, M.D., he asked me whom I liked in modern art and Franz Kline came to mind. This turned out to be the right answer. When I bought this memoir, intrigued by the title (Frank O’Hara called Sandler “the balayeur des artistes” in one of his poems) because I think we psychiatrists are sweepers up after patients, I learned that Kline’s Chief was the “eye opening revelation” (p. 111) that led Sandler to switch from history to modern art criticism. Since then he has been a ubiquitous critic and promoter of abstract expressionist action painting, whose “pantheon” of most admired artists include Willem de Kooning, Philip Guston, Mark Rothko, Ad Reinhardt, David Smith, and Franz Kline.

We get an inside view of many artists from Sandler, their critic and friend, who joined them in their clubs and pubs and heard what they thought they and their fellow artists were intending. For example, Kline’s powerful black strokes derived from his love of the Lehigh Valley locomotives and railroad bridges where he grew up, a fondness for trains I shared without knowing I was perceiving it in his work. And we learn that the proper way to experience a Rothko painting may be in tears. These artists trudged to John Cage concerts they disliked, to support experimental music. Cage told Sandler, “Well, if they didn’t, there’d be nobody there” (p. 36).

About halfway through the book Sandler moves from fondly mounting his artist-specimens to the equally juicy critical wars. “If Harvard-bred WASPs controlled the Museum of Modern Art, heterosexual Jews (the sons of immigrants) dominated writing about avant-garde art in the late 1940s and 1950s” (p. 179). The latter and the artists were formerly Marxist, in search of their identities, moralist, high-art, and anti-kitsch; the former, “aspiring aristocrats,” “would entertain any idea of art the bourgeoisie found shocking or repulsive” (p. 180), even the campy, chichi surrealist, pop, outsider, or

outré. Critics Harold Rosenberg and Clement Greenberg were in the artist-writer camp, yet they had their differences. Rosenberg admired content and emotional discovery but sneered at “apocalyptic wallpaper” (p. 183), and Greenberg favored purely abstract formal qualities, purged of literary content. Sandler “never liked Clem” (p. 185) and also disliked Lincoln Kirstein (while appreciating his bringing Balanchine to America), who launched a “vendetta” against Alfred Barr and the Museum of Modern Art and “was obsessed with discipline and hated freedom” (p. 126). Sandler was “perversely amused” by Hilton Kramer’s “diatribes” (p. 320).

The abstract artists struggled with the fear that their work might be irrelevant to social concerns. A chapter describes the trauma of Vietnam and its aftermath, but the political protest art, like socialist realism and the drug culture art of the 1960s, did not amount to much.

Schildkraut et al. (1) documented the high rate of mood disorders in 15 of the mid-20th-century abstract expressionists of the New York School:

Over 50% of the 15 artists in this group had some form of psychopathology, predominantly mood disorders and preoccupation with death, often compounded by alcohol abuse. At least 40% sought treatment and 20% were hospitalized for psychiatric problems. Two committed suicide; two died in single-vehicle accidents while driving; and two others had fathers who killed themselves. Many of these artists died early deaths, and close to 50% of the group (seven of 15) were dead before the age of 60....By bringing the artist into direct and lonely confrontation with the ultimate existential question, whether to live or to die, depression may have put these artists in touch with the inexplicable mystery that lies at the heart of the “tragic and timeless” art that the Abstract Expressionists aspired to produce. (p. 482)

The conclusion of Schildkraut et al., after taking account of genetic links between depression and creativity (p. 487), is very much in the spirit of Sandler's fleshed-out account of the motivations and foibles of his favorite artists who were responsible for the triumph of American art.

Reference

DAVID V. FORREST, M.D.
New York, N.Y.


This serious book on an important matter forms part of the admirable Cornell Studies in the History of Psychiatry series, edited by Sander L. Gilman and George J. Makari. Armstrong is a classicist, and he has succeeded in writing the most thorough investigation yet on Freud’s relation to the ancient world. Of course, others have noted bits and pieces connected to the problem of Freud’s involvement with the ancient world, but Armstrong is in a position to execute a thorough and comprehensive investigation of many sides to this important question within the history of ideas.

Today’s clinicians will probably be inclined to note aspects of the story that Armstrong skips. For example, the tomb-like atmosphere of Freud’s consulting room was bound to have a suggestive impact on his patients. Certainly the presence of Freud's archological antiquities was unusual enough to prevent his office from becoming anything like a neutral laboratory setting. The atmospherics of Freud’s collection of ancient objects, as well as the books, helped make Freud’s presence a realistic one. Armstrong also does not raise the issue of how psychological clients could collaborate with Freud in using the past as a defense against current problems, pointed out by Carl Jung and others.

Where Armstrong is at his strongest is in dealing with just about every possible classical reference in Freud’s writings. The extensiveness of Freud’s appeal to the ancient world has to be considered remarkable, and the substance of Armstrong’s book securely complements those inclined to see Freud primarily in a biological context. Armstrong is correct in seeing how this fascination of Freud’s with pagan life can be related to both his critique of Christianity and his complicated relation to Judaism; the topic of Freud and religion is not likely to be exhausted in the near future. Armstrong is also especially interesting in understanding Freud within the context of late-19th-century history-writing.

No scholar can be expected now to rely on the entire corpus of Freud studies, but as far as I can tell Armstrong has done a remarkably conscientious job of covering his topic. Intellectual historians will be grateful for this path-breaking humanistic exploration of a subject that has been unduly neglected until now.

PAUL ROAZEN, Ph.D.
Cambridge, Mass.

I’m Your Father, Boy: A Family Memoir of Barbados, by Ezra E.H. Griffith. Tucson, Ariz., Hats Off Press, 2004, 188 pp., $40.95; $15.95 (paper).

Not too long ago Ezra Griffith really knocked my socks off. After a decade of study and writing I had completed my magnum opus (an opus to be sure, and magnum to me) on religion and behavior. A prepublication copy was sent to a host of colleagues whom I respected, hoping for comments that could be used to promote the book. Included on my list was Ezra Griffith because of his seminal studies on the role of religion in African American mental health and because of his reputation as beloved Professor of Psychiatry and of African American Studies at Yale. Much wonderful commentary arrived, but what Dr. Griffith wrote stunned me. Not only did he read the entire book, but he truly understood what I meant to communicate; every jot and tittle. He concluded that my book was written for “those courageous enough to explore their connection—or lack of it—to God and religion.” Holy Toledo! He was fully in synch with what I had written.

On discovering that he had just published a family memoir of his life in Barbados titled I’m Your Father, Boy, I eagerly obtained a copy. After reading it I understood why he comprehended my book so well. His book is about his relationship with his father, Vincent, the pastor of a village church in Barb-
bados whose preaching became famous throughout the island. Before every sermon he intoned a brief prayer asking for God’s blessing: “May the words of my mouth and the meditation of my heart be always acceptable in thy sight, O Lord, my strength and my redeemer.”

As a youngster Vincent had a precocious sense of independence, rebelliousness, and stubbornness. His life changed in the 1930s when he married and converted to Christianity at the end of a week-long revival meeting (such meetings were common on the island). He was part of the Christian Missionary movement, a Caribbean-wide form of Pentecostal church. As he matured, Vincent preached from the pulpit of many denominations. Ezra notes,

Over the years, my most memorable discussions with him would center on our differing interpretations of Biblical passages or our views of what God expected from his followers in certain contexts.... He also constantly monitored where I was in my relationship with his Savior. He thought it the duty of an attentive Barbadian father to verify incessantly whether his son was hewing to the right religious path. (pp. 21–22)

The book contains myriad anecdotes that provide a vivid sense of life on the island. Women selling peanuts and sugar cakes and roast corn and pig’s liver fried hard as nails that took on an “indescribable consistency.” The service-of-song parties that brought people together as the rum flowed, pork chops and chicken were abundant, and participants sang hymns and chanted psalms to their heart’s delight. The philosophical tidbits offered by a woman who helped rear Ezra and taught him that a married Christian woman should never wear pajamas to bed instead of a nightgown. The discovery that for many women pregnancy need not necessarily be linked with marriage. Ezra’s childhood confirmation in the Anglican Church followed by attendance at both Methodist and Moravian churches so that he would get a wider exposure to Bible lessons and learn how to recite poems in public. The freedom that came with owning a bicycle.

For economic reasons, Vincent moved to Brooklyn in 1954 and proudly moved his family there 2 years later. Ezra was introduced to cold weather, apartments that allowed little sunlight to enter, and singing in the choir of his father’s church. He met the bishop of the African Orthodox Church in America, who explained that it was a symbolic rebellion of blacks against the white control of the Episcopal and Anglican churches. Ezra came to realize that the hype of New York City was not realized for many immigrants:

New York is full of those who have been worn out just by the winter alone. It can be a city that tears unashamedly and with vindictiveness at the unprepared who cannot meet the city on its own terms. Mamma was among the unprepared, something my father suspected, although he may not have wished to acknowledge it. (p. 128)

Fortunately the author thrived at Boys High School, although he almost lost his eyesight due to an injury at a soccer game. He was fortunate to encounter two instructors who loved to teach through dialogue before, during, and after class. The fact that neither of these teachers discussed God or religion was puzzling because Vincent harped on one’s relationship with God as a major justification for doing good and doing right. The two teachers seemed to have some other basis for their moral compass, a possibility that my father was not willing to consider, although he couldn’t make it disappear. (p. 133)

The most delightful portrayal in the book is that of the annual Calypso dance held at the Gayheart Ballroom. All the men wore serious Sunday suits, and the women were lovely in evening dresses. Everyone ate fish cakes and drumsticks and ham covered with hot sauce while the dancing was fueled by Mount Gay rum. As the evening progressed the atmosphere became more frenetic, leading up to the last dance, where you could grab onto anything moving, old or young, church girl or not. Church deacons danced in the middle of the dance floor where they couldn’t be seen, and married women made moves under cover of the crowd that only their husbands ever saw before. My father was always sensitive about the business of public behavior. So I never saw him join in what we all used to call this “las’ lap,” this chance to dance the very last calypso for the night. (p. 148)

The book ends with Vincent’s death back in Barbados and a recounting of his views on race, politics, and religion. At his funeral a tape of one of his sermons was played:

It was a spectacular moment. There he was lying in his casket, and his voice, strong and insistent, was filling up this enormous edifice with his pleas that we get right with God and prepare for the next life. He was flat on his back, dead, and he was still able to prick our consciences with his living voice.

At least he was buried in a warm Barbadian earth, while his wife’s body was laid to rest in the freezing ground of a cemetery in Brooklyn. In a sort of postscript, the author puts into perspective the story of his life with his father and laments some of the changes of modern life that essentially work to eradicate the memories that hold him and his father together. All in all, this is a gentle book, neatly constructed and never boring. It is honest, truthful, and lacking in hyperbole or sensationalism, so it will not make the bestseller list. But it is a book that lingers in a reader’s mind like one of Vincent’s favorite Biblical quotes: “The eternal God is thy refuge, and under the everlasting arms” (Deuteronomy 33:27).

ARMANDO R. FAVAZZA, M.D., M.P.H.
Columbia, Mo.


When an Oscar-winning movie actress writes a memoir, most readers expect it to be a kiss-and-tell recounting of dalliances with celebrities. They are not surprised if the exposition includes details (when and how who did what to whom and with which) that would fit well into Popular Mechanics.
magazine. Accordingly, depending on appetite for titillating triviality, a reader will either gorge or pass.

I felt confident that Jane Fonda's autobiography would not fit that description. I anticipated that whatever was described would be as revealing of personality as of salacity and therefore really exciting. So I bought an early copy. I was prepared to find an account of a life that moved from gifted artist to passionate antiwar activist and beyond, but I was completely astonished to find that this volume also has other and major importance.

First, the writing is completely accessible and clear and yet often displays the grace and charm of poetic prose. Fonda's prose particularly favors the accounts of the shooting of the long roster of her marvelous films. Second, the content reveals in full perspective a child trying to surmount the terrible deprivation of a cold distant father and a psychotic mother who committed suicide when Fonda was 11 years old. Fonda discloses how she metamorphosed into a creative and mature woman who loved three husbands, her offspring and theirs, many friends and co-workers, and people in general.

Fonda's life illustrates the deeply ingrained drive of all people who survive into adulthood to try to emulate the caregivers who succored them as babies by later on "making it better" in some way. Jane's tenacious personality, not discounting the later neglect she experienced by both parents, must have stemmed from someone's devoted nurturance in early infancy. This toughness is exemplified in her selection, as a belated transitional object, of her older sister's English saddle, the cleaning and preservation of which she took on when she was 11. The need to make it better did not diminish as she grew older and was transferred onto people—her spouses and their offspring, as well as her own children.

Fonda's account shows how still later this nurturance has become an abiding calling, being invested into helping impoverished young girls and teenagers avoid teen pregnancy. Along the way, this remarkably insightful writer recounts how she developed a lifelong trend to surrender her inner self to please her man, the all-too-common adaptation of women to a world in which male dominance is a near universality. Finally, Fonda shows how with sustained effort this crumbling of personality can be unlearned and relinquished.

In so doing, she beautifully illustrates the rewards of individuation and generativity expressed in Erikson's concepts in *Childhood and Society* (1).

Especially interesting for psychoanalysts is her description of her early psychosexual development of a "tomboy" and "Lone Ranger" identity. Fonda's account shows how this phase, while manifesting phenomena especially stimulated in our "more is better" patriarchal societies (and often dismissed as mere penis envy), eventually was supplanted by another that was enriched by the childbearing and nurturing compensations of femininity along with their sometimes unavoidable sacrifices.

Here are a few brief illustrations, out of scores that are relevant, of Fonda's remarkable native talent in psychodynamic probing. (In the interest of brevity, I'll leave to the reader any further probing.)

In the midst of her "Lone Ranger" identification at age 11:

I was furious that Peter could whip out his little penis and write his name in the snow, so I tried to do the same by taking off my panties and running as fast as I could with my legs far apart trying to spell "Jane" as I peed. Needless to say, it was indecipherable—and I got very cold. (p. 64)

If I felt a boy was cute, he'd be the one I'd beat up.... Teddy, the stable boy, was blond and very cute and ... I...kicked him in the balls....Seemed to me like a perfectly reasonable way to flirt. (p. 68)

Regarding her personality after her mother's suicide:

The kudos I got for appearing strong satisfied a need for approval and locked me into a modus vivendi: Jane the strong one. The shell that formed around my heart served a purpose of keeping me on my feet, but it solidified my superficiality and independence. (p. 73)

About her teenage social-sexual experience:

My classmates started having parties where...post office and spin-the-bottle were de rigueur....I don't remember if I was more scared that someone would "get me" and "try to go too far" or that no one would want to. As other girls became more feminine, I seemed to be a lump of androgyny, always behind trying to catch up. What happened to the girl who saw herself as heroic? She had slipped away so quietly that I never said "Goodbye, see you again in 50 years." (p. 74)

Writing about her decision to have a child, Fonda recounts,

I felt the first wave of nausea. I didn't need a pregnancy test to tell me. I broke into a cold sweat, returned to my car to sit down—and was overcome with a sense of dread. I felt I had to muster all my forces against an unknown terror that seemed to have invaded me. Why? I wanted this....And then I knew: the pregnancy was incontrovertible proof that I actually was a woman—which meant victim. Which meant that I would be destroyed, like my mother. It was one of those strange moments when I was feeling what I was feeling while simultaneously standing outside of myself analyzing the feeling—and being shocked by what it meant. (p. 183)

Whatever may be their personal politics, fair-minded readers cannot fail to be impressed and touched by the determined and persistent struggle this woman has made to become fully her own self. That includes the courageous solitary trip to Vietnam to photograph the terrible damage inflicted by American bombing of dikes that threatened to cause the deaths of hundreds of thousands from starvation and/or drowning. She does not accept as warranted the complete rejection by some critics of the justification for the trip, her feeling that she had the special responsibility of celebrity to unmask lies by the U.S. government. Particularly convincing, however, is her heartfelt self-criticism and apology for the "betrayal" of allowing a photographer to depict her sitting on a North Vietnamese antiaircraft gun rig.

Jane Fonda's autobiography has great illustrative-educational value for professionals in the fields of helping others to find their way to their own maturity. I have recommended it to several psychiatric resident supervisees and psychoanalytic trainees as well as to patients. All have reported that reading it...
has engaged them emotionally with the joys and agonies—and wisdom—of the author and thereby conferred profound personal benefits. Not the least of this book's wonders is that, scattered among many passages that evoke tears, there are others leavening every illuminating chapter with hilarious humor.

Since this inspired book covers only Jane Fonda's life "so far," I am left wondering what creative achievements are yet to come and in what medium they will be expressed.

Reference


RODERIC GORNEY, M.D., Ph.D.

Los Angeles, Calif.

NOVELS


"Writer's writer" is a supreme compliment that contains a trace of pity. The label is reserved for stylists who are admired by their fellow authors and whose work is taught in graduate schools—but otherwise goes unread. I don't know that I've seen a review of James Salter's fiction that fails to mention sentences that are the Salter trademark:

"...from her husband's perspective, in a series of the straightforward sentences that are the Salter trademark:"

She had embarked. My God, he thought, My God. He had known her when she was in her twenties, long-legged and innocent. Now he had slipped her, as in a burial at sea, beneath the flow of time.

Inevitably, something goes awry. The wife does not die. What she discovers the next morning lays bare the hypocrisy of the man. The revelation deprives him of any remaining comfort and justifies the dry conclusion: "That was just the way it was." Those sentences. The temptation, reviewing Salter, is to quote them endlessly:

"Philip married Adele on a day in June. It was cloudy and the wind was blowing. Later the sun came out...They were married in her house, the one she'd gotten in the divorce. All her friends were there. She believed strongly in friendship. The room was crowded.

You can see the author's eye twinkle as he inserts the slight change in viewpoint—"She believed..." It's a trick borrowed from Chekhov, underlining through subtle incongruity. That's the other temptation with Salter—analyzing how it's done.

The same is true when discussing Amy Hempel, another writer's writer. Here, too, the sentences are perfect, the emotional effects are strong—and the pitfall, for a fellow writer, is envy: how does she pull it off? The same men who inhabit Salter's stories drive Hempel's, but her interest is in the women who allow those predators and narcissists to injure them. One victim turns to friends for comfort:

"The women advised long walks. They told the wife to watch the sun rise and set, to look for solace in the natural world, though they admitted there was no comfort to be found in the world and they would all be fools to expect it.

Hempel has found a readership in glossy magazines, in part through a genre called "short shorts." Here is one of her stories, "Memoir," in its entirety:

"Just once in my life—oh, when have I ever wanted anything just once in my life?

The sentence asks to be unpacked. We imagine a woman about to complain, and then taking responsibility, perhaps for repeated disasters—and yet holding fast to pride in her hunger for experience. Repetition is not only compulsion; it is also energy, hope, character, and comedy. Hempel's books, too, deserve wider circulation. The Dog of the Marriage is a good starting point. In the title story, a woman wavers between stances:

"I suppose there are many things one should try not to take personally. An absence of convenient parking, inclement weather, a husband who finds that he loves someone else.

Later, she considers the inclination to self-blame:

"Did I invite this? It is like sitting in prayers at school when the headmistress says "Who dropped lunch bags on the hockey field?" and although you went home for lunch, you think, I did, I did."
These two collections have an odd resonance. Hempel discusses cancer. Salter builds a narrative around a woman's attachment to a dog. Dogs are Hempel's signature detail. When she is not writing or teaching, she trains guide dogs, and she brings mention of dogs into every essay or book. Dogs stand for decency in a world half-peopled, almost, by James Salter's males.

*The Dog of the Marriage* displays the range possible in short fiction devoted a single topic—sensitive women and their social surround. One bravura story has the form of a protest, over an unfair ticket, to the New York City Parking Violations Bureau. Although she has coped with some of life's major insults, the complainant does not take minor ones in stride. The letter ends:

I want what is fair. I don't want a fight. But the truth is, I'm shaking—right now, writing this letter. My hand is shaking while I write. It's saying what I can't say—this is the way I say it.

This is the way we say it, whether self-confident man or self-aware woman: in short, declarative sentences, with words of one or two syllables, adjectives and adverbs all but banned.

**Reference**


Peter D. Kramer, M.D.

**Amador, by Fernando Savater; translated from the Spanish by Alistair Reid. New York, Henry Holt & Co., 1994, 190 pp.**

After 8 years of college, medical school, and internship, I began my residency in psychiatry in 1953 having never taken a single course in psychology! While that pedagogic travesty has been remedied, another of near-equal importance has not. Psychiatrists rarely study ethics; yet our clinical work is perfused with ethical concerns.

Psychopathology defines mental illness, but mental health is more than the absence of illness, and what makes for a good life and how to achieve it is the topic of ethics. It would also seem that ethics and certain psychopathologies bear an inverse relation to each other, better knowledge of which might throw light on both.

Fernando Savater is the Spanish academician who, in 1991, turned his effort to introduce his 15-year-old son to the subject of ethics into a book that sold 150,000 copies in Spain (the equivalent of a million copies in the United States) and remains in print around the world (although not in the United States). It is no small literary feat to present time-worn principles into a bestseller. The reader is welcomed into the world of ethics by an invitation to think (an activity that strongly favors ethical behavior) along some basic lines of inquiry: What is the importance of choice, and the relation between freedom of choice and responsibility? What is the difference between selfishness and taking proper care of oneself? What is the core of a moral compass and what guides us to it? And what are human beings for anyway? Each chapter addresses such a question and concludes with related quotations from philosophy, literature, or the movies.

Savater writes in the form of letters to his son; thus the book is a primer on ethics written in an engaging manner and format suitable for the serious beginner. He projects an intimacy with both his subject and his audience as he starts off with a deceptively simple assertion: The key to a good life is to know what suits you, and do it. “Do what pleases you” is a refrain that permeates the book, but Savater quickly thickens the soup by showing that what distinguishes humans from other creatures is our ability to think and to make choices. In fact, making choices is a necessity, for we cannot not choose, and in the harsh words of Sartre, “We are condemned to freedom.” Since we are inexorably responsible for the consequences we create, it serves us best to choose with deliberation and care, that is, thoughtfully. We must take our freedom seriously in order to use and preserve it, for it is the possession that distinguishes us from everything else in the universe.

How do we find what suits us? We must notice that some things make us feel good while others do not, and often there are legitimate and serious conflicts between our different needs and wants that we must think about in order to make good decisions. Then we must attend to consequences and remember them so that we can learn from our mistakes. The protagonist of *Citizen Kane* surrounded himself with material possessions, but in the end he longed for Rosebud and the sentiments that attended his childhood sled. Savater concludes that the most basic thing about humans is not the drive to accumulate material possessions but the need for human relationships. Value and interest in life derive ultimately from human interactions, and seeking material wealth in order to win love enslaves us and compromises our freedom.

In connection with doing what suits us, since conscience is part of our being, treating others poorly causes remorse. Conversely, nothing gives more pleasure than winning appreciation and love through actions taken with a good heart. If others take advantage of our goodwill we must protect ourselves, but if we react vindictively we recycle the damage. On the other hand, if we react to bad treatment with forgiveness and forbearance we have a shot at breaking the cycle of negativity—that is, doing good. Antagonism begets antagonism, mistrust begets mistrust, and it is easy to see the rewards of decency and a healthy respect for the dignity of each individual. We have more in common with the worst criminal—a common humanity—than with any nonhuman animal or material thing, and we share the potential to influence that person and to be influenced by him. Forgiveness and generosity move mountains.

To show how these ethical principles apply, Savater focuses on two of life's major playing fields: sex and politics. He reserves his harshest criticism for Puritans who equate sexual pleasure with immorality, and he asks, “If we do not provide for the satisfaction of the needs of our bodies how can we possibly have any kind of good life, for we are our bodies.” The giving and taking of pleasure is essential to our well-being, and there is neither shame nor harm in taking pleasure that is at no one's expense. But the Puritans seem able to take pleasure only in depriving others of theirs, and they hold that suffering is evidence of living rightly. Rubbish! says Savater. It is sometimes necessary to bear suffering, but it is never a virtue.
in and of itself. Virtue is doing good for others, and so Puritanism is about as opposite as can be from an ethical view of life.

Savater commends all pleasure but warns that excessive attention to one pleasure over all others can prove a distraction from the complexities of life. He calls for a balance in all things and temperament, which is an intelligent friendship with all forms of pleasure. Ethics is not about scrupulosity.

Regarding politics, it follows that the goal of all governing is to serve and preserve the dignity and freedom of each individual while safeguarding the common good. Individuals and groups who would undermine the common good must be opposed, but it is crucial in opposing them not to be corrupted by them. When we compromise our own ethics in dealing with criminals (or terrorists) we lose our battle against them, for we have then allowed them insidiously to invade us. It is one thing to protect ourselves from harm, but something quite else to deprive someone of his humanity and dignity in the service of protecting ourselves.

The ideal political system is one with minimal restraints on the freedom of individuals; one that treats all people with dignity, respect, and equality; one that provides help for those who need it, with equitable and intelligent distribution of resources (“war no more” is a given); and one with respect for the planet we inhabit.

Returning to the question of the relationship between ethics and psychiatry, in my own clinical work I have come to see that dishonesty, broken promises, and irresponsibility are always self-defeating, whatever the underlying psychopathology. Lies undermine the trust essential to any good relationship, and varieties of inconsideration are manifold. In treatment I can often make a tacit confrontation in a nonjudgmental clinical context that enables patients to see that they are unwittingly engaging in behavior which neither suits nor serves them, and seeing that leads to relational repair and clinical improvement. Contrition heals wounds.

Savater’s counsel to find what suits you and do it elaborates on the idea that there is a self to which one needs to be true—a True Self. Here, ethics and psychotherapy share common ground, for we always aim to improve patients’ ability to be honest and accepting of themselves and others. Familiarity with ethics enlightens that path. Should ethics ever become insinuated into psychiatry, then DSM-VI or DSM-IX might include “relational psychopathology,” which would address a whole range of interactional considerations like selfishness and abuses of power, to expand our nosology and techniques of treatment.

In his epilogue, Savater advises his son to take the words of this book seriously, but not too seriously:

Seriousness is not an unequivocal sign of wisdom, as serious people believe, and intelligence must know how to laugh. I have no magic words for you. Each of us must find our own meanings and create our own lives.

Everyone could use a father, or psychiatrist, like Savater, who would express these thoughts:

Seek out and think for yourself, in full freedom, responsibly, with no tricks. Choose what opens things for you, to other people and to new varieties of experience.

Avoid what encloses you and buries you. Keep your nerve! Have confidence in yourself! Good Luck!

Thanks, Fernando.

JUSTIN SIMON, M.D.
Berkeley, Calif.


It’s a fairly safe bet that many psychiatrists take themselves too seriously. This overly serious demeanor may even extend to their choice of reading matter. In fact, some may feel that when granted some rest and recreation during the holidays, they should spend their time with the likes of Henry James, A.S. Byatt, or (this may be an exaggeration) Thomas Pynchon. But guilty pleasures deserve a place on the bedside tables of our somber colleagues as well. Good storytelling stripped of lofty ambitions has its place and should warrant a visit to one’s favorite bookstore or web site no less than the latest annotated version of Ulysses.

Carl Hiaasen’s 11th novel, Skinny Dip, may provide pleasant diversion from unpleasant holiday houseguests who threaten to stir up ancient intrafamilial conflicts. We meet Chaz Perrone, a priapic marine biologist whose primary interest lies in scamming federal environmental agencies. Chaz takes bogus water samples in the endangered Florida Everglades and finds them surprisingly pollutant-free, in exchange for which he receives handsome payoffs from a thoroughly corrupt (and over-the-top) entrepreneur. Chaz is not troubled by pangs of conscience about his sullying of nature, however, because he just can’t understand all the fuss about the Everglades. Why would anyone care about a breeding ground for vermin ranging from the pesky (mosquitoes) to the dangerous (alligators and water moccasins).

Chaz fears that Joey Perrone, his rich and ravishing wife, knows a little too much about his scam and concludes that he has no alternative but to eliminate her. He tosses her overboard after a romantic evening on a cruise ship. Not known for thinking through the consequences of his actions, Chaz forgets that his wife was a champion swimmer in college. By turning the fall off the deck into a dive and swimming to a nearby island, Joey survives the murder attempt with the help of a handsome stranger who is at the right place at the right time.

Much of the remaining story is a tale of revenge and revolves around how Joey can terrorize her husband without letting him know that she is still alive. Some hilarious and improbable action ensues that will keep all but the most prudish reader amused and engaged. Colorful characters stroll across the stage as the narrative moves at breakneck speed toward its denouement. Hiaasen combines the noirish crime genre of Michael Connelly with the outrageous sense of humor of Dave Barry—not a bad combination for holiday reading at 35,000 feet during a transcontinental flight.

Hiaasen’s effort to incorporate a political message about saving the Everglades is a bit misplaced in this romp of a novel. Also, the poetic justice meted out in the end flies in the face of the basic axiom that life is unfair. Nevertheless, readers...
who refuse to think too hard are in for a satisfying—if instantly forgettable—ride.

GLEN O. GABBARD, M.D.
Houston, Tex.


As a biological virus infects a cell, producing copies of itself, a computer virus is a self-replicating program that attaches itself to a program within the hardware of a computer’s operating system and takes it over for its own purposes. Some viruses are delayed and act as a bomb, either a time-bomb that comes out at given moment, or a logic-bomb, which comes out when the computer executes a given operation. The computer virus analogy is complete, in that it infects computers through their e-mail connections, just as a human virus is spread through any entrance to the body. A virus is part of another program, but a worm, on the other hand, is a self-contained program that can exploit the phenomenon of transmission. Thus the name of this fascinating novel about a computer programmer from India who escapes to Silicon Valley from his cramped apartment in New Delhi.

Arjun Mehta, one of millions of aspiring technocrats clamoring to fill the thousands of positions generated by the computer revolution, received his Bachelor of Science at North Okhla Institute of Technology. He succeeds in winning his dream job after meeting an uncouth and supercilious headhunter at a job fair for U.S. firms. He is transported from his Third World Warren to the paradise of the West Coast of the United States, where Silicon Valley is booming with instant millionaires of the information age.

Thomas Friedman, of the New York Times, expounded a thesis that the Internet has resulted in globalization and the leveling of the playing field for technology (1). An Indian in Bangalore, a technician in Shanghai, and a computer hacker in Moscow are equal when it comes to access to information and technology. When a student at Yale calls the help line to deal with a problem on his laptop, he may be speaking to a consultant in India who has been trained to speak in an American idiom.

When a small hospital in the Midwest cannot find a local radiologist to be on-call in the night, it can transmit magnetic resonance images of the brain to a radiologist on the Asian subcontinent to get an instant reading for a trauma patient in the emergency room. Of course, no one knows the credentials of that radiologist, where he trained, if she is board-certified (although he may have gone to Harvard), or if she has a medical license. Electronic transfers of money from financial institutions and brokerages flash around the world faster than optical license. Electronic transfers of money from financial institutions and brokerages flash around the world faster than gold can be moved from one holding area to another in vaults beneath the streets of the city of Zurich. Scientific discoveries may be posted, reviewed, replicated, or refuted with the speed of a keystroke from M.I.T. to Berkeley to Taipei.

This amazing electronic revolution has grown exponentially faster than the opening of the New World by the sailing ship, the opening of the American West by the railroad, or the destruction of the Iron Curtain by MTV and CNN. The children of the current generation are more likely to turn to Wikipedia.com than to Encyclopedia Britannica in the library. Wikipedia allows for postings and modification by anyone, allowing the billions of people on earth to share information, much like the billions of neurons in the brain that connect through synapses. Our computer servers and web-cams and e-mails interconnect us into one huge fabric of humanity.

This entire electronic network makes one wonder if Lewis Thomas (2) was right when he suggested that human beings may simply be cells in a larger organism and that humanity’s delusion of individuality is only a dream. Science fiction is no longer a genre less than “literature” but has become mainstream. As Jules Verne predicted space travel, William Long’s Seetee Ship (3) predicted antimatter, and Jonathan Brunner’s The Shockwave Rider (4) anticipated and named computer viruses, it was inevitable that someone would write a novel about destructive software (malware).

Hari Kunzru is the highly acclaimed author of The Impressionist (5), also a novel about a man from India, a Dickensian journey reminiscent of Oliver Twist or Great Expectations, which was short-listed for a number of international prizes. Transmission, on the other hand, is as up-to-date as an iPod and more amusing than TiVo. One need not be a computer nerd to be swept along by Kunzru’s skewering prose. No knowledge of programming is necessary for the reader to become immersed in pop culture and social satire. Arjun Mehta’s team leader observes,

What my team has come to realize is that in the 21st century the border is not just a line on the earth anymore. It is so much more than that. It is about status. It is about opportunity. Sure, you are either inside or outside, but you could be on the inside and still be outside, right?

Transmission, of course, refers to a computer worm, which Mehta releases on the world in a misguided attempt to save his job. He generates a destructive virus in the hope of achieving acclaim by eliminating it and thereby proving his worth to his employers. Like the man who introduced the rabbit to Australia, Mehta falls victim to the law of unintended consequences when his computer worm has an Indian movie actress dancing on computer screens the world over.

This is a love story, a social satire, and a book about justice, vengeance, and fate. There is a journalistic quality to the prose, and the dialog may have been overheard from a cell phone conversation in a public place. Kunzru knows how to tell a story that is entertaining yet poignant, full of psychological insight. The Indian characters, the California characters, and the fantasy land of Indian movie making in “Bollywood” are all expertly drawn and meshed together like the gears in the latest BMW. It would not be surprising if the reader hesitates a moment before opening his next e-mail attachment after reading this exciting novel.

Kunzru has succeeded in writing a technological love story wrapped in a cloak-and-dagger thriller, a style reminiscent of Ludlum, with the wit of Sedaris or Celine. The computer revolution is not about computing, it is about communication. It is about communication as commerce, which is now less about production than about distribution. Hari Kunzru has raised the comic novel to its psychological endpoint in this book in which the entertainment is wrapped in elegant prose. I would not presume to predict the topic of his next novel, but I look forward to reading it.

Dr. Paul Lucas is a tormented soul. The protagonist of Roderick Anscombe's third novel, The Interview Room, suffers from posttraumatic nightmares replaying the fatal car accident that claimed the life of his toddler son only a year ago. His lovely wife, Abby—a social worker who declares she is not ready to surrender her grief—has become remote and secretive. His chairman calls to give him a heads-up: Craig Cavannah, the grandson of a superwealthy financier (he's on the hospital's board of trustees and has donated its Cavannah Pavilion) will be admitted to Paul's unit at the Sanders Institute, the forensic facility run by the hospital.

Craig's crime is simple. He's a Harvard undergraduate who's been stalking Natalie Davis, a teaching assistant for Craig's creative writing class, and he's violated a restraining order. Before he heads to Sanders Institute to interview Craig, Dr. Lucas meets Natalie—who looks a bit like Abby and is just a little flirtatious—to hear her side of the story. He then proceeds to Sanders, and his sessions with Craig begin in Interview Room One. The sessions are filled with tension, and it's never quite clear who's interviewing whom; their interactions are reminiscent of those between Clarice and Dr. Lecter, and there are clear who's interviewing whom; their interactions are reminiscent of those between Clarice and Dr. Lecter, and there are moments when it seems Craig has the upper hand and Dr. Lucas just might crumble. Finally, Craig's commitment hearing comes, and the judge agrees to release him—aft all, he's only suffering from an adjustment disorder with mixed emotional features—and condition that he continue in outpatient therapy with Dr. Lucas.

From here, all hell breaks loose. Craig, now free in society, is able to violate every boundary. He deftly weaves his way in and out of Dr. Lucas' life, and the therapy sessions are a series of mind games. Although Dr. Lucas is proclaimed as the world's expert on lying, it isn't clear how he will conquer Craig, and the voice of Dr. Lucas lends insights unique to thriller fiction. His portrayal of the stalker's tenacity is chilling. Worth the sleep you'll lose staying up to finish it.

As a quick read with a self-propelled plot, The Interview Room is the perfect "vacation read." That's not to say, however, that the novel is fluff. Its author, Roderick Anscombe, is a forensic psychiatrist at Bridgewater State Hospital, and the story is told with calculating, lyrical prose and meticulous detail to tone, character, and setting. The Interview Room is not as elegant a work; the writing is more clipped and less detailed, and everything about this book moves at a faster pace designed to entertain. Anscombe is a gifted, versatile writer whose literary voice glides in a variety of genres.

References

JAMES R. MERIKANGAS, M.D.
Washington, D.C.

DINAH MILLER, M.D.
Baltimore, Md.
the scale of New York’s Central Park. Every vista of this garden and the particular beauty of its brooks, mountains, etc., are described in unsparing detail. An exact reproduction of that fictional garden has been built in Shanghai and elsewhere in China. But I confess that the beauty of the language describing the beauty of the garden did not hold my interest, despite the fact that there is every reason to believe that the translation from Chinese to English is superb.

For those like myself who are psychologically minded in their approach to literature, the problem is that one meets hundreds of characters, most of whom are two-dimensional. The two young teenage girls who play central roles, Lin Daiyu and Xue Bochai, are said to be archetypal females, antithetical to each other. Devoted readers of the novel debate which they prefer. But they seemed imaginary creatures, not real humans, to me. People whose opinions I respect and who know much more of Chinese culture than I do assure me that I have missed the greatness of the novel and that if I can get past the second volume I will be in literary Nirvana. I am sure they are correct, but I would nonetheless advise even the most ambitious readers to borrow a library copy of the first volume and read it before ordering the complete set as I did.

ALAN A. STONE, M.D.
Cambridge, Mass.


The question of why tyrants develop as they do has never been satisfactorily answered. The extensive published scrutiny provided the 20th century’s most murderous despots seems yet to describe an ontogeny sufficient to our understanding of their bizarre motivations and behaviors. Details of their personal lives have been largely, and often deliberately, obscured, and any sense of personal intimacy with these social aberrants has eluded almost all of their chroniclers. This particular biography does quite a bit better than its predecessors on the subject of one Joseph Vissarionovich Djugashvili. It is not fully successful in illuminating the darkest of places within the most dangerous of our species, but it notably doesn’t content itself with the usual clichés like “enigmatic” and “satanic” with which such men are often despairingly summarized.

Stalin may be responsible for more intentional human slaughter than any single person in recorded history. The “Man of Steel” rose to succeed Lenin in 1929, ruled the Soviet Union for 24 years, and died still in office in 1953. From early in his revolutionary career he was careful to hide his intellectual leanings from some of the Bolshevik leadership, including his own revolutionary friends and colleagues, his “comrades.” But he was even more devastating in the Russian countryside, where millions of peasants, farmers, and laborers, the people from whom he came (his father was a cobbler), starved and fell to enforcers of his policies of collectivization. After World War II, uncounted others were cut down as Stalin continued to murder, assassinate, and imprison in his slave-labor camps to consolidate his power within the Soviet Union and the states abandoned by the conquering Allies to his absolute control.

This is a thoroughly researched and masterfully conceived work. Simon Montefiore is able to find in Stalin “a more understandable and intimate character, if no less repellent.” Mining previously unrecovered archives, memoirs, personal letters, family interviews, and testimony from the dictator’s inner circle, Montefiore considers Stalin as a son, husband, and father, and tries to locate a credible personality within the public persona. The intimate access to Stalin’s life and era one feels is sometimes exhilarating, but the enormity of the coldly determined cruelty ultimately becomes numbing, and the vital question of how a man becomes motivated to destroy millions, finally remains a mystery. No one yet knows many details of Stalin’s early life. He apparently saw to that, as he manipulated historical records, even altered photographs before it was easy, to serve what he thought was his advantage throughout his career. Whipped as a child? Raised by wolves? Brain-washed by crazed religious sects at school? Montefiore is better than many biographers in resisting the urge to write fiction when he doesn’t know the facts, so we won’t learn the answers to those questions here. He is able to tell us that Stalin was a mutter-kind (like Hitler), an Orthodox seminarian, a scholarly intellectual, a dedicated revolutionary and disciple of Lenin, and a skilled politician who ascended the pyramids of Bolshevik leadership and matured from a canny apparatchik to a tyrant of terrible malignity. When he’s writing about things the surviving record can support, Montefiore has a lot of fascinating things to say. He is an excellent historian.

Stalin: In the Court of the Red Tsar begins at a 1932 party in the Kremlin apartment Stalin shared with his second wife, Nadya. The couple and their guests were celebrating the anniversary of the great revolution. Before dawn, Nadya was dead, a suicide. All surviving accounts agree that Stalin was devastated. Always secretive and suspicious, he became even more emotionally isolated, bitter, and suspicious. With his indelible memory, especially for slights and insults, he would later make those colleagues and their wives whom he believed were somehow involved in Nadya’s death pay with their own lives. Montefiore proceeds to bring us into Stalin’s life: into his office, apartment, dacha, committee rooms, war councils, politburo sessions, meetings with heads of state, hunting and fishing trips, private railroad cars, command bunkers, and then into the room where he died. The text is amply laced with conversation and is richly anecdotal. Though Stalin believed himself an intellectual and read voraciously (he boasted of 500 pages a day), early in his revolutionary career he was careful to hide his intellectual leanings from some of the coarser, less educated Bolsheviks whose support he needed on his way to the top. When he got there, he commandeered the role of sole arbiter of Soviet literary, musical, and artistic tastes. Montefiore never betrays admiration for Stalin, but he...
does recognize and detail the man's astonishing capacity for hard work in leading a backward and crippled country, kicking and screaming, into the 20th century: "Stalin was already famous for his Sphinxian inscrutability and phlegmatic modesty... Far from being the colourless bureaucratic mediocrity disdained by Trotsky, the real Stalin was an energetic and vainglorious melodramatist who was exceptional in every way." His ascension to power drank from the same well: "No one alive was more suited to the conspiratorial intrigues, theoretical runes, murderous dogmatism and inhuman sternness of Lenin's Party."

Montefiore's anecdotes and metaphorical snapshots study the working parts of Stalin's seductive personal charm and how he used it to manipulate colleagues and world leaders throughout his career. He was a formidable politician by any standard. He was also unimaginably crude, always the peasant beneath the costumes of power. Traveling on a lonely and unpaved Russian back road, he ordered his driver to stop the Zis limousine, dismounted, unbuckled, squatted, and unburdened himself on the side of the road. He joined in singing and enjoyed entertaining visitors with his fine tenor voice. He also loved to humiliate his party guests, especially the Magnates and visiting communist officials, by commanding performance of the high-kicking peasant dances that their well-fed figures just couldn't manage any more. He was a living parody of the mad dictator.

But the inescapable theme of this work is Stalin's retributive and arbitrary cruelty. The exploits of his secret police chiefs, appalling sadists who tortured and murdered as much for their own amusement as for Stalinist policies, will sicken and infuriate most everyone who reads about them here. Torture in the infamous Lubianka Prison was prolonged and vicious. Eyeballs were left hanging from their sockets as "confessions" were routinely beaten from the hapless innocent. Lavarentia Beria, architect, husband, father, murderer, sadist, rapist, and philanderer, was the last and worst of the legendary secret police directors; his predecessors had all been dispatched, many by him, as they inevitably became troublesome.

Montefiore tells a fine tale of Stalin in his last days, commanding a new "Terror" with the Anti-Semitic Doctors' Plot, arbitrarily imprisoning some of the country's finest scientists, artists, and physicians and ordering the secret police to "Beat, beat, beat, and beat again." On the evening of Sunday, March 1, 1953, he was found paralyzed and semiconscious on the floor of his dacha. The devouring terror he had generated in every will, his co-workers redounded in delicious irony, as no one dared to put caring hands upon him for fear of implication in this ultimate horror. The country's best doctors were all behind bars, and Stalin was left virtually untreated on a sofa for the several days it took him to die.

This intimate portrait of Stalin is also a portrayal of the excesses and corruption of unchallenged power. Though we still do not know all the psychological and social influences that shaped this shoemaker's son into one of history's greatest monsters, this consummate exposition will confirm for many of us the intuitive intelligence of our Founding Fathers and the care they took to protect themselves and us from the worst excesses of governmental power. This is what can happen, and this is what did happen. If for no other reason than to help us keep our political faith Montefiore's opus is a rewarding and edifying experience.

WILLIAM EDWIN FANN, M.D.
Houston, Tex.


This is a thought-provoking book, both because and in spite of itself. The author seeks to explain "the dramatic shift away from social concern and toward competitive self-interest that occurred during the closing decades of the 20th Century" (p. 257). He feels that "the same tools and technologies that have enabled America to achieve Adam Smith's 'universal opulence' have also compromised the social anchors" (p. 258). He sees our challenge as "transforming those pleasures into mass happiness" and "forging from the affluence of commercial success a balanced and equitable society" (p. 263).

Whybrow, himself a British immigrant, advances the anthroposophical hypothesis that we Americans are a nation of "migrants" who are equipped not just with the selfish genes of all biology but also with the restless, adventurous genes of those who have risked all to seek a new land of opportunity. The book's argument relies on the work of Chuan Sheng Chen and his co-workers (1), who correlated a greater percentage of certain dopamine-4 alleles, specifically 7-repeats and long alleles (5- to 11-repeats), with exploratory human nature—novelty-seeking personality, hyperactivity, and risk-taking—in human groups who have "macro-migrated." The 4-repeats, according to this argument, are more prevalent in sedentary groups. It is the opinion of Chen (whose name is misspelled) and his co-workers that the gene variation is the natural selective result rather than the cause of migratory culture over thousands of years and that recent U.S. immigrants (Japanese, Chinese, and Europeans) do not show the effect. The effect is seen, for example, in Colombian Indians who migrated far across the Bering land bridge 10,000 and 5,000 years ago. Granted, North American Native Americans, who are rich in 7-repeats, have contributed to the casino bubble, but this is not necessarily a sign of risk-taking, because this business has not proved very risky for them and, in any case, they have had help from other Americans of European ancestry. The savings and loan and Internet collapses, caused by risk-taking, were associated with Americans of European ancestry, who have the same number of repeats and alleles as their forebears.

The book has a booming tone, as though it were written standing at a lectern with a feather pen. In keeping with the author's admiration of de Tocqueville, he seems to view Americans at a remove that is mercifully relaxed only in discussing his interviewees. Whybrow argues that highly migrant people are novelty seekers, restless and optimistic risk-takers, and poor farmers. An overload of social stimulation engages and "hijacks" the same dopamine "superhighways" as caffeine, nicotine, cocaine, and amphetamines (p. 93), with the result that the United States bubbles over with unusual "irrational exuberance" and "infectious greed" (in Federal Reserve Board Chairman Alan Greenspan's words [p. 127]), which Whybrow likens to hypomania and florid mania. This is aggravated by "globalization and greed," resulting in a "manic society" in which...
more is never enough and time pressures increase. As examples he parses the Internet bubble of 2000 as if the United States were going through the phases of a manic illness—namely, optimism, hypomania, full flower, and bubble bursting. In America, failure is no shame, only not to try is.

“Turbocapitalism” (p. 187) and “a giant casino” (p. 240) have resulted, with the growth of megacorporations like Wal-Mart, which accounts for 6% of American retail sales, undermining the cohesive and supportive personal social microcosms of the rural villages with their butchers and bankers in a Faustian exchange that substitutes the impersonality of “multinational companies with little accountability” (p. 255). Whybrow amply illustrates the loneliness and isolation that result and bemoans Thomas Jefferson's substituting the “pursuit of happiness” for “property” in the Declaration of Independence. As one of Whybrow’s interviewees puts it, “Happiness cannot be pursued. Happiness is something that wells up inside” (pp. 46, 240). This book has rich chapters on the evolutionary psychobiology of American malnurtient obesity and our sleep-indebtedness to our suprachiasmatic nuclei. Whybrow notes that coffee is second only to oil as a traded commodity (p. 162).

Fearing that as a New Yorker I am unqualified to critique this book because manic time pressures seem normal here, I refer the reader to Jacobs and Gerson (2), who showed that the temporal pressure of staying long at work is confined to us professionals and that the national uptick in annual work hours is the result of more women working. Thomas Friedman (3), the New York Times correspondent, argues that with globalization, the “flattening” of the world economy has allowed India, China, and many other countries to join in the global supply chain of manufacturing and services, which will require us to run even faster to stay in place than Dr. Whybrow notes we are already running. In my opinion, the best evidence for Whybrow’s case is the multiplication of slot machines, the growth of Las Vegas, and the meretricious, pseudoreligious sanctimony of the new sanitized, family-friendly casinos.

Whybrow’s book is provocative in spite of itself because it not only revisits the possibility of psychodiagnosis in the axis of culture, venturing beyond the pieties of anthropology and identity politics that all cultures are equally beneficial, but also challenges us to reconsider the inadequacy of our psychiatric descriptors for that task. Identifying a culture as manic or paranoid or obsessive or dyssocial is somewhat informative, but we could seek better concepts and terms for a culture’s role in psychogenesis. Kardiner (4), attempting to avoid the idea of national character, proposed a basic personality as the repository of a society’s values, from which individual character differentiates. Specifying the concepts and nomenclature awaits our ingenuity.

References

DAVID V. FORREST, M.D.
New York, N.Y.


When I took my children on their first grand tour of Europe around 1970, I insisted they go over a book by Frederick Hartt called History of Italian Renaissance Art (1). At that time I considered it the best available introduction to Renaissance art. Now I might go about it differently because there is a spectacularly good series of taped lectures by William Kloss put out by The Teaching Company (2), far the best introduction to the artists of the Italian Renaissance. So today I would have my children start with that and then move on to the book under review here, which is Hartt’s book in its fifth edition and is without a doubt the basic text for this subject.

David G. Wilkins, Professor of the History of Art and Architecture at the University of Pittsburgh, revised the fourth edition, and in this fifth edition he adds more color reproductions and some new works to expand and enhance the late Professor Hartt’s original vision but still retains the organization of the text. Wilkins adds secular works to Hartt’s emphasis on religious art and also adds a series of portraits of important patrons and personalities of the period as well as some extracts from Renaissance texts. The book is logically divided into three parts: the Late Middle Ages, the Quattrocento, and the Cinquecento. Hartt’s model was the classic by Vasari (1511–1574), Lives of the Artists, but he also attempts to unfold the story in an interesting and integrated way. I cannot praise this book highly enough; it is a work of art in itself. It belongs in the library of every educated person.

The fifth edition is in marked contrast to the first edition, in which the photos were in black and white. An example of the wonderful changes from the first to the fifth edition is evident immediately on opening the latter, which begins with what is called a “Portfolio,” containing some magnificent color pictures of Florence and several other Italian cities and sites prominent in Renaissance art. We are then offered a “Prelude” that consists of chapter 1 of the first edition, titled “Italy and Italian Art,” sensibly separated out by Wilkins from the first section of the earlier book, which deals with the late Middle Ages. Subsequent chapters in the book have subsections on painting, sculpture, and architecture. The addition of many colored illustrations enhances what Hartt first attempted to present in his initial edition, even though Hartt’s written text is mostly preserved.

Of course we do not really have a good explanation for the incredible upsurge of painting that took place at the time of the early Trecento. Cimabue, who was praised by Vasari as a great painter, appeared just before that period. However, Hartt declares, “In reality Cimabue belongs not at the beginning of a development but at the end. His altarpieces show him to be the last Italo-Byzantine painter” (p. 70). This runs against the popular conception of Cimabue and surprised me; I cannot say that I am in agreement with it. The ultimate test is to look at Cimabue’s paintings. Of course, Hartt still considers him a great painter.
I was fascinated by the reproduction, included at this point, of a painting by Cavallini done about 1290 showing the first interest in the function of light to realize form, which Hartt considers “a fundamental revolution in artistic vision” (p. 77). For Hartt, the centerpiece of the opening of Renaissance art is the work of Giotto. We are told that his best preserved works are in the Scrovegni chapel, sometimes called the Arena chapel because it was built over a Roman arena. (This wonderful chapel is not always open to tourists; when I took my grandchildren to see it in 2001 we arrived in Padua only to find it closed for “renovation.”) It is interesting that Hartt refers to the rivalry between Cimabue and Giotto cited by Dante in the Divine Comedy (xi, 94–96). Dante makes it clear that Giotto stole Cimabue’s fame, and, according to Hartt, Dante’s statement is literally true. It is frightening to realize that in World War II the Arena chapel was almost destroyed by an Allied bomb that fell close to it and left a big crater.

One of the great assets of this textbook is that the pictures referred to in the text are either on or very close to the page of the text in which they are discussed, and the colored pictures in this fifth edition are really wonderful. There is a leisurely tone to the text and a careful examination of a great number of magnificent pictures, which, if closely followed, could lead to a great improvement in the reader’s capacity to appreciate and analyze a painting.

How odd it is that the Early Renaissance, the 15th century, “was an era of bitter conflict and of challenges never more than partly met; seldom, however, in history is the gap between human problems and their solutions more evident” (p. 181). Throughout that time Florence, for example, was in an atmosphere of crisis and in a stalemate with enemy forces on many sides of the city; yet the great advances in Renaissance art were made at the same period. Hartt writes,

The Florentine humanists thought that geometric principles could unlock mysteries at the heart of the universe and reveal the intentions of a God who was, if one only knew how to go about it, eminently understandable and had created the universe for human enjoyment. (p. 195)

It is hard to reconcile this statement with the fact that Florence was under military siege for a long period of time. Many of the 15th-century Florentine sculptures, Hartt points out, represent propaganda for the Republic of Florence. Although the statues produce ideals of the virtues demanded in a crisis that it was hoped the populace would develop and follow, according to Kloss, “More significant artists were born and more epochal art created in Italy in the 15th century than in any comparable place and century in the history of art” (2, part 2, p. 1). This all seems a dramatic mystery to me.

One of the most interesting sections of the book moves from Masaccio to his contemporaries Fra Angelico and Fra Filippo Lippi and on to those painters who, “with Alberti embody the ideals of the second Renaissance style” (p. 293): Paolo Uccello, Domenico Veneziano, Andrea del Castagno, and, most remarkable of them all, Piero della Francesca. Hartt threads the narrative together very nicely and continues in this fashion all the way to the end of the book, which, appropriately, closes with examination of some of the famous mannerist painters of the 16th century.

The final third of the Quattrocento was dominated by five artists, Antonio del Pollaiuolo, Andrea del Verrocchio, Sandro Botticelli, Filippino Lippi, and Domenico del Ghirlandaio (p. 359). Hartt discusses them as well as all the well-known and the not so well-known Italian renaissance painters and their work, in addition to the sculpture and architecture of their time. Because there are so many painters, sculptors, and architects mentioned, it would have been helpful to the reader to include a much more detailed table of contents, and I hope the sixth edition of this book will contain one.

Piero is my favorite mid-Quattrocento painter. I agree with Hartt’s description of Piero della Francesca: “The artist who seems to us today to fill the Albertian ideal of absolute and perfect painting in nearly every respect” (p. 310). I remember standing for a long time in front of Piero’s Flagellation of Christ (reproduced in this book on page 322) hanging in the palace of the paradoxical Duke Federico da Montefeltro in Urbino, trying to figure out the significance and the identity or the allegorical meaning of the apparently preoccupied group of three men at the right of the scene. Hartt discusses this, although it is mostly speculation. There are many other enigmas in the Italian renaissance paintings. For example, Hartt discusses the famous Tempestuous Landscape With the Soldier and the Gypsy painted around 1505 to 1510 by Giorgione (p. 633), a painting filled with mystery. An unconventional 1510 painting with a similarly unclear meaning, Fête Champêtre (p. 635), by either Giorgione or Titian, had a tre-
mendous influence on Edouard Manet when he created his notorious Déjeuner sur l’herbe. Or look at The Baptism of Christ, painted by Verrocchio (p. 367) but with parts done by the young Leonardo da Vinci, his pupil. When it is first presented under the topic of Verrocchio in the book (pp. 366–371), there is only a passing remark, and I thought the mystery of it would be ignored, but later in the book Hartt explains Leonardo’s contribution to the picture:

About most of the painting there can be little doubt, for Verrocchio’s hand is everywhere apparent. But of the two kneeling angels, the curly headed boy at the right…is in sharp contrast to his companion at the left, who looks out from deep, luminous eyes and whose hair streams from his forehead to his shoulders with the mysterious patterns of Leonardo’s water patterns. The water behind him, whose shimmering surface breaks into rapids over underlying shoals and whose juncture with the surrounding rocks is masked by mists, is by the artist who was to later paint the watercourses in the landscape of the Mona Lisa. (p. 484)

Leonardo, says Hartt, was the first to have produced a painted female portrait in which the sitter turns toward the viewer and the first to include her hands. Yet in thousands of pages of his writing, Leonardo never gave any “hint that he ever cared deeply for another human being” (p. 478). Whether the fact that he was an illegitimate child contributed to this is unknown. There is a beautiful long paragraph describing the contrast between the characters and styles of Michelangelo and Leonardo (p. 501).

It is interesting that Luca Signorelli and Piero di Cosimo in their Orvieto frescoes depict what Hartt says was most repugnant to Michelangelo, “humanity in a subhuman stage and subject to the depredations of antihuman creatures and forces” (p. 523). The artists of the High Renaissance, originating in the work of Leonardo (p. 477), attempted to raise humanity to a divine level. Chapter 16 is the highlight of the book, covering in great detail Leonardo da Vinci, Michelangelo, and Raphael, the foremost painters of the High Renaissance in the early Cinquecento.

St. Anthony of Padua predicted that the heart of a famous miser would be found after his death not in his body but in his money chest. Tullio Lombardo, influenced by the Venetians’ belief they were the true heirs of ancient Imperial Rome, made a remarkable relief sculpture of this in about 1525 for the tomb chapel of St. Anthony in Padua (p. 653). Hartt concludes that, “by the middle of the Cinquecento in central Italy the Renaissance, in its etymological sense of ‘rebirth,’ was over” (p. 702). This wonderful book, which must have been a labor of enormous extent to produce and revise through five editions, closes with a very helpful glossary, bibliography, and index.

References

RICHARD D. CHESSICK, M.D., PH.D.
Evanston, Ill.

Reprints are not available; however, Book Forum reviews can be downloaded at http://ajp.psychiatryonline.org.