In This Issue

Perspectives

Editorials

Jeffrey L. Cummings
Searching for Methods to Detect, Prevent, and Treat Alzheimer’s Disease

Reviews and Overviews

Glen O. Gabbard
Mind, Brain, and Personality Disorders
The Efficacy of Light Therapy in the Treatment of Mood Disorders: A Review and Meta-Analysis of the Evidence
Stephan Heckers and Debra Titone

**Hippocampus, IV: Relational Memory**

---

**Introspections:**

Lloyd Benjamin

**Gaze**

---

**Images in Psychiatry:**

Ana Maria G. Raimundo Oda, Walmor Piccinini, and Paulo Dalgalarrondo

**Juliano Moreira (1873–1933): Founder of Scientific Psychiatry in Brazil**

---

**New Research**

---

**Articles:**

Pedro J. Modrego, Nicolás Fayed, and Miguel A. Pina

**Conversion From Mild Cognitive Impairment to Probable Alzheimer’s Disease Predicted by Brain Magnetic Resonance Spectroscopy**

Mamoru Hashimoto, Hiroaki Kazui, Keiji Matsumoto, Yoko Nakano, Minoru Yasuda, and Etsuro Mori

**Does Donepezil Treatment Slow the Progression of Hippocampal Atrophy in Patients With Alzheimer’s Disease?**

Kristine Yaffe, Kathryn Krueger, Steven R. Cummings, Terri Blackwell, Victor W. Henderson, Somnath Sarkar, Kristine Ensrud, and Deborah Grady

**Effect of Raloxifene on Prevention of Dementia and Cognitive Impairment in Older Women: The Multiple Outcomes of Raloxifene Evaluation (MORE) Randomized Trial**

Michael A. Rapp, Karen Dahlman, Mary Sano, Hillel T. Grossman, Vahram Haroutunian, and Jack M. Gorman

**Neuropsychological Differences Between Late-Onset and Recurrent Geriatric Major Depression**

Anthony F. Jorm, Kaarin J. Anstey, Helen Christensen, Greg de Plater, Rajeev Kumar, Wei Wen, and Perminder Sachdev

**MRI Hyperintensities and Depressive Symptoms in a Community Sample of Individuals 60–64 Years Old**

Dan G. Blazer, Celia F. Hybels, Gerda G. Fillenbaum, and Carl F. Pieper

**Predictors of Antidepressant Use Among Older Adults: Have They Changed Over Time?**
Wenhui Wei, Usha Sambamoorthi, Mark Olfson, James T. Walkup, and Stephen Crystal

*Use of Psychotherapy for Depression in Older Adults*

George S. Alexopoulos, Ira R. Katz, Martha L. Bruce, Moonseong Heo, Thomas Ten Have, Patrick Raue, Hillary R. Bogner, Herbert C. Schulberg, Benoît H. Mulsant, Charles F. Reynolds, III, and The PROSPECT Group

*Remission in Depressed Geriatric Primary Care Patients: A Report From the PROSPECT Study*

Jane S. Paulsen, Karin Ferneyhough Hoth, Carissa Nehl, Laura Stierman, and The Huntington Study Group

*Critical Periods of Suicide Risk in Huntington’s Disease*


*Symptom Fluctuation in Eating Disorders: Correlates of Diagnostic Crossover*

Virginia V.W. McIntosh, Jennifer Jordan, Frances A. Carter, Suzanne E. Luty, Janice M. McKenzie, Cynthia M. Bulik, Christopher M.A. Frampton, and Peter R. Joyce

*Three Psychotherapies for Anorexia Nervosa: A Randomized, Controlled Trial*

Hans W. Hoek, Peter N. van Harten, Karin M.E. Hermans, Melanie A. Katzman, Glenn E. Matroos, and Ezra S. Susser

*The Incidence of Anorexia Nervosa on Curaçao*

Harry R. Millar, Fiona Wardell, Juliet P. Vyvyan, Simon A. Naji, Gordon J. Prescott, and John M. Eagles

*Anorexia Nervosa Mortality in Northeast Scotland, 1965–1999*

Erland W. Schubert and Thomas F. McNeil

*Neuropsychological Impairment and Its Neurological Correlates in Adult Offspring With Heightened Risk for Schizophrenia and Affective Psychosis*

Alan S. Brown, Catherine A. Schaefer, Charles P. Quesenberry, Jr., Liyan Liu, Vicki P. Babulas, and Ezra S. Susser

*Maternal Exposure to Toxoplasmosis and Risk of Schizophrenia in Adult Offspring*

Traolach Brugha, Nicola Singleton, Howard Meltzer, Paul Bebbington, Michael Farrell, Rachel Jenkins, Jeremy Coid, Tom Fryers, David Melzer, and Glyn Lewis

*Psychosis in the Community and in Prisons: A Report From the British National Survey of Psychiatric Morbidity*
Cathaleene Macias, Paul Barreira, William Hargreaves, Leonard Bickman, William Fisher, and Elliot Aronson

**Impact of Referral Source and Study Applicants’ Preference for Randomly Assigned Service on Research Enrollment, Service Engagement, and Evaluative Outcomes**

---

**Brief Reports:**

Paul J. Moberg, David R. Roalf, Catherine C. Balderston, Stephen J. Kanes, Raquel E. Gur, and Bruce I. Turetsky

**Phenylthiocarbamide Perception in Patients With Schizophrenia and First-Degree Family Members**

Gabriele Ende, Petra Hubrich, Sigrid Walter, Wolfgang Weber-Fahr, Nina Kämmerer, Dieter F. Braus, and Fritz A. Henn

**Further Evidence for Altered Cerebellar Neuronal Integrity in Schizophrenia**

Christopher R. Bowie, Irene Tsapelas, Joseph Friedman, Michael Parrella, Leonard White, and Philip D. Harvey

**The Longitudinal Course of Thought Disorder in Geriatric Patients With Chronic Schizophrenia**

Gordon Parker, Heather Brotchie, and Kay Parker

**Is Combination Olanzapine and Antidepressant Medication Associated With a More Rapid Response Trajectory Than Antidepressant Alone?**

Jitschak G. Storosum, Tamar Wohlfarth, Christine C. Gispen-de Wied, Don H. Linszen, Berthold P.R. Gersons, Barbara J. van Zwieten, and Wim van den Brink

**Suicide Risk in Placebo-Controlled Trials of Treatment for Acute Manic Episode and Prevention of Manic-Depressive Episode**

David A. Axelson, James M. Perel, Boris Birmaher, George Rudolph, Sharon Nuss, Linda Yurasits, Jeffrey Bridge, and David A. Brent

**Platelet Serotonin Reuptake Inhibition and Response to SSRIs in Depressed Adolescents**

Alexander Neumeister, Peixiong Yuan, Theresa A. Young, Omer Bonne, David A. Luckenbaugh, Dennis S. Charney, and Husseini Manji

**Effects of Tryptophan Depletion on Serum Levels of Brain-Derived Neurotrophic Factor in Unmedicated Patients With Remitted Depression and Healthy Subjects**

Maria J. Portella, Catherine J. Harmer, Jonathan Flint, Philip Cowen, and Guy M. Goodwin

**Enhanced Early Morning Salivary Cortisol in Neuroticism**
Letters to the Editor:

DAVID I. MAYERHOFF, JEFFRY NURENBERG, SNEHAL SHAH, and STEVEN J. SCHLEIFER

**Neurotoxicity Associated With Free Valproic Acid**

SHINSUKE KITO, TORU NAKAJIMA, HIROSHI YAMADERA, YOSHIHIKO KOGA, SHINJI KOSUGI, and NORITAKA HAI

**Multiple Endocrine Neoplasia Type 1 Presenting as Psychosis**

MASANORI SAITO, YOSHIKI MATSUI, YOSHIMASA OTANI, and HITOSHI MIYAOKA

**Liepmann’s Phenomenon During Benzodiazepine Withdrawal**

SARA DORIS BIENENTREU and KLAUS-THOMAS HELMUT KRONMÜLLER

**Increase in Risperidone Plasma Level With Lamotrigine**

RYOHEI MATSUMOTO, YURINOSUKE KITABAYASHI, YASUHITO NAKATOMI, HIDETO TSUCHIDA, and KENJI FUKUI

**Neuroleptic Malignant Syndrome Induced by Quetiapine and Fluvoxamine**

MARCUS C. ROSENHAGEN, MANFRED UHR, PETRA SCHÜSSLER, and AXEL STEIGER

**Elevated Plasma Ghrelin Levels in Night-Eating Syndrome**

FLORENCE VORSPAN, DOMINIQUE WAROT, ANGÈLE CONSOLI, DAVID COHEN, and PHILIPPE MAZET

**Mania in a Boy Treated With Modafinil for Narcolepsy**

KARIM SEDKY, RITA SHAUGHNESSY, TIFFANY HUGHES, and STEVEN LIPPMANN

**Clozapine-Induced Agranulocytosis After 11 Years of Treatment**

ABHAY SINGH, ROBERT ALTHOFF, R. JARED MARTINEAU, and JAMES JACOBSON

**Pramipexole, Ropinirole, and Mania in Parkinson’s Disease**

AKIFUMI IKEDA, KANAKO SEKIGUCHI, KENICHI FUJITA, HIROSHI YAMADERA, and YOSHIHIKO KOGA

**5-Methoxy-N,N-Diisopropyltryptamine-Induced Flashbacks**

DAVID E. ROSS, MICHELE THOMAS, MARIA BOOTH, and MICHAEL WEINBORN

**Rate of Tardive Dyskinesia in Hospitalized Patients**
FUMIO MORIYA and YOSHIKAI HASHIMOTO

Do Smokers Who Commit Suicide Have High Blood Levels of Nicotine?

MICHAEL SERBY and STEVEN C. SAMUELS

Dementia With Lewy Bodies, Visual Hallucinations, and Medications

ANDRÉS MARTIN, WALTER S. GILLIAM, JEFFREY Q. BOSTIC, and JOSEPH M. REY

Child Psychopharmacology, Effect Sizes, and the Big Bang

RÉMY P. BARBE

Child Psychopharmacology, Effect Sizes, and the Big Bang

MAJU MATHEWS, BABATUNDE ADETUNJI, JOANNE MATHEWS, BIJU BASIL, VINU GEORGE, MANU MATHEWS, KUMAR BUDUR, and SHINY ABRAHAM

Child Psychopharmacology, Effect Sizes, and the Big Bang

KAREN DINEEN WAGNER, ADELAIDE S. ROBB, ROBERT L. FINDLING, and JIANQING JIN

Dr. Wagner and Colleagues Reply

Book Forum:

RICHARD H. GIBSON and LAURA WEISS ROBERTS

Morality and Ethics in Theory and Practice

C. ROBERT CLONINGER

Character Strengths and Virtues: A Handbook and Classification

BEATA ZOLOVSKA and HAROLD J. BURSZTAJN

"Are You There Alone?" The Unspeakable Crime of Andrea Yates

OWEN D. BUCK

PsychoBible

REMI J. CADORET

European Cases of the Reincarnation Type

RICHARD D. CHESSICK

Understanding Dissidence and Controversy in the History of Psychoanalysis

PETER BARGLOW

Misunderstanding Freud

MARCIA WEBB

Interpersonal Foundations of Psychopathology
SOPHIA VINOGRAVOD
Catatonia: A Clinician’s Guide to Diagnosis and Treatment

MARK H. FLEISHER
Neurodevelopmental Mechanisms in Psychopathology

HAROLD ALAN PINCUS
Mental Health Services: A Public Health Perspective, 2nd ed.

ISRAEL KATZ and ALBERT C. GAW
School-Based Mental Health Services: Creating Comprehensive and Culturally Specific Programs

DONALD M. HILTY and MALATHI SRINIVASAN
Evidence-Based Practice Manual: Research and Outcome Measures in Health and Human Services

Corrections:

Correction
Correction
Correction
Correction
Predicting Alzheimer’s Disease

Some elderly people have memory problems beyond those expected for their age. Is this mild cognitive impairment a harbinger of Alzheimer’s disease? The answer could lead to early drug therapy of mild Alzheimer’s disease and slow its progression. Modrego et al. (p. 667) measured several chemical compounds in the brains of patients with mild cognitive impairment, using magnetic resonance spectroscopy. After 3 years, 29 of the 53 patients had developed probable Alzheimer’s disease. The only baseline value that predicted this conversion was the ratio of N-acetylaspartate to creatine in the left occipital lobe. The demented patients’ average ratio was 1.46, whereas for nondemented patients it was 1.64. If confirmed in larger studies, this finding would provide an objective tool for predicting conversion of mild cognitive impairment to probable Alzheimer’s disease.

Wider Antidepressant Use by the Elderly

With the development of newer medications, use of antidepressants began to increase in the late 1980s. That coincided with the start of a long-term study of 4,162 elderly community residents in North Carolina. Blazer et al. (p. 705) report characteristics in 1986/1987 that predicted use of antidepressants 3 years later, in 1989/1990, and features in 1992/1993 that were associated with antidepressant use in 1996/1997. The relationships for the two intervals showed some similarities and some differences. Negative affect and previous antidepressant use predicted current use in both intervals. Somatic problems were a stronger predictor in the first interval than in the second, whereas chronic illness, cognitive dysfunction, and low positive affect were stronger factors in the second interval. Over the study period, antidepressant use rose from 4.5% of respondents to 8.1%, but the rates were lower for blacks than for whites, and chronic illness, cognitive dysfunction, and low positive affect were stronger factors in the second interval. Over the study period, antidepressant use rose from 4.5% of respondents to 8.1%, but the rates were lower for blacks than for whites, and this discrepancy grew over time.

Crossover of Eating Disorders

Many patients with anorexia nervosa or bulimia nervosa eventually develop the other eating disorder. To help identify who is likely to switch disorders, Tozzi et al. (p. 732) assessed 88 girls or women with anorexia nervosa, restricting subtype, and 350 with bulimia nervosa, purging subtype. Of those with anorexia nervosa, 36% developed bulimia nervosa over 15 years. Of those with bulimia nervosa, 27% developed anorexia nervosa. The crossover usually occurred within the first 5 years. For subjects who switched from anorexia nervosa to bulimia, perceived parental criticism was a salient predictor. Crossover from bulimia to anorexia nervosa was linked to less novelty seeking and less alcohol abuse/dependence. Both types were associated with a low level of self-directedness and family communication, therefore, may stabilize the diagnosis as well as address core psychological issues.

Affective Psychosis: Another Neurodevelopmental Disorder?

Schizophrenia has been dubbed a neurodevelopmental disorder, and many patients and their relatives have brain, neurological, and neuropsychological abnormalities. Schizophrenia shows some genetic connection to affective psychosis, and patients with affective psychosis also have neurological abnormalities and neuropsychological impairment. To determine whether these deficits are associated with risk for the disorder, Schubert and McNeil (p. 758) examined young adult offspring of mothers with schizophrenia, mothers with affective psychosis, and healthy mothers. There was only one neuropsychological test on which both groups of high-risk offspring differed from the normal-risk offspring. Also, the affective-risk group was less likely than the schizophrenia-risk group to score above the normal subjects’ 90th percentile on multiple tests, and their level of neurological abnormalities was similar to that of the normal-risk offspring. These findings suggest that schizophrenia and affective psychosis belong to different biological categories.

Psychosis Goes to Prison

Are the high rates of psychosis found among prisoners real, or do they reflect prisoner drug abuse or unrepresentative surveys? They are real, according to Brugha et al. (p. 774). National samples totaling 13,250 adults in British households and prisons were screened for psychosis, and a diagnostic interview was administered to those with positive results. The prevalence of psychosis was more than 10 times higher among the prisoners than among the community subjects, 52 per 1,000 versus 4.5 per 1,000. The difference could not be explained by the prisoners’ lower socioeconomic status or age and was minimally related to alcohol and drugs. Psychotic symptoms were similar in the two groups, and the rates of antipsychotic medication in the community and in male prisoners were the same (34%). Fewer prisoners had seen primary care physicians about mental health problems during the previous year, however, suggesting a need for greater recognition and treatment for these prisoners.
Searching for Methods to Detect, Prevent, and Treat Alzheimer’s Disease

Alzheimer’s disease becomes more prevalent with increasing age, and the world’s population is aging at a rapid rate. Unless means are found to prevent or retard the emergence of cognitive impairment in aged individuals, the personal and public health burdens posed by Alzheimer’s disease and other cognitive impairment syndromes of the elderly will grow dramatically. This issue of The American Journal of Psychiatry includes three contributions that address important facets of this evolving health challenge. Modrego and colleagues provide evidence that magnetic resonance spectroscopy may assist in determining which patients with mild cognitive impairment will go on to develop Alzheimer’s disease. Yaffe and co-workers demonstrate that the selective estrogen receptor modulator raloxifene may reduce the development of mild cognitive impairment and other cognitive deficit syndromes in postmenopausal women. Hashimoto et al. examine the important question of whether treatment of Alzheimer’s disease with cholinesterase inhibitors reduces disease progression. Taken together, these contributions provide new evidence that technology may assist in identifying patients with Alzheimer’s disease before the emergence of Alzheimer’s dementia and that medications may help maintain cognition, prevent or delay the emergence of cognitive impairment, and have disease-modifying as well as symptomatic effects.

Mild cognitive impairment is a syndrome characterized by relatively marked impairment in a single cognitive domain such as memory or moderate impairment in several cognitive domains, but patients continue performing activities of daily living normally and do not meet criteria for dementia (1, 2). Patients with mild cognitive impairment commonly progress to Alzheimer’s disease, converting from one diagnosis to the other at a rate of approximately 15% per year on average. Some patients, however, appear to remain in the mild cognitive impairment state for long periods, and occasionally patients recover from mild cognitive impairment to normal cognition (3). Foreknowledge of which patients with mild cognitive impairment suffer from Alzheimer’s disease and will progress to Alzheimer’s dementia would allow the appropriate application of disease-modifying treatments to prevent further progression at a point when clinical manifestations are limited. Modrego et al. studied the sensitivity and specificity of magnetic resonance spectroscopy in distinguishing patients with mild cognitive impairment that converted to Alzheimer’s disease compared with patients who did not progress to Alzheimer’s disease over a 3-year follow-up period. An occipital cortex N-acetylaspartate/creatine ratio ≤1.61 predicted conversion from mild cognitive impairment to dementia with 100% sensitivity, 75% specificity, a positive predictive value of 83%, and a negative predictive value of 100%. It is surprising that the N-acetylaspartate/creatine ratio did not have significant predictive value in voxels derived from the hippocampus or parietal cortex, areas more commonly associated with classic and severe changes of Alzheimer’s disease. N-Acetylaspartate is considered a surrogate marker of neuronal integrity, and diminished levels of this compound as detected by proton spectroscopy suggest a local decrement in neuronal number. Modrego and colleagues divide their patient...

“These contributions provide new evidence that technology may assist in identifying patients with Alzheimer’s disease...and that medications may...have disease-modifying as well as symptomatic effects.”
group into converters and nonconverters on the basis of their 3-year observations. A more cautious terminology would be “early converters” versus “early nonconverters,” since a substantial number of individuals whose cognitive impairment did not convert within 3 years may go on to develop Alzheimer’s disease.

Yaffe and coworkers examined the data from the Multiple Outcomes of Raloxifene Evaluation trial, which included 5,386 women examined for the presence of dementia 3 years after being randomly assigned to receive placebo, 60 mg of raloxifene, or 120 mg of raloxifene. Those receiving the higher raloxifene dose had a 33% lower risk of developing mild cognitive impairment relative to those receiving 60 mg or placebo. There was also a tendency, although not significant, for those receiving the higher dose to have a lower risk of developing Alzheimer’s disease or any cognitive impairment syndrome. Raloxifene is a selective estrogen receptor modulator used to prevent and treat osteoporosis. Women with emergent cognitive impairment were generally older and postmenopausal for longer periods of time, were more likely to have had a hysterectomy, were less well educated, drank less alcohol, and had more depressive symptoms at baseline relative to women without cognitive deficits. These observations reinforce an emerging epidemiology indicating that late-onset depression is often a precursor of dementia and that moderate alcohol consumption may reduce the risk of dementia.

There is an apparent contradiction between the results of the current study and those of the Women’s Health Initiative Memory Study (4), which showed that estrogen plus progestin increased the incidence of dementia in postmenopausal women. However, the current study used a selective estrogen receptor modulator and that may be the critical difference between the two studies. Differences in selection criteria and trial methodology may also contribute to differential outcomes. Potential mechanisms by which raloxifene might defer the onset of cognitive compromise include stimulation of choline acetyl transferase activity in the hippocampus, stimulation of neurite outgrowth, and increasing the numbers of 5-hydroxytryptamine-2A receptors in the cingulate and frontal cortex (5–7). Furthermore, raloxifene has not been associated with adverse cardiovascular or cerebrovascular outcomes that may contribute to cognitive impairment (8). The rate of emergence of mild cognitive impairment or Alzheimer’s disease was low in this trial, reflecting the general well-being of the clinical trial participants. The extent to which these apparently beneficial effects of raloxifene can be generalized to other populations awaits further study.

Hashimoto and co-workers investigated the potential disease-modifying effects of donepezil, a commonly used cholinesterase inhibitor. Comparing the rate of brain atrophy as revealed by magnetic resonance imaging measures of the hippocampus taken at 1-year intervals, the investigators compared patients treated with donepezil to historical control subjects assessed identically prior to the availability of donepezil therapy. There was a significant difference between the two groups: the mean annual rate of hippocampal atrophy in the treated group (3.82%) was significantly lower than that in the historical control group (mean=5.04%). The 5% annual decrease in hippocampal volume found by Hashimoto et al. is close to the 4.9% annual change reported by Jack and colleagues (9) in a recent placebo-controlled trial. The authors cite in vitro evidence that muscarinic receptor stimulation decreases beta-amyloid protein production. Nicotinic cholinergic receptor stimulation also appears to protect neurons from degeneration induced by amyloid beta protein (10, 11). It is also possible that the enhanced cognition observed with donepezil therapy may be reflected in relative maintenance of synaptic integrity, which in turn may influence measurable atrophy. The use of historical control subjects in this study requires that the results be regarded as tentative. Historical control subjects appear to have progressed more rapidly than those currently recruited into clinical trials (12, 13). These cohort differences may contribute to the differences reported by Hashimoto et al. A small double-blind, placebo-controlled trial with parallel groups also found a beneficial effect of donepezil on maintenance of hip-
These findings support the reported results. Larger randomized trials with parallel groups will be necessary before final conclusions can be drawn regarding the potential neuroprotective effects of donepezil or other cholinesterase inhibitors.

These three studies provide important new information indicating that advances in technology may assist in identifying patients at risk for Alzheimer’s disease. In addition, therapies such as raloxifene may prevent or defer the onset of mild cognitive impairment, thus reducing the number of patients who may progress from mild cognitive impairment to Alzheimer’s disease. Cholinesterase inhibitors represent the standard of care for patients with mild to moderate Alzheimer’s disease and may have disease-modifying as well as symptomatic effects. These three individual studies represent small but important steps in our march toward more effective prevention and treatment of Alzheimer’s disease.

References


JEFFREY L. CUMMINGS, M.D.

Address correspondence and reprint requests to Dr. Cummings, Reed Neurological Research Center, David Geffen School of Medicine at UCLA, 710 Westminster Plaza, Suite 2–238, Box 951769, Los Angeles, CA 90095–1769; jcummings@mednet.ucla.edu (e-mail). Dr. Cummings is supported by the National Institute on Aging (PS0 AG-16570), the Alzheimer’s Disease Research Centers of California, and the Sidell-Kagan Foundation. Dr. Cummings has served as a consultant to Pfizer Inc. and Eisai Pharmaceuticals.
Mind, Brain, and Personality Disorders

Glen O. Gabbard, M.D.

Objective: The use of the terms “mind” and “brain” in psychiatry is often associated with a set of polarities. Concepts such as environment, psychosocial, and psychotherapy are linked with “mind,” while genes, biology, and medication are often associated with “brain.” The author examines these dichotomies as they apply to personality disorders.

Method: Research on antisocial and borderline personality disorders that is relevant to these dichotomies is evaluated. The implications of the findings for the understanding of pathogenesis and treatment are reconsidered.

Results: In the clinical setting, it is problematic to lump together terms such as “genes,” “brain,” and “biological” as though they are separate and distinct from terms such as “environment,” “mind,” and “psychosocial.” These dichotomies are problematic, because genes and environment are inextricably intertwined in the pathogenesis of personality disorders, psychosocial experiences may result in permanent changes in the brain, and psychotherapy may have its effect by altering brain structure and function. The “theory of mind” is a useful construct for bridging “mind” and “brain” in the treatment of personality disorders.

Conclusions: Severe personality disorders are best understood and treated without “either-or” dichotomies of brain and mind. Each domain has a different language, however, and the language of the mind is necessary to help the patient develop a theory of mind.

The mind-brain relationship has vexed philosophers for centuries and continues to be the subject of controversy. In psychiatric discourse, we often refer to “mind” and “brain” as though they are separate entities, even though most psychiatrists in the post-Cartesian era regard the mind as the activity of the brain (1). The persistence of these terms in contemporary psychiatric discussions reflects the fact that references to “mind” and “brain” have become a form of code for different ways to think about patients and their treatment. As Cloninger (2) noted, “biomedical” and “psychosocial” define two discrete paradigms, and the division into these separate models has had a stagnating effect on the science of mental health. Polarities such as genes versus environment, medication versus psychotherapy, and biological versus psychosocial are often glibly subsumed under categories of “brain” and “mind” (Figure 1).

Using what we know about personality disorders, we can begin to deconstruct some of these problematic dichotomies while still preserving the broadly biopsychosocial framework of diagnosis and treatment that is essential to the provision of comprehensive and effective intervention for patients with these disorders. Virtually all major psychiatric disorders are complex amalgams of genetic diatheses and environmental influences. Genes and environment are inextricably connected in shaping human behavior. Experience shuts down the transcriptional function of some genes, while turning on that of others (3). As Michael Rutter has emphasized, “Genetic influences, as they apply to individual differences in the liability to show particular behaviors, are strong and pervasive but rarely determinative” (4, p. 996). Similarly, psychosocial stressors, such as interpersonal trauma, have profound effects of a biological nature by changing the functioning of the brain. Teasing apart biological and psychosocial phenomena may be a formidable challenge when treating a patient. Finally, to think of psychotherapy as a treatment for “psychologically based disorders” and medications as a treatment for “biological or brain-based disorders” is to make a specious distinction. The effect of psychotherapy on the brain is well established (5).

A review of recent research on personality disorders suggests that these constructs can be dichotomized only in the abstract. In clinical work with patients, mind and brain are intimately connected and can never be separated. Nevertheless, each domain has its own language (6). In clinical work with personality disorders, a “bilingual” psychiatrist who understands both the language of “mind” and the language of “brain” may be in a better position to develop a comprehensive biopsychosocial formulation and implement an overall treatment plan. The “theory of mind” (7, 8) is a particularly useful bridge between the two constructs in the treatment of personality disorders, because it emphasizes the role that psychotherapy plays in creating a sense of mind in the patient.
Genes and Environment

Because of space considerations, this partial review will be limited to antisocial and borderline personality disorders, the two axis II entities with the greatest bodies of research data. Although systematic genetic studies of borderline personality disorder are in short supply, well-designed studies of antisocial personality disorder and criminality consistently show that a genetic diathesis acted on by adverse environmental experience is necessary for the development of the disorder (9–12). In an elegant study from New Zealand (12), for example, a birth cohort of 1,037 children were followed prospectively and assessed at ages 3, 5, 7, 9, 11, 13, 15, 18, and 21 years. Ninety-six percent of the cohort was intact at age 26. Between the ages of 3 and 11 years, 8% experienced “severe” maltreatment, 28% experienced “probable” maltreatment, and 64% experienced no maltreatment. A functional polymorphism in the gene occurring in the neurotransmitter-metabolizing enzyme monoamine oxidase A (MAO A) was found to moderate the effect of maltreatment. Males with the low MAO A activity genotype who were maltreated in childhood had elevated scores on measures of antisocial behavior. Males with high MAO A activity did not have elevated antisocial behavior scores, even when they had experienced childhood maltreatment. Eighty-five percent of males with both the low MAOA activity genotype and severe maltreatment developed antisocial behavior. These findings suggest that genotypes moderate children’s sensitivity to environmental stressors.

Primate studies have yielded a heuristically valuable animal model that resembles human patients with a diagnosis of borderline personality disorder and antisocial personality disorder. Monkey research has also made it possible to vary specific rearing effects to gain more specific knowledge about the influence of environmental factors. Between 5% and 10% of field populations of rhesus monkeys are unusually impulsive, insensitive, and overtly aggressive in their interactions with other troop members (13). They make dangerous leaps in trees that result in self-injury. They harass juveniles who are younger and physically weaker. They are also socially inappropriate and may self-destructively challenge a dominant adult male. Males with these characteristics are likely to be expelled from the group before puberty, while females are likely to end up at the bottom of the social hierarchy and are incompetent, neglectful mothers.

Rhesus monkeys, who share approximately 95% of their genes with human beings, also show commonalities in the linkage between impulsive aggression and measures of serotonergic metabolism (14). An inverse relationship exists between measures of CSF 5-hydroxyindoleacetic acid (5-HIAA) concentration and measures of impulsive aggression. However, the inherited propensity to develop patterns of impulsive aggressiveness can be modified substantially by early experiences involving social attachment relationships. Monkeys reared by peers consistently demonstrate lower CSF concentrations of 5-HIAA, compared to those reared by mothers.

The serotonin transporter (5-HTT) gene has length variation in its promoter region that results in allelic variation in 5-HTT gene expression. A “short” allele (ls) confers low transcriptional efficiency to the 5-HTT promoter, relative to the “long” allele (ll), suggesting that low 5-HTT gene expression may result in decreased serotonergic function.

Bennett et al. (15) found that CSF 5-HIAA concentrations did not differ as a function of 5-HTT status for mother-reared subjects, whereas among peer-reared monkeys, individuals with the ls allele had significantly lower CSF 5-HIAA concentrations than those with the ll allele. Being reared by one’s mother appeared to buffer any potential deleterious effects of the ls allele on serotonin metabolism. Conversely, peer-reared monkeys with the ls polymorphism exhibited much higher levels of impulsive aggression than their peer-reared counterparts with the ll polymorphism, who exhibited low levels of impulsive aggression similar to those of both ll and ls mother-reared monkeys, again suggesting a buffering effect of maternal rearing.

Rhesus monkeys with low CSF concentrations of 5-HIAA are also prone to consume more alcohol in a “happy hour” situation where a 7% ethanol aspartame-flavored beverage is available (13). Here, the data on maternal buffering effects strikingly reflect the role of environment on the influence of genes. Peer-reared monkeys with the ls allele consumed more alcohol than peer-reared monkeys with the ll allele. Exactly the reverse was true if the subjects were reared by mothers: the monkeys with the ls allele had less alcohol consumption than the ll monkeys. Investigators concluded that the short allele of the 5-HTT gene may well lead to psychopathology among rhesus monkeys who have adverse early rearing histories but could possibly be adaptive for those monkeys who have secure early attachment relationships with their mothers (13). Both nature and nurture appear to be at play in the development of most, if not all, of the biobehavioral aspects of rhesus monkey impulsive aggressiveness.

The implications for psychotherapy from this sophisticated understanding of gene-environment interaction are
provocative. With the knowledge that genetic “hard-wiring” is a questionable assumption, we have reason to be optimistic about potential consequences of altering early parental and caregiver interactions with children.

Although randomized, controlled trials of therapy with individuals or families at risk for antisocial personality disorder have yet to be reported, a long-term follow-up study of the effect of home visitation by a public health nurse on children’s antisocial behavior (16) is highly suggestive. The investigators randomly assigned a visiting home nurse to high-risk new mothers. The visits started during pregnancy and continued through the child’s second birthday. The comparison group consisted of mother-child pairs who received standard prenatal and well-child care in the public health clinic. Eighty-five percent of the mothers enrolled were young, unmarried, or from households with low socioeconomic status. The nurses visited an average of nine times during pregnancy and 23 times from birth through the child’s second birthday. Three aspects of maternal functioning were the focus of the home visits: health-related behaviors, competent care of children, and maternal personal development. At 15-year follow-up, adolescents born to women who had received the nurse visits had significantly lower rates of antisocial behavior, relative to the comparison subjects. They also had lower rates of substance abuse and fewer lifetime sex partners.

Results of this nature raise the possibility that early psychotherapeutic interventions might serve to influence the expression of genes that lead to antisocial behavior. A neglected benefit of individual psychotherapy is its positive effect on the offspring of the patient. After extensively studying genetic and environmental influences on adolescent antisocial behavior, Reiss et al. (17) made the following observation: “The encoding of genetic information into family processes might rival in importance, and serve in tandem with, the much better known process of RNA encoding—the critical intracellular transduction of genetic information on the road to protein synthesis” (p. 386). In any case, genes and environment are inextricably connected in the pathogenesis of antisocial behavior; a strict “either/or” dichotomy is specious.

The “Biological/Psychosocial” Distinction, Trauma, and Borderline Personality Disorder

The etiology of borderline personality disorder is probably multifactorial, but extensive research supports the notion that early abuse and neglect may be a significant factor in many cases (18). Early childhood separations, chaotic home environments, insensitivity to the child’s feelings and needs, emotional discord in the family, and trauma of varying degrees have all been implicated in the etiology (see reference 18 for a review of this literature). A genetically based temperament is also probably a key factor (19, 20). Certain temperamental predispositions may increase the likelihood that negative life events will occur. Borderline personality disorder research documenting early abuse has been controversial because much of it has relied on retrospective methods that must confront the vagaries of memory. However, an elegantly designed prospective study of 639 youths and their mothers demonstrated strong linkage between trauma and borderline personality disorder (21). These investigators found that both neglect and sexual abuse were associated with greater symptoms of borderline personality disorder.

The effect of early trauma on the developing brain has been a subject of increasing interest in psychiatry. The hippocampus appears to be vulnerable to the effects of stress, in part because it has many glucocorticoid receptors (22). Some imaging studies have demonstrated reduced hippocampal volume in adult patients with borderline personality disorder (23, 24).

Preliminary data suggest that early trauma may promote hemispheric lateralization and adversely affect integration of the right and left hemispheres. Auditory-probe- evoked potential attenuation was measured as an index of hemispheric activity in 10 subjects with a history of childhood trauma and 10 matched comparison subjects without such history while they recalled a neutral memory and then a traumatic memory (25). Abused children used their left hemisphere when thinking about neutral memories, and their right hemisphere for frightening memories. The comparison group used both left and right sides equally regardless of the memory content.

This failure of hemispheric integration may be reflected in the use of splitting as a major defense mechanism by borderline personality disorder patients. To deal with the concern that hate and aggression will destroy all positive qualities, they tend to compartmentalize self and object representations into “all good” and “all bad” categories (26).

A growing body of research suggests that another consequence of early childhood trauma is persistent sensitization of the hypothalamo-pituitary-adrenal (HPA) axis (27–29). Women with a history of childhood abuse and major depression have shown a more than sixfold greater ACTH response to stress than age-matched comparison subjects (30). The researchers concluded that a persistent consequence of childhood abuse is hyperreactivity of the HPA axis and autonomic nervous system. They inferred that this hyperreactivity was related to hypersecretion of corticotropin-releasing factor (CRF).

These findings have now been confirmed in studies of patients with borderline personality disorder who had sustained childhood abuse. Rinne et al. (31) studied 39 female borderline personality disorder patients who were given combined dexamethasone/corticotropin-releasing hormone (CRH) tests and compared with 11 healthy subjects. Twenty-four of these women had histories of sustained childhood abuse. The chronically abused borderline personality disorder patients had a significantly enhanced ACTH and cortisol response to the dexametha-
sone/CRH challenge, compared with nonabused subjects. The researchers concluded that a history of sustained childhood abuse is associated with hyperresponsiveness of ACTH release. Their findings suggest that this hyperactive physiological state is relevant to a subgroup of borderline personality disorder patients, but not all. Sustained childhood abuse appears to increase the CRH receptors’ sensitivity.

Several implications can be derived from these findings. First, these data illustrate why it is problematic to lump together terms such as “genes,” “brain,” and “biological,” as though they are separate and distinct from terms such as “environment,” “mind,” and “psychosocial.” Psychosocial events may result in persisting biological alterations in the brain. Second, because the HPA axis is intimately linked with serotonergic function, these data suggest the possibility of understanding the mechanism of action of serotonin reuptake inhibitors in patients with borderline personality disorder. Third, because internal object relationships are created from the building blocks of self representations, object representations, and the affects that link the two (32), we can infer that a hypervigilant and anxious affect state will be linked to a perception of objects as persecuting and the self as victimized. Appreciation of this internal object relationship and its affect connection may influence the clinician’s psychotherapeutic approach.

The Role of Psychotherapy: Mentalization and Theory of Mind

The capacity to mentalize, or have a “theory of mind,” involves being able to recognize that someone else has a different mind from one’s own (33, 34). These terms also imply the ability to infer what is going on inside someone else’s mind by their facial expression, tone of voice, and other nonverbal communications. In essence, it is the ability to understand one’s own and others’ behaviors in terms of mental states such as beliefs, feelings, and motivations (35). Inherent in mentalization are an appreciation and recognition that the perceived states of one’s self and others are fallible and subjective and are representations of reality that reflect only one of a range of possible perspectives. Mentalization is created in the context of secure attachment with a caregiver who ascribes mental states to the child, treats the child as a mental agent, and helps the child to create internal working models (35). In other words, one automatically reads the expression on another’s face and knows what that person is feeling without extensive conscious effort to figure out the meaning of the facial expression. Hence it is not the same as conscious introspection. Neither is it identical with empathy. Mentalization refers specifically to the capacity to represent mental states of self and other. Empathy implies emotional resonance with another person. One can conceive of the mental state that may drive a person to murder without feeling empathy for that person.

In the absence of secure attachment, children have difficulty discerning their own mental states or those of others. A securely attached caregiver passes on this secure attachment and capacity to mentalize to the infant. Research has linked borderline caregiver passes on this secure attachment and capacity to mentalize to the infant. Research has linked borderline personality disorder patients with categories of insecure attachment—either preoccupied or unresolved/disorganized attachment (36–39). The failure to resolve trauma appears to distinguish the borderline personality disorder group from others. Early childhood trauma leads to a defensive withdrawal from the mental world on the part of the victim. Hence some patients with borderline personality disorder who have had severe trauma cope with the abuse by avoiding reflection on the content of the caregiver’s mind, which prohibits its resolution of abusive experiences (39, 40). One patient whose mother threatened to cut her hands off when she made a mess said that she stopped thinking about why her mother yelled at her because she was afraid her mother hated her and regarded her as a monster.

Fonagy et al. (41) studied an inpatient group that consisted predominantly of female patients with severe personality disorders. Using a reflective functioning scale that was developed to measure the capacity for mentalization (42), Fonagy et al. were able to quantify this dimension. Ninety-seven percent of the subjects with abuse and low reflective functioning met the criteria for borderline personality disorder. However, only 17% of the subjects reporting abuse in the group who had high reflective functioning met the criteria for borderline personality disorder. Hence patients with mentalizing capacity could understand the caregiver’s mind and process what happened so as to resolve the trauma. On the other hand, those who coped with abuse by refusing to think about what was going on in the caregiver’s mind failed to mentalize and therefore could not resolve the abuse experience.

In normal development, mentalization is a psychological achievement. A child younger than age 3 years operates primarily in a psychic equivalence mode (33). In this mode, the child assumes that perceptions of reality are identical to the reality itself. Around age 4 or 5 years, the child begins to integrate the pretend mode with the psychic equivalence mode of thinking. The 5- or 6-year-old child understands that one’s perception is influenced by subjective factors. This understanding allows for the phenomenon of play, where a child and a playmate can pretend to be others and perceive each other in those roles even though they are aware that the perception is different from the reality. Patients with borderline personality disorder often have great difficulty shifting from the psychic equivalence mode to the pretend mode, and this difficulty interferes with their capacity to recognize transference in psychotherapy. They often hold on to their perception as an absolute fact rather than viewing it as one of several possible alternatives, as the following case vignette illustrates:
Ms. A was a 28-year-old patient with borderline personality disorder in dynamic psychotherapy. About 6 months into the process, an apparently minor event in the therapy session triggered a major reaction in Ms. A. With about 5 minutes left in the therapy session, Ms. A was talking about having visited her family during the Thanksgiving holidays. She felt unimportant to her father because he seemed much more interested in her brother’s activities than in hers. In the course of this discussion, I looked at the clock on my wall because I knew the time was running out and I wanted to see if I had time to make an observation about her assumption regarding her father’s feelings about her. Ms. A stopped talking and looked at the floor. I asked her what was wrong. After a few seconds of silence, she burst into tears and said, “You can’t wait for me to get out of your office! I’m sorry if I’m boring you! I’ve known for a long time that you can’t stand me, and you just do this for the money. I’ll leave now if you want me to.” I was taken aback and replied, somewhat defensively, that I was simply monitoring the time because I wanted to be sure I had time to say something before the session was over.

Ms. A replied by saying, “Nice try to get out of it. You think I’m going to believe that?” Escalating in my defensiveness, I stated emphatically, “Whether you believe it or not, that’s the truth.” Ms. A was adamant: “I saw what I saw.” Placing her hand firmly on the wooden table next to her chair, she raised her voice: “It’s like you’re telling me that this table is not made out of wood!” Feeling as misunderstood as she was, I continued: “All I’m saying is this: it’s possible that I looked at the clock for reasons other than the ones you attribute to me—just like you may make assumptions about your Dad.” Ms. A became even more insistent in response to my efforts to offer other possibilities: “Now you’re trying to say I didn’t see what I saw! At least you could admit it!”

One of the greatest challenges for a psychotherapist is managing this almost delusional conviction of some patients with borderline personality disorder that their perception is a direct reflection of reality rather than a representation of reality based on their internal beliefs, feelings, and past experiences. This failure to mentalize may make it extremely difficult for them to work on transference issues because they are convinced that their view of the therapist is “correct” rather than one of a number of possible interpretations of the therapist’s behavior, facial expression, or comments. Fortunately, mentalization occurs on a continuum, and at times patients with borderline personality disorder may be capable of entering the pretend mode and reflecting on their own internal world and that of others. Whereas states such as autism are characterized by complete absence of mentalization on a neurological basis, a patient with borderline personality disorder often retains partial ability to mentalize under some circumstances, particularly when there is not an affectively intense involvement in an attachment relationship.

This vignette reflects how the misreading of the therapist’s mind led to an activation of a trauma-based internal object relationship associated with a hyperactive HPA axis. I became a potentially malevolent persecuting object; she became a victimized self; and a hypervigilant, anxious, humiliated affect state linked self and object. In this state of feeling terrorized, one cannot think or reflect. The intensity of Ms. A’s accusation also eroded my capacity to think, and I escalated my defensiveness to the point where I actually became a version of the persecuting object that she feared. This projective identification process, where the therapist is coerced under pressure from the patient into playing a role in the patient’s internal drama, can cause therapists to temporarily lose their capacity for mentalization, such that they cannot think their own thoughts in a psychotherapeutic role (43). In other words, I was insisting that only my version of reality was valid. Patients with borderline personality disorder colonize the minds of others as a way of extruding and controlling perceived danger from within. They unconsciously coerce the therapist into taking on the characteristics of an abusive internal object. I had become “bad” in two senses of the word—a bad object and a bad therapist. Although my interpretation that Ms. A’s misreading of me was similar to her misreading of her father may have been accurate, my timing was poor. Neither the patient nor I was in a reflective state of mind where meanings could be entertained. Our research in the Menninger Clinic Treatment Interventions Project (44) found that transference interpretation may need to be postponed under such circumstances until the patient’s ability to reflect returns.

Neural imaging studies suggest that mentalization entails several different brain structures working in concert (45–49). Most of these studies involved asking the subject to perform mental activities that require an understanding of someone else’s inner world. Recently Calarge et al. (48) asked 13 healthy volunteers to place themselves in another person’s place and attribute mental states to that person by having them describe the experience of a crying stranger met during a chance encounter on a park bench. The authors noted that these capacities are necessary in psychoanalytic psychotherapy practice. As in other studies, the medial frontal region was activated when the subjects attributed mental states to others. One of the most significant findings was that the largest activation during the task occurred in the right cerebellum. Like Frith and Frith (45), these investigators suggested that it is best to think of a “theory of mind” system or network that is widely distributed and made up of interactive nodes, probably in the medial prefrontal regions, the superior temporal sulcus, the inferior frontal region, and the cerebellum.

Mirror neurons may also play a role in a neurobiological basis understanding of mentalization. These neurons in the premotor cortex, first identified in monkey studies, respond when a primate observes certain hand movements performed by another primate or by a human or when the animal performs the same movements itself. In other words, these neurons are encoding object-oriented actions, whether they are performed or observed. This group of neurons in the ventral premotor cortex is activated during observation of an agent acting in a purposeful way...
upon objects (50). Fogassi and Gallese (50) suggested that mirror neurons may have a crucial role in goal detection and therefore in action understanding. They point out that the “mind reading” associated with theory of mind studies involves a series of explicit behavioral signals. They proposed that the capacity to understand another person’s internal world is related to the activation of this shared representation through mirror neurons. In other words, these neurons recognize intrinsically meaningful behavioral signals.

The fact that certain brain areas are activated during theory of mind experiments does not help the psychotherapist much when a patient with borderline personality disorder is sitting in the consulting room. However, the theory of mind construct helps bridge the domains of brain and mind. Within this conceptual model, mind becomes a sense of a subjective internal world accompanied by the recognition that others have internal worlds different from our own. There is no resort to Cartesian dualism in this model, but there is a recognition that subjectivity is extraordinarily complex and involves a language of meanings, perceptions, feelings, intentions, beliefs, and motivations that are not readily reducible to neuroscience constructs. Although the brain is an objective, observable entity, the mind of another is known through empathic connection. The vocabulary of the psychotherapist draws on the lexicon of the mind. To create a “mind” in the patient with borderline personality disorder requires the language of the mind.

Treatment Implications

The American Psychiatric Association (APA) practice guideline for the treatment of borderline personality disorder (51) suggests that a combination of psychotherapy and medication is the optimal treatment approach. Even though no randomized, controlled trials have compared medication alone with this combination or psychotherapy alone with this combination, the conventional wisdom is that medication may facilitate the psychotherapy of borderline personality disorder patients. The research reviewed earlier suggests ways in which medication and psychotherapy may work synergistically in the subgroup of patients with borderline personality disorder who have severe early trauma. The APA guideline notes that both dialectical behavior therapy and psychoanalytic/psychodynamic therapy are helpful for patients with borderline personality disorder. Here, I will confine my comments to the latter.

The findings of four double-blind, placebo-controlled trials (52–55) suggest that patients with serious personality disorders (principally borderline personality disorder) respond to optimal dosages of selective serotonin reuptake inhibitors (SSRIs) with improvements in anger, impulsive aggressive behavior (particularly verbal aggression), and affective lability. These agents may facilitate psychotherapy by reducing “affective noise”—such as intense anger, hypervigilant anxiety, or dysphoria—that prevents patients from reflecting on their internal world and the inner experiences of others. There is also growing evidence that SSRIs may actually stimulate neurogenesis, particularly in the hippocampus, leading to improved verbal declarative memory (56). In addition, SSRIs may reduce the hyperreactivity of the HPA axis by reducing hypersecretion of CRF. Nemeroff and Owens (57) reported that in rats, paroxetine reduces the increase in CRF mRNA expression after 3 weeks. As a result, CRF concentration and the increased HPA axis response to stress secondary to early life trauma are also reduced. When treatment with paroxetine stops, these same indices return to their usual abnormal values.

Rinne et al. (58) studied the effect of fluvoxamine on the HPA axis in 30 female patients with borderline personality disorder. They were given a combined dexamethasone/CRH hormone test. The test was administered before and after treatment with 150 mg/day of fluvoxamine. Seventeen of the patients had a history of sustained childhood abuse, and 13 had no abuse history. Both 6 and 12 weeks of fluvoxamine treatment were associated with a significant reduction of ACTH and cortisol response to the dexamethasone/CRH test. The magnitude of the reduction was directly related to the presence of sustained childhood abuse but not dependent on the presence of comorbid posttraumatic stress disorder or major depression. The investigators concluded that fluvoxamine reduces HPA hyperresponsiveness in borderline personality disorder patients who have a history of sustained childhood abuse.

The reduction of HPA axis hyperresponsiveness may directly affect the patient’s capacity to reflect. As noted earlier, the hypervigilant, anxious affective state is linked to a specific object relations unit in the patient that involves the perception of others as potentially malevolent persecutors and the self as a victim. In the state of being terrorized, one cannot think clearly. One can only react. Turning down the hyperreactivity with an SSRI facilitates thinking and reflecting. Without the intense affective state that existed before treatment with the SSRI, the patient can more easily consider other motives in the therapist. The patient can also have the luxury of reflecting on his or her own internal state. The patient can begin to see the therapist as someone there to help rather than persecute. Similarly, when the hypervigilant state in the patient is reduced, the therapist’s capacity to think psychotherapeutically is less likely to be eroded. When the therapist is on the defensive, as in the case reported earlier, defending oneself may become more prominent than working psychotherapeutically. Indeed, through projective identification the therapist develops a mental state that is similar to the patient’s (43, 59).

Use of an SSRI may help facilitate psychotherapeutic changes in the brain. The patient’s capacity to perceive the therapist as a helpful and caring figure instead of a persecuting and malevolent figure will serve to build up
new neural networks of internal representations while weakening the old ones (60). Splitting may be reduced as well, because lowered anxiety makes the defense less necessary. Hence better integration of left and right hemispheres may be promoted. Findings from a pilot study in Finland (61) that used single proton emission computed tomography imaging suggested that psychotherapy may even alter serotonergic metabolism in borderline personality disorder. Kandel (3) has emphasized that psychotherapy should be regarded as a “biological” treatment in that it leads to anatomical changes at the microcellular level.

These therapeutic changes take time. There is no “quick fix” for borderline personality disorder. Research in which this mentalization-based psychotherapeutic approach was used in conjunction with medication in a day hospital program indicated that it is highly efficacious but requires at least 12 months (62). However, the time investment is well worth it. A follow-up assessment at 18 months found that patients continue to make improvements after the therapy ends (63). The duration of treatment necessary in less intensive outpatient treatment will require further study.

In conclusion, personality disorders are best understood and treated without “either-or” dichotomies of brain and mind. Environmental influences on genes make nature-nurture distinctions difficult. Psychosocial factors produce biological changes in the brain. Medication and psychotherapy work synergistically to make changes in the brain over time. The languages of “brain” and “mind” are both necessary in the treatment of personality disorders. The language of “mind” is necessary for psychotherapy, but its effect is on the brain as well as on the creation of a “theory of mind” in the patient. Nemeroff et al. (64) found that psychotherapy is essential in the treatment of patients with chronic forms of major depression and a history of childhood trauma. The same is probably true for traumatized patients with borderline personality disorder. The model described here is applicable primarily to the subgroup of borderline personality disorder patients with childhood abuse and neglect. Further research is needed to identify models for those without that history.

Presented as the Adolf Meyer Award Lecture at the 157th annual meeting of the American Psychiatric Association, New York, May 1–6, 2004. Received Feb. 19, 2004; revision received May 20, 2004; accepted May 28, 2004. From the Menninger Department of Psychiatry, Baylor College of Medicine. Address correspondence and reprint requests to Dr. Gabbard, Menninger Department of Psychiatry, Baylor College of Medicine, One Baylor Plaza, MS HMC500, Houston, TX 77030; ggabbard@bcm.tmc.edu (e-mail).

References

14. Higley JD, Suomi S, Linnola M: CSF monoamine metabolite concentrations vary according to age, rearing and sex, and are influenced by the stressor of social separation in rhesus monkeys. J Psychopharmacol 1991; 103:551–556
23. Driessen M, Hermann J, Stahl K: Magnetic resonance imaging volumes of the hippocampus and the amygdala in women with borderline personality disorder and early traumatization. Arch Gen Psychiatry 2000; 57:1115–1122
34. Dennett DC: Beliefs about beliefs. Behav Brain Sci 1978; 1:568–569
57. Nemeroff CB, Owens MJ: Pharmacologic differences among the SSRIs: focus on monoamine transporters and the HPA axis. CNS Spectr 2004; 9:23–31
The Efficacy of Light Therapy in the Treatment of Mood Disorders: A Review and Meta-Analysis of the Evidence

Robert N. Golden, M.D.
Bradley N. Gaynes, M.D., M.P.H.
R. David Ekstrom, M.A., M.P.H.
Robert M. Hamer, Ph.D.
Frederick M. Jacobsen, M.D., M.P.H.
Trisha Suppes, M.D., Ph.D.
Katherine L. Wisner, M.D.
Charles B. Nemeroff, M.D., Ph.D.

Objective: The purpose of this study was to assess the evidence base for the efficacy of light therapy in treating mood disorders.

Method: The authors systematically searched PubMed (January 1975 to July 2003) to identify randomized, controlled trials of light therapy for mood disorders that fulfilled predefined criteria. These articles were abstracted, and data were synthesized by disease and intervention category.

Results: Only 13% of the studies met the inclusion criteria. Meta-analyses revealed that a significant reduction in depression symptom severity was associated with bright light treatment (eight studies, having an effect size of 0.84 and 95% confidence interval [CI] of 0.60 to 1.08) and dawn simulation in seasonal affective disorder (five studies; effect size=0.73, 95% CI=0.37 to 1.08) and with bright light treatment in nonseasonal depression (three studies; effect size=0.53, 95% CI=0.18 to 0.89). Bright light as an adjunct to antidepressant pharmacotherapy for nonseasonal depression was not effective (five studies; effect size=−0.01, 95% CI=−0.36 to 0.34).

Conclusions: Many reports of the efficacy of light therapy are not based on rigorous study designs. This analysis of randomized, controlled trials suggests that bright light treatment and dawn simulation for seasonal affective disorder and bright light for nonseasonal depression are efficacious, with effect sizes equivalent to those in most antidepressant pharmacotherapy trials. Adopting standard approaches to light therapy’s specific issues (e.g., defining parameters of active versus placebo conditions) and incorporating rigorous designs (e.g., adequate group sizes, randomized assignment) are necessary to evaluate light therapy for mood disorders.

The development of light therapy in psychiatry is closely intertwined with the original description of the syndrome of seasonal affective disorder. Two decades ago, Rosenthal and colleagues (1) described a series of patients with histories of recurrent depressions that developed in the fall or winter and spontaneously remitted during the following spring or summer. Their initial report also included preliminary findings indicating that bright artificial light, administered in a manner that would in essence extend the photoperiod, was more effective than dim light in treating seasonal affective disorder. The article presented an underlying hypothesis about the pathophysiology of the syndrome (i.e., depressogenic effects of melatonin), which in turn shaped the selection of treatment parameters: the intensity, duration, and timing of bright light exposure were designed to suppress the release of melatonin and lengthen the photoperiod.

Both seasonal affective disorder and bright light therapy quickly captured considerable attention, both in the scientific community and with the general public. Several research groups launched clinical trial programs, and soon this experimental treatment was extended to other conditions, including nonseasonal mood disorders, Alzheimer’s disease, circadian-related sleep disorders and jet lag, eating disorders, and other behavioral syndromes (2, 3). An international organization (the Society for Light Treatment and Biological Rhythms) was created, and several journals that emphasized phototherapy and biological rhythms emerged. Despite the growth in clinical and research programs, there remained an absence of recognition and support for light therapy within many segments of the psychiatric treatment community. Most insurers do not offer reimbursement for this treatment, most residency training programs do not provide clinical training in phototherapy, and there is a sense that “the biological psychiatry establishment has regarded light therapy with a certain disdain and relegated it to the edge of the paradigm” (4).

The American Psychiatric Association (APA) Council on Research requested that the APA Committee on Research on Psychiatric Treatments use the principles of evidence-based medicine to examine the efficacy of light therapy (J. Greden, personal communication). A work group was formed from members of the committee as well as outside consultants with expertise and experience in relevant disciplines. The work group completed a comprehensive literature review and meta-analyses. This report contains our findings about the efficacy of light therapy in the treatment of mood disorders in adult patients.
Method

Search Strategy

We searched PubMed for medical literature published from Jan. 1, 1975, to July 25, 2003. The search terms included “phototherapy” (which was the original term applied to light therapy) and any of the following terms: 1) “seasonal affective disorder”; 2) “depressive disorder”; 3) “bipolar disorder”; 4) “sleep” or “sleep disorder”; 5) “circadian rhythm” or “jet lag” or “melatonin”; 6) “Alzheimer’s disease” or “dementia”; 7) “premenstrual dysphoric disorder” or “premenstrual syndrome” or “late luteal phase dysphoric disorder”; 8) “eating disorder” or “bulimia” or “obesity”; 9) “serotonin”; and 10) “attention” or “vigilance” or “reaction time.”

Selection Criteria

Study groups were limited to adults who met a criterion-based mood disorder diagnosis. We restricted the age range to 18–65 years in an effort to define a standard for adequate treatment. We recognized that at each end of the age spectrum, the requirements for light therapy dosing may differ. For example, children and adolescents may differ from adults in the needed dose of light therapy, while elderly patients may require a higher dose of photons given the normal age-related clouding of the lens and ocular media, as well as possible reduction in the number of retinal photoreceptors in that population. Accepted criterion standards were DSM-III, DSM-III-R, DSM-IV, the Research Diagnostic Criteria (5), and the Rosenthal criteria (1). Subsyndromal diagnoses were excluded.

The included studies were required to be randomized, controlled trials of patients in the acute phase of treatment and to have a credible placebo control condition. Defining a minimum treatment dose (lux by time) for the experimental bright light treatment intervention was complicated by the absence of an accepted standard definition of adequate dosing. In some studies, we found that the “placebo condition” consisted of light exposure that was greater than that for the “active condition” in other studies. After we reviewed standard textbooks and consulted with expert clinicians in the field, we arrived at a priori definitions of adequate light therapy dose and duration. For bright light treatment of seasonal affective disorder, the definition was a minimum of 4 days of at least 3,000 lux-hours (e.g., 1,500 lux for 2 hours or 3,000 lux for 1 hour). We required placebo comparison groups to receive a maximum of 300 lux. For dawn simulation studies, we required that the active intervention consist of increasing light exposure from 0 to 200–300 lux over 1.0–2.5 hours and that the placebo condition consist of an increase that was less than 5 lux and/or less than 15 minutes in duration. For studies of bright light augmentation, we applied the same minimum lux criteria as those for bright light treatment of seasonal affective disorder and required that bright light therapy was the primary adjunct to the standard treatment under investigation.

We required that the outcomes be psychiatric symptom measures, e.g., Hamilton Depression Rating Scale, Seasonal Affective Disorders Version (6). Studies were excluded if gross protocol violations occurred (e.g., the study design was changed during the course of the trial).

Study Selection Process

Selection of the studies for inclusion involved two steps. First, two authors (R.N.G. and B.N.G.) independently reviewed the abstracts of all articles identified by the literature searches and excluded those for which they agreed that the eligibility criteria were not met. Next, the remaining articles were abstracted in detail (as described in the following) and the two authors made a final decision about inclusion or exclusion by consensus.

A total of 173 articles were identified and reviewed according to the selection criteria. This first step produced potentially relevant articles reporting 64 randomized, controlled trials (Figure 1). Of these, 50 involved patients with seasonal affective disorder and 14 involved patients with other depressive illnesses. These 64 trials received detailed abstraction, including clarification of subject group and selection, study design characteristics, and intervention parameters, to determine final eligibility. In addition, baseline demographic information and psychiatric outcome data were collected to allow direct comparisons of the experimental and control groups.

Data Analysis

All studies meeting the inclusion criteria were grouped by disorder and treatment type to produce four categories: bright light for seasonal affective disorder, bright light for nonseasonal depression, dawn simulation for seasonal affective disorder, and bright light as adjunctive treatment combined with conventional antidepressant pharmacotherapy for nonseasonal depression. The statistical information required for inclusion of a trial consisted of the mean score and standard deviation on a psychiatric symptom outcome measure and the number of subjects for each treatment condition; a trial was also included if the report contained sufficient information from which we could calculate the preceding data. For the articles in which the reported statistical detail was inadequate to allow meta-analytic procedures, several attempts were made to contact the corresponding author to gain...
<table>
<thead>
<tr>
<th>Treatment, Diagnosis, and Study</th>
<th>Treatment</th>
<th>Duration of Trial (days)</th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Condition</td>
<td>Illuminance (lux)</td>
<td>Time (hours/day)</td>
<td>Number of Patients</td>
</tr>
<tr>
<td>Bright light</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seasonal affective disorder</td>
<td>Avery et al., 2001 (9)</td>
<td>42</td>
<td>White light</td>
<td>10,000</td>
</tr>
<tr>
<td></td>
<td>Eastman et al., 1998 (10)</td>
<td>24</td>
<td>White light</td>
<td>6,000</td>
</tr>
<tr>
<td></td>
<td>Michalon et al., 1997 (11)</td>
<td>14</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Rosenthal et al., 1984 (1)</td>
<td>14</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Rosenthal et al., 1985 (12)</td>
<td>7</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Rosenthal et al., 1987 (13)</td>
<td>7</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Schwartz et al., 1997 (14)</td>
<td>21</td>
<td>White light</td>
<td>10,000</td>
</tr>
<tr>
<td>Nonseasonal depression</td>
<td>Baumgartner et al., 1996 (16)</td>
<td>7</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Kripke et al., 1992 (17)</td>
<td>7</td>
<td>White light</td>
<td>&gt;2,000</td>
</tr>
<tr>
<td></td>
<td>Volz et al., 1991 (18)</td>
<td>7</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td>Dawn simulation for seasonal affective disorder</td>
<td>Avery et al., 1992 (19)</td>
<td>7</td>
<td>White light (&quot;gradual dawn&quot;)</td>
<td>0–275</td>
</tr>
<tr>
<td></td>
<td>Avery et al., 1993 (20)</td>
<td>7</td>
<td>White light (&quot;gradual dawn&quot;)</td>
<td>0–250</td>
</tr>
<tr>
<td></td>
<td>Avery et al., 1994 (21)</td>
<td>7</td>
<td>White light (&quot;gradual dawn&quot;)</td>
<td>0–250</td>
</tr>
<tr>
<td></td>
<td>Avery et al., 1998 (22)</td>
<td>7</td>
<td>White light (&quot;gradual dawn&quot;)</td>
<td>0–250</td>
</tr>
<tr>
<td></td>
<td>Avery et al., 2001 (9)</td>
<td>42</td>
<td>White light (&quot;gradual dawn&quot;)</td>
<td>0–250</td>
</tr>
<tr>
<td>Bright light as adjunctive treatment of nonseasonal depression</td>
<td>Beuchum and Hays, 1997 (23)^a</td>
<td>7</td>
<td>White light</td>
<td>10,000</td>
</tr>
<tr>
<td></td>
<td>Fritzche et al., 2001 (24)^b</td>
<td>14</td>
<td>White light</td>
<td>2,500</td>
</tr>
<tr>
<td></td>
<td>Holsoer-Trachsel et al., 1994 (25)^c</td>
<td>16</td>
<td>White light</td>
<td>5,000</td>
</tr>
<tr>
<td></td>
<td>Muller et al., 1997 (26)^d</td>
<td>28</td>
<td>White light</td>
<td>5,000</td>
</tr>
<tr>
<td></td>
<td>Neumeister et al., 1996 (27)^d</td>
<td>6</td>
<td>White light</td>
<td>3,000</td>
</tr>
</tbody>
</table>

^a The other treatments were lithium, valproate, lorazepam, fluoxetine, desipramine, sertraline, clonazepam, haloperidol, benztpine, paroxetine, flurazepam, temazepam, amitriptyline, venlafaxine, zopiclone, diazepam, and pimozide; one patient received no drugs. The subjects continued their medications throughout the trial.

^b The other treatments were tricyclic antidepressants (N=32), selective serotonin reuptake inhibitors (SSRIs) (N=15), neuroleptics in low doses (N=15), and lithium or carbamazepine (N=6); 10 of the patients were treated with monotherapy and 30 with combinations. The medications were started before the light therapy and kept constant during the trial as much as possible.

^c All patients received trimipramine beginning 1 week before the light therapy period and continuing throughout the trial.

^d The other treatments were monotherapy with tricyclic antidepressants (N=6) or SSRIs (N=11) and combinations of these (N=3). The medications were started at least 3 weeks before the beginning of the study and kept constant during the trial.
access to the necessary raw data. For articles presenting outcomes from several treatment or control conditions, we pooled and analyzed the resulting statistics within conditions in order to use the largest study group available.

We employed standard meta-analytical methods, as described by Lipsey and Wilson (7). For each study, the standardized mean difference effect size and its 95% confidence interval (CI) were calculated and adjusted with the Hedges correction for small sample bias (8). For each of the four meta-analyses, the weighted mean of the component effect sizes and its 95% CI was calculated with the inverse variance weights from each study. The statistical significance of each weighted mean effect size was tested nondirectionally at the 0.05 level by using the z test.

If the number of subjects was sufficient, we performed homogeneity analyses of our results. If the required data were available, we calculated odds ratios for the likelihood of remission, defined a priori as a final score on the Hamilton Depression Rating Scale of 8 or less. We also applied the Q test, which uses the chi-square distribution to test the homogeneity of effect sizes across studies.

Results

Only 23 of the 173 studies identified during our literature search met our selection criteria (Figure 1). Of these, 20 distinct articles had sufficient data to allow inclusion in our meta-analysis (Table 1). One article (18) appeared to include a subset of patients who had been the subjects in a trial reported earlier (28), so the latter paper was excluded from this analysis. An additional report (9) contained the results of two experiments, one involving light therapy and another involving dawn simulation. Each was included in the appropriate, separate meta-analysis.

The results of the meta-analyses are shown in Table 2. We demonstrated significant effect sizes for bright light treatment of seasonal affective disorder (Figure 2), dawn simulation for seasonal affective disorder (Figure 3), and bright light treatment of nonseasonal depression (Figure 4). The effect size for bright light as an adjunctive treatment for nonseasonal depression was not significant (Figure 5).

Homogeneity analysis of the eight studies of bright light treatment of seasonal affective disorder was performed (for the other three meta-analyses, the number of studies was not large enough to support a Q test). The Q test indicated significant (p<0.0001) heterogeneity among studies; however, the effect sizes were consistently positive.

Odds ratios and their 95% CIs were calculated for the subset of four studies of bright light treatment of seasonal affective disorder for which the number of subjects who experienced remission was known. The summary estimate of risk of remission given treatment was an odds ratio of 2.9 (95% CI=1.6 to 5.4). The four study-specific odds ratios were significantly heterogeneous (p<0.04) but consistently positive.

Discussion

In our literature review, we found that most of the published research reports on the effects of light therapy in mood disorders did not meet recognized criteria for rigorous clinical trial design. There are several potential explanations for this observation. First, there are inherent challenges in creating an acceptable placebo (or even an active control) condition for light therapy. While it is relatively easy to create a placebo pill or capsule that is identical in appearance to an active medication formulation, it is more difficult to “blind” a subject when broad-spectrum intense white light is the active experimental intervention.
The pharmaceutical industry, which has considerable resources devoted to research and development activities, funds much of the clinical trial research for potential new antidepressant pharmacotherapies. In contrast, there has not been a similarly endowed industry nor as sizable a market in place to support the development and testing of light therapy treatments. The history of the development of light therapy, which is inextricably interwoven with the development of the concept of seasonal affective disorder, not surprisingly was dominated in its early stages by a series of relatively small, investigator-initiated pilot projects. These researchers did not have access to resources of the magnitude of those available when pharmaceutical companies seek recognition by the U.S. Food and Drug Administration of the safety and efficacy of a new medication. Unfortunately, a consensus about standard approaches to study design issues (such as lux parameters for the active treatment, duration of an adequate light therapy trial, and characteristics of placebo control conditions) was not established in the early years of light therapy research. In too many cases, high standards of research design (such as random assignment to treatment conditions, adequate reporting of results statistics) were not followed. Not surprisingly, these conditions produced inconsistencies in the research literature. We found substantial variability in the selection of study groups and in the doses of both the active and control interventions for those trials meeting our selection criteria.

All of these factors have limited the conclusions that can be drawn from light therapy research. Past efforts to synthesize the available body of literature have been as challenging as the proverbial comparison of apples and oranges. More important, the limitations in much of the literature on light therapy research may have created the unsubstantiated impression that the treatment itself has limitations in terms of its efficacy.

When we analyzed the data from all available randomized, controlled trials that met our a priori standards, we demonstrated a significant reduction in depression symptom severity following bright light therapy in seasonal affective disorder and in nonseasonal depression, as well as a significant effect with dawn simulation in seasonal affective disorder. In other words, when the “noise” from unreliable studies is removed, the effects of light therapy are comparable to those found in many antidepressant pharmacotherapy trials (29).

Earlier reviews of light therapy yielded similar findings. Terman et al. (30) pooled data from 14 research groups that collectively studied 332 patients who received bright light therapy for seasonal affective disorder over a 5-year period, and they applied a pooled clustering technique in their analysis. Twenty-nine data sets were included. Unfortunately, the vast majority were not available for inclusion in our current analysis, because they consisted of personal communications, unpublished posters presented at meetings, and book chapters, as well as a few additional reports that did not meet our inclusion criteria. Thus, only two of their 29 data sets overlap with the 20 studies included in our meta-analyses, i.e., two studies by Rosenthal et al. (1, 12). Terman et al. (30) found that 2,500-lux light exposure for at least 2 hours/day for 1 week resulted in significantly more remission when administered in the early morning than in the evening or at midday. Treatments at each of these three administration times were significantly more effective than control treatments with dim light. Tam et al. (31) concluded that bright light therapy that utilized at least 2,500-lux white light for 2 hours/day and treatment with 10,000 lux for 30 minutes/day had comparable response rates and that both treatments were efficacious. They noted that more studies were needed before conclusions could be drawn about the efficacy of dawn simula-
tion. They highlighted the methodological limitations of the literature, which included brief treatment periods, small study groups, and lack of replication.

Several caveats and limitations in our review and analyses should be noted. First, we limited our focus to efficacy and did not study the other key feature of all treatments, safety. Very few reports of the controlled studies contained data on side effects or toxicity. Several side effects of bright light therapy have been described elsewhere, including headache, eye strain, nausea, and agitation (32, 33). To our knowledge, there have been no reports to date of retinal toxicity in association with bright light treatment, and a 5-year follow-up study showed no adverse ocular effects (34). However, some psychotropic medications may increase photosensitivity, and further study of potential adverse effects of combined pharmacotherapy and light therapy is indicated. Light therapy, like other antidepressants, may be associated with a switch to hypomania or mania in vulnerable bipolar patients (35). Other rare potential side effects from bright light treatment may emerge only after the treatment has become more widely applied. Thus, any potential recommendation of light therapy for mood disorders, based on findings of efficacy in our meta-analyses, must be tempered by the acknowledgment that safety must also be considered. This important aspect of light therapy merits careful examination with additional long-term follow-up studies.

Another limitation in this study, as described in our methods section, is that we restricted our analyses to studies of a relatively homogeneous, clearly defined population (i.e., nongeriatric adult patients). There are published reports of light therapy for seasonal affective disorder in children (e.g., references 36 and 37) and for mood disorders in the elderly (38). These important patient populations merit separate consideration, and there is a need for a larger evidence base in these areas. An additional potential application of light therapy lies in the treatment of depression during pregnancy and in the postpartum period, when safe and effective alternatives to pharmacotherapy without potential toxicity for the fetus or newborn would be clearly desirable (39, 40). It should be noted that all of the studies of dawn simulation in our meta-analysis came from a single research group, and confirmation of their findings by others at different locations would be especially important in determining the generalizability of their results. Finally, by setting a reasonably high standard for study inclusion in our meta-analysis, we excluded many of the published reports in this area. One could argue, as Smith et al. (41) did in another context, that many small, “imperfect” studies can “converge on a true conclusion.” However, we agree with those who believe that meta-analyses based on flawed studies are not useful and that some bodies of data are inadequate for supporting a proper meta-analysis (42).

This study suggests that certain types of light therapy are effective in the treatment of seasonal affective disorder and other forms of depression. Much of the available literature is limited in terms of study design, and additional randomized, controlled trials with appropriate numbers of subjects are needed. Remaining questions of efficacy, safety, optimum dose, and the proper place of light therapy in the psychiatrist’s toolbox may be answered only after investigators in the field define and consistently adhere to standard approaches to essential components, such as definitions of acceptable parameters for active treatment and control conditions.

References

15. Terman M, Terman JS, Ross DC: A controlled trial of timed bright light and negative air ionization for treatment of winter depression. Arch Gen Psychiatry 1998; 55:875–882
Studies of brain lesions in humans have established an important role of the hippocampus in learning and memory. For example, the ability to remember facts and events, referred to as declarative memory, relies on an intact hippocampus in humans. More recently, neuroimaging studies have further clarified the role of the hippocampus for human memory. One such aspect of memory, demonstrated via blood flow changes as measured with functional magnetic resonance imaging (fMRI), is relational memory.

The ability to learn relationships between items is an essential property of human intelligence. For example, having learned in panel A of the figure that A>B and B>C, we can infer from these two relationships that A>C. This ability, referred to as transitive inference, has previously been linked to the hippocampus in animal and human experiments. The image in panel B shows right anterior hippocampal activation during a transitive inference task for arbitrary visual stimulus patterns. The hippocampus was specifically activated during the recognition of novel pairs of stimuli (such as A and C in panel A), for which the relationship between items could be correctly identified on the basis of prior learning. Transitive inference can be demonstrated in both humans and animals and may have evolved throughout phylogenesis as a beneficial strategy for survival, e.g., to infer the hierarchy of inhabitants in an environment. Recent experiments indicate that specific functions of relational memory, such as transitive inference, are impaired in psychiatric patients, while other aspects of declarative memory remain intact. Such evidence points to anatomically distinct lesions within the human hippocampus, impairing some but not all of its functions.

STEPHAN HECKERS, M.D.
Belmont, Mass

DEBRA TITONE, PH.D.
Montreal
When residents report on patients, I emphasize repeatedly that although the history and many aspects of the mental status examination are undeniably important, I think of the gaze, our patients’ and ours, as particularly important. Some of our gazes must serve as a diagnostic tool, but much of them are a point of connection to patients, and I am reminded of a much earlier experience, when I was a medical student.

I had been assigned to work in the pediatric oncology clinic, and on one of those days, I was asked to evaluate a young girl, about 9 years old, who had come in with her parents for evaluation of her malignant astrocytoma. The attending physician prefaced my going into the room with just the slightest comment, insinuating that the visit was perfunctory and that she was not doing well at all. As I entered the examination room, the parents were seated rather still, with a somewhat hunched posture, conveying a level of fatigue, their very cushingoid daughter nearby. I explained that I was a medical student and would be first examining their daughter, to which the mother nodded as if to say “yes” but did so very reflexively. And her husband, a good bit more solemn in expression, never said anything. Thereafter, with ophthalmoscope in hand, I bent over, in a sense brushing past the child’s gaze for the sake of looking into the back of her eyes.

For all of the times one is told of a cardinal feature of one disease entity or another, it is a shock to finally see something as bold as the feature itself in unbridled form. There in the back of her eyes was the most glaring example of papilledema I had ever seen, and it was startling. I almost had to back away for a bit to reconsider what I had seen, as if to ponder its veracity only to approach again to convince myself that it was so. But in likely just as short an amount of time, I knew in the most exacting of terms that I had seen what other students and doctors had been observing and that it spoke clearly of disease, deterioration, and pending death. The ophthalmoscope having translated to a virtual crystal ball, I was looking at evidence of a child who would soon expire.

As I backed away from the girl, righting myself somewhat slowly, she scanned my face, somehow in search of the answer to which we, as doctors, were so directed. “What is happening to me?” she seemed to say. She stood so passively, her long hair shrouding her swollen face, quietly licking her lips very anxiously and fidgeting a bit, exuding enormous fright and vulnerability. I felt entirely overwhelmed, and my inclination was to hold her, although I didn’t. Because I couldn’t convey anything good, I averted my gaze a bit toward her parents, who simply looked back, the mother’s eyes scanning me slightly. And then the mother smiled, just slightly and rather weakly, I would add, conveying that she already knew the answer. I recall stumbling over a few words to explain the findings.

“Well, uh…there is still swelling. I, uh…I think that my attending will be right in.”

It was impossible to know what else to say. For what felt like a very long time, our silence persisted, and we simply exchanged glances. That was all. We transcended ophthalmoscope readings to now speak silently of the balance between life and death. I’m not sure that words would have done it nearly enough justice, perhaps only proving cluttering, disturbing a reverence for our collective struggle.

Without a doubt, the young girl has passed away by now, her parents devastated by her death. They can look at photos of when she was younger, colors entrapped in a photo gel that speak of a time when everything seemed just so much brighter, but they...
are undoubtedly crushed under the weight of her absence. In many respects, I am crushed by her absence, too, and I also write this as a long-overdue eulogy to her, but I can no longer exchange gazes with them to tell them so. Our gazes can now only be directed within, to help us reflect on all that we’ve seen and felt.

Certainly sans implements, one has to face patients and tell them in so many words, gaze included, what our eyes have allowed us to see. You, as the reader, are currently using your own gaze to read prose conveying something of what I’ve seen and felt, expectantly devoid of the exact experience of what I saw in a certain moment of time. One is thereby solely reliant on descriptive prose, an amalgamation of words, as a means of conveyance of sight, but it can feel a bit lacking. Although we place an enormous emphasis on verbal expression, and it is vital, some conduit to connection with patients, words will sometimes be faltering, nondescript, laundered, or simply beside the point. This does seem to speak, as it were, to the impasse of words alone, written or spoken, not encompassing enough of the human experience. When I emphasize gaze to residents, therefore, it is with the hope that it will be more far-reaching than simple use of a diagnostic gaze and an admonishment to not be solely reliant on the exchange of words. Our interchanges with patients extend much beyond verbal expression. If we really allow ourselves to look within, our gazes and those of patients will reveal some very hidden and profound feelings, and they will serve a pivotal role in empathic connection. Sometimes, we may simply have to speak in silence. We hope our eyes will say it all.

LLOYD BENJAMIN, M.D.

Address correspondence and reprint requests to Dr. Benjamin, UCDMC, 2230 Stockton Blvd., Sacramento, CA 95817; lloben@earthlink.net (e-mail).
Juliano Moreira (1873–1933): Founder of Scientific Psychiatry in Brazil

At the beginning of the 20th century, the Afro-Brazilian doctor Juliano Moreira, adopting many of Kraepelin's ideas on classification, started a theoretical renovation of Brazilian psychiatric ideas, which up to this point had a marked French accent. In 1903, he was invited by the new republican government to take charge of the National Mental Hospital. While serving there as director, he undertook an ambitious project of renovation, which changed considerably the way the asylum operated. Under his influence, for example, lumbar puncture started to be done in Brazil, as a “modern” way of diagnosing cerebral syphilis and other neuropsychiatric disorders (1).

Moreover, he devoted much time to the study of cross-cultural psychiatry (Kraepelin’s “comparative psychiatry”). He especially investigated dementia caused by syphilis, concluding, against dominant views, that the racial condition would neither immunize against nor favor the emergence of such condition. According to him, regardless of climate or race, “dementia paralytica” (and mental disease in general) was equally present in every human being, due mainly to factors related to previous health conditions, concomitant diseases, and sociocultural status (2).

In Moreira’s time, most Brazilian intellectuals endorsed racial biases and prejudices regarding medical and scientific ideas (3). This was not the case for Moreira, who took a clear and courageous stance, both personally and scientifically, by carefully investigating mental diseases in the different Brazilian ethnic groups and fighting the intense and widespread color prejudice of his time.

References

Conversion From Mild Cognitive Impairment to Probable Alzheimer’s Disease Predicted by Brain Magnetic Resonance Spectroscopy

Pedro J. Modrego, M.D.
Nicolás Fayed, M.D.
Miguel A. Pina, M.D.

Objective: Mild cognitive impairment has been regarded as a pre-Alzheimer condition, but some patients do not develop dementia. Given the available therapies for Alzheimer’s disease, early diagnosis is of paramount importance. The authors’ objective was to determine whether findings from magnetic resonance spectroscopy (MRS) of the hippocampus and other cortical areas would predict conversion from amnestic mild cognitive impairment to probable Alzheimer’s disease.

Method: A longitudinal inception cohort of 53 consecutive and incident subjects fulfilling the criteria of amnestic mild cognitive impairment was followed for a mean period of 3 years. At baseline, a neuropsychological examination (Mini-Mental State Examination, Blessed Dementia Rating Scale, Clinical Dementia Rating, verbal fluency test, and memory tests) and standard blood tests were performed, and three cortical areas were examined by proton MRS: left hippocampus, right parietal cortex, and left occipital cortex. The patients were evaluated periodically to detect conversion to probable Alzheimer’s disease. The statistical analysis of predictions was based on receiver operating characteristic curves.

Results: By the follow-up assessment that occurred on average after 3 years, 29 patients (55%) had developed probable Alzheimer’s disease. An occipital cortex N-acetylaspartate/creatine ratio ≤1.61 predicted dementia at 100% sensitivity and 75% specificity (area under the curve=0.91, 95% CI=0.80–0.97). The positive predictive value was 83%, and the negative predictive value was 100%, with an overall cross-validated classification accuracy of 88.7%. None of the values in the hippocampus and parietal cortex had significant predictive value.

Conclusions: MRS of the brain performed on patients with mild cognitive impairment is a valuable tool in predicting conversion to probable Alzheimer’s disease. Occipital values were more reliable than hippocampal values in this prediction.

(Mem J Psychiatry 2005; 162:667–675)

Mild cognitive impairment may be defined as a transitional state between normal aging and Alzheimer’s disease in which memory impairment is greater than expected for age, but general cognitive function and daily living activities are preserved (1). The rate of progression to dementia is considerable according to estimates from various longitudinal studies: 12% per year for 4 years (1), 40% over 2 years (2), 53% over 3 years (3), 34%–100% over 4–5 years (4, 5), and 100% over 9.5 years (6).

Currently, there has been open discussion about terminology and on whether mild cognitive impairment must be considered either a separate nosological entity or a prodrome of Alzheimer’s disease (7). Whatever the option, it is of paramount importance to develop tools that enable us to predict conversion to dementia.

The availability of drug therapy for mild-to-moderate Alzheimer’s disease has led to an emphasis on the early detection and management of this disorder. As therapies with the potential to halt progression of Alzheimer’s disease are developed, identification of people with pre-Alzheimer conditions has become increasingly important to determine whether such individuals can benefit from interventional agents.

Previous reports have pointed to the major role of the hippocampal complex in memory (8, 9). Temporal atrophy and especially hippocampal atrophy have been detected in early Alzheimer’s disease (10, 11). Another report concluded that memory dysfunction was highly correlated with atrophy of the amygdala–hippocampal complex and of the temporal lobes (12). Magnetic resonance imaging-based hippocampal volume measurements in patients with mild cognitive impairment have demonstrated that anterior hippocampal formation is most affected in comparison with normal aging, and that the left parahippocampal gyrus shrinks disproportionately to the right (13). Longitudinal assessments of hippocampal atrophy rate support the hypothesis of a continuum from normal aging to mild cognitive impairment and Alzheimer’s disease (14). Hippocampal atrophy has been regarded as a
significant predictor of dementia in subjects with mild cognitive impairment (15).

Proton magnetic resonance spectroscopy (1H-MRS) studies have found decreased N-acetylaspartate and increased myo-inositol in the occipital, temporal, parietal, and frontal regions of patients with Alzheimer's disease even at the early stages of Alzheimer's disease (16). N-Acetylaspartate is one of the most abundant amino acids in the central nervous system, located predominantly in neurons, axons, and dendrites. Its function is not well known so far and remains speculative. It has been thought to play a role in lipid and protein synthesis, but it also may be a product of N-acetylaspartyl-glutamate degradation or an osmolyte. Because this metabolite is decreased in some neurological diseases such as degenerative dementia, it has been labeled as a surrogate marker of neuronal integrity (17). The sugar-alcohol compound myo-inositol may act as an osmoregulator, intracellular messenger, and detoxication agent; it is also regarded as marker of glial cells (18).

Relative to healthy subjects, a reduced concentration of N-acetylaspartate has been found in the medial temporal lobe (21% difference) and the cortex of the parietal lobe (13% difference) of patients with Alzheimer's disease as well as a smaller hippocampus (29% difference) without significant differences between both sides (19). Poor performance on recognition memory tests was correlated with low gray matter N-acetylated compounds and increased choline concentrations in Alzheimer's disease patients (20). Significantly lower N-acetylaspartate/creatine ratios were found in both hippocampi of 34 Alzheimer's disease patients in comparison with 22 healthy subjects (21), but in another study comparing Alzheimer's disease patients and healthy subjects the differences were only significant for the left hippocampus (22). However, to our knowledge there have been no longitudinal studies that used 1H-MRS in patients with mild cognitive impairment as a predictor of conversion to Alzheimer's disease.

On the basis of these reports, we hypothesized that hippocampal alterations in patients with mild cognitive impairment detected with 1H-MRS would predict conversion to Alzheimer's disease when the patients are followed longitudinally. It is also likely that these alterations may appear early in the occipital and parietal cortices.

**Method**

The patients we see in our outpatient clinics are usually referred from the community by family physicians. Given the high proportion of elderly people in our area, mild cognitive impairment and dementia of Alzheimer type are relatively frequent diagnoses in our outpatient clinics.

Before the beginning of the study we asked the general practitioners to refer to us all elderly patients with recent memory complaints. From this group we recruited a cohort of 59 consecutive incident patients who fulfilled the criteria of amnestic mild cognitive impairment proposed by Petersen et al. (1): memory complaints, normal activities of daily living, normal general cognitive function, and abnormal memory for age but no dementia. We also gave special importance to the fact that the memory inefficiencies of the patient had to be corroborated by an informant (wife/husband or relative).

In the clinical history we asked for family history of dementia, vascular risk factors, depression, and other diseases. The final diagnosis was made by taking into account as much information as possible from all sources (general interview, neuropsychological examination, informant report). We only considered patients with a disease duration of up to 12 months as being incident cases of mild cognitive impairment.

Neuropsychological examination at baseline included the Spanish version of the Mini-Mental State Examination (MMSE) (which has a maximum score of 35 points and a cutoff point of 23 for elderly subjects (23)), Blessed Dementia Rating Scale, and category fluency test (Set Test). Memory was explored with subscales of the 144 Battery of Signoret (Spanish validated version) (24):
of vitamin B12, folic acid, and TSH; a proteinogram; and serologic tests; measurement of sedimentation rate, cell count, and levels Depression Scale (25). Laboratory tests included standard blood poses. Depression was initially assessed by means of the Geriatric viation from the mean of healthy subjects.

Every spectroscopy was performed on a 1.5-T clinical scanner (Signa Horizon, GE Medical Systems, Milwaukee). Sagittal T1-derwent 1H-MRS of the brain in three different areas (Figure 1):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients With Mild Cognitive Impairmenta</th>
<th>Normative Group (N=20)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini-Mental State Examinationb score</td>
<td>Mean 27.6 SD 3.6 Range 20–34</td>
<td>Mean 24.5 SD 3.8 Range 20–34</td>
<td></td>
</tr>
<tr>
<td>Set Test score</td>
<td>33.4 3.3 27–40</td>
<td>32.2 3.1 25–40</td>
<td></td>
</tr>
<tr>
<td>Blessed Dementia Rating Scale score</td>
<td>2.5 1.0 2–4</td>
<td>2.1 1.0 1–3</td>
<td></td>
</tr>
<tr>
<td>Geriatric Depression Scale score</td>
<td>8.6 5.2 0–26</td>
<td>6.0 3.5 0–25</td>
<td></td>
</tr>
</tbody>
</table>

Memory performance
1. Immediate verbal recall of a maximum of 12 words with three trials, as well as delayed recall of those words. The maximum possible score is 12 points for immediate recall and 12 for delayed recall.
2. Immediate visual recall: the patient has to reproduce a complex geometric figure after 1 minute of observation. Delayed visual recall was also evaluated. The maximum possible score is 24 points for each.
3. Logical memory: immediate and delayed recall of a brief story. Maximum score=24 points for each.
4. Digit span (number of digits forward).

Memory functions were also assessed in a random comparison group consisting of 20 healthy elderly subjects recruited from the same population so as to demonstrate baseline differences between patients with mild cognitive impairment and subjects without mild cognitive impairment and to establish the cutoff point. We defined abnormal memory as a total memory score <60, which represented a difference greater than one standard deviation from the mean of healthy subjects.

The patients were also evaluated with the Clinical Dementia Rating Scale and the Global Deterioration Scale for staging purposes. Depression was initially assessed by means of the Geriatric Depression Scale (25). Laboratory tests included standard blood tests; measurement of sedimentation rate, cell count, and levels of vitamin B12, folic acid, and TSH; a proteinogram; and serologic tests for syphilis. The final diagnosis of amnestic mild cognitive impairment was made when the patients fulfilled the criteria of Petersen et al. (1) and an informant’s report corroborated the memory complaints. We interpreted with caution some low scores yielded in neuropsychological examination, since some patients had a low level of education. Scores higher than 60 on the memory scale did not exclude the diagnosis of mild cognitive impairment if the clinical history and caregiver report were clearly compatible with memory loss.

Upon completion of the clinical history and neuropsychological examination, the patients with mild cognitive impairment underwent 1H-MRS of the brain in three different areas (Figure 1): left hippocampus, right parietal cortex, and left occipital cortex. Every spectroscopy was performed on a 1.5-T clinical scanner (Signa Horizon, GE Medical Systems, Milwaukee). Sagittal T2-weighted topogram and T2-weighted axial localizing series were used to locate volumes (8 cm3) in cortical areas. Single-voxel 1H-MRS was performed by means of an echo time of 30 msec and a repetition time of 2500 msec with spin-echo technique that uses selective excitation with gradient spoiling for water suppression. The mode of spectral acquisition was Probe-p (PRESS technique).

Data were transferred to a Sun Microsystems workstation and analyzed in an automated manner. Free induction decays were zero filled, exponentially filtered (line broadening=1 Hz), and transformed. Each spectrum was automatically fitted to four peaks corresponding to levels of N-acetylaspartate (2.02 ppm), total creatine (3.03 ppm), choline-containing compounds (3.23 ppm), and myo-inositol (3.56 ppm). We also calculated the ratios of the peak amplitude of the metabolites relative to creatine and the N-acetylaspartate/myo-inositol ratio in each of the three areas of exploration.

For reproducibility purposes we repeated the exploration after 2 months in 10 patients with the same localizing protocol.

Every patient was followed and reevaluated every 6 months—or earlier if required by caregivers or family physicians—for a mean period of 3 years (range=2.5–3.5 years). Patients with depressive symptoms were adequately treated in the psychiatry unit and reevaluated after 2 months so as to clarify that those patients really met the criteria for mild cognitive impairment. The main end point to be considered was the development of dementia.

Criteria for probable Alzheimer’s disease per the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) (26) were applied to the patients who became demented.

The demographic, clinical, and radiological data of patients whose baseline mild cognitive impairment did or did not convert to probable Alzheimer’s disease were analyzed separately and compared statistically. Quantitative variables were compared with t tests for independent samples. The predictive value of the different measurements obtained in baseline 1H-MRS findings until the end of the follow-up.

Criteria for probable Alzheimer’s disease per the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) (26) were applied to the patients who became demented.

The demographic, clinical, and radiological data of patients whose baseline mild cognitive impairment did or did not convert to probable Alzheimer’s disease were analyzed separately and compared statistically. Quantitative variables were compared with t tests for independent samples. The predictive value of the different measurements obtained in baseline 1H-MRS findings was analyzed by means of receiver operating characteristic curves. For this purpose we used the Med-Calc software, which selects the best cutoff for prediction taking into account sensitivity, specificity, likelihood ratio, and predictive values. Complementary statistical analysis was based on a cross-validation technique with the option discriminant analysis/classify/leave-one-out technique of the SPSS, version 10 (SPSS, Inc., Chicago).

Informed consent was obtained from patients and relatives.
Results

At the beginning of the study we had to exclude four patients who did not tolerate the practice of MRI. Therefore, 55 patients were included in the final follow-up assessments. The average age was 72.7 years (SD=5.3, range=61–84). There were 34 women and 21 men. The mean follow-up length was 3 years (range=2.5–3.5), and the clinical assessments were done every 6 months. Each patient was examined six times on average (SD=1).

The baseline clinical variables are summarized in Table 1, including the memory performance of a normative group of 20 healthy subjects (11 women and nine men, mean age=71.4 years). We found significant between-group differences in scores on most memory subscales and total memory score according to two-tailed t tests. All patients with mild cognitive impairment scored 0.5 on the Clinical Dementia Rating and 3 on the Global Deterioration Scale. Six patients scored below 23 on the Spanish version of the MMSE, but this was due to low educational level and not to mental deterioration.

In Table 2 we present the mean metabolite values and ratios from the first and second scans of 10 patients and...
found no significant changes. In Table 3 are reported the baseline mean values of the different metabolite ratios found in the three areas of exploration. In Figure 2, scatter-plots representing the N-acetylaspartate/creatinine ratios for the three areas explored are shown.

After 1 year, one patient was lost, and another nondemented patient had died from cardioembolic stroke. The remaining 53 patients completed the follow-up period. Of these 53 patients with mild cognitive impairment at baseline, 13 (25%) converted to a diagnosis of probable Alzheimer’s disease per NINCDS-ADRDA criteria. At this stage we found that an occipital cortex N-acetylaspartate/creatinine ratio equal to or lower than 1.61 predicted conversion to dementia at 100% sensitivity and 46% specificity; the area under the curve was 0.69 (95% CI=0.55–0.80). In the hippocampus, a myo-inositol/creatinine ratio higher than 1.04 predicted conversion at 66.7% sensitivity and 72% specificity, with an area under the curve of 0.66 (95% CI=0.51–0.79). None of the remaining variables predicted conversion to dementia with statistical significance as the value 0.5 was included in the confidence intervals.

After a mean follow-up period of 3 years, 29 (55%) out of 53 patients converted to a probable Alzheimer’s disease diagnosis, and 24 remained with memory loss but did not exhibit dementia. We did not encounter any other type of dementia (Lewy body disease, frontotemporal dementia, or vascular dementia). After this period two patients died after developing dementia. None of the patients reverted to normality from mild cognitive impairment or reverted to mild cognitive impairment from dementia. Treatment for depression in 15 patients barely improved memory performance.

Only the occipital cortex N-acetylaspartate/creatinine ratio at baseline predicted conversion to dementia with confident values (Table 3 and Table 4). The mean ratio for demented patients was 1.46 (SD=0.11) and for patients free of dementia it was 1.63 (SD=0.08), a significant difference. Figure 3 shows spectrum examples with the metabolite peaks for a nonconverter and converter, respectively. N-Acetylaspartate was significantly decreased in relation to creatine in converters. The mean occipital cortex myo-inositol/creatinine ratio was 0.69 for nonconverters and 0.72 for converters, which was not significant (p=0.50). In the receiver operating characteristic curve analysis, an occipital cortex N-acetylaspartate/creatinine ratio equal to or lower than 1.61 predicted conversion at 100% sensitivity and 75% (95% CI=53%–90%) specificity. The area under the receiver operating characteristic curve (Figure 4) was 0.91 (95% CI=0.80–0.97), with a positive predictive value of 83% and a negative predictive value of 100%. The cross-validation analysis yielded a classification accuracy of 88.7% (canonical correlation=0.8, Wilk’s lambda=0.37, $\chi^2=49.1$, p=0.0001). The addition of other variables such as the MMSE score or other metabolite ratios did not improve the model of predictions in the discriminant analysis.

If the cutoff point were ≤1.46 for the N-acetylaspartate/creatinine ratio, then specificity would be 95% (95% CI=78%–99%), but sensitivity would decrease to 55% with a positive predictive value of 94% and a negative predictive value of 64%.

The remaining occipital cortex ratios and ratios obtained from the hippocampal and parietal regions did not predict conversion significantly (Table 3 and Table 4). A choline/creatinine ratio equal to or lower than 0.63 in the right parietal cortex disclosed a 42.9% sensitivity and a 90.2% specificity with a positive predictive value of 85.7%. The absolute metabolite concentrations were not of value for predicting purposes. All predictions were based on baseline $^1$H-MRS scans.

We also observed that none of the following baseline variables predicted conversion to dementia with confidence: MMSE score, memory performance, verbal fluency, age, or family history of dementia. After 1 year, a score equal to or lower than 26 on the MMSE (Spanish version with a maximum of 35 points) predicted conversion at 61.5% sensitivity and 83.3% specificity, with an area under the curve of 0.71 (95% CI=0.55–0.84). At the 3-year follow-up assessment, the values of prediction for a cutoff point of 29 were as follows: 76.9% sensitivity, 68.2% specificity, and an area under the curve of 0.76 (95% CI=0.60–0.88). The values were below 70% for verbal fluency scores and memory performance scores and, therefore, far from being clinically useful.
Memory impairment is a frequent complaint in the elderly, but hitherto we did not have any parameter objective enough to predict conversion to Alzheimer’s disease. Numerous studies have focused on this issue; some of them have used neuropsychological tests as predictors, others have used neuroimaging techniques. Although most studies are interesting, the matter has not been completely resolved because of either insufficient sensitivity or specificity of the tools used.

A cohort of 195 patients with questionable dementia underwent comprehensive neuropsychological examination to predict conversion to dementia, but these predictions achieved at most 64% sensitivity and 76% specificity for verbal recognition and verbal fluency (27). There have been some reports suggesting that hippocampal volume changes in patients with mild cognitive impairment may help predict conversion to Alzheimer’s disease (15, 28, 29). However, the subject of which region of the brain, entorhinal cortex or hippocampus, demonstrates the earliest abnormalities in early Alzheimer’s disease has not yet been elucidated because of the large variability of imaging measures, the intersubject variability of brain anatomy, and the imaging artifacts in the medial temporal region (30). In addition, some degree of hippocampal atrophy may be seen in up to one-third of cognitively normal old people (31). It is clear that we need more powerful objective pa-

**TABLE 4. Regional Metabolite Ratios at Baseline as Predictors of Conversion to Probable Alzheimer’s Disease in 53 Subjects With Mild Cognitive Impairment**

<table>
<thead>
<tr>
<th>Region and Metabolite Ratio</th>
<th>Baseline Level</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Likelihood Ratio</th>
<th>Predictive Value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hippocampus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Acetylaspartate/creatine</td>
<td>≤1.29</td>
<td>63.0</td>
<td>52.6</td>
<td>1.33</td>
<td>0.70</td>
</tr>
<tr>
<td>Choline/creatine</td>
<td>≤0.97</td>
<td>55.6</td>
<td>78.0</td>
<td>2.64</td>
<td>0.56</td>
</tr>
<tr>
<td>myo-Inositol/creatine</td>
<td>&gt;1.02</td>
<td>61.5</td>
<td>70.6</td>
<td>2.09</td>
<td>0.54</td>
</tr>
<tr>
<td>N-Acetylaspartate/myo-inositol</td>
<td>≤1.22</td>
<td>70.4</td>
<td>75.0</td>
<td>2.80</td>
<td>0.40</td>
</tr>
<tr>
<td>Parietal cortex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Acetylaspartate/creatine</td>
<td>≤1.46</td>
<td>60.7</td>
<td>72.7</td>
<td>2.23</td>
<td>0.54</td>
</tr>
<tr>
<td>Choline/creatine</td>
<td>≤0.63</td>
<td>42.9</td>
<td>90.2</td>
<td>4.72</td>
<td>0.63</td>
</tr>
<tr>
<td>myo-Inositol/creatine</td>
<td>&gt;0.69</td>
<td>60.7</td>
<td>77.3</td>
<td>2.67</td>
<td>0.51</td>
</tr>
<tr>
<td>N-Acetylaspartate/myo-inositol</td>
<td>≤2.13</td>
<td>71.0</td>
<td>65.0</td>
<td>2.05</td>
<td>0.44</td>
</tr>
<tr>
<td>Occipital cortex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Acetylaspartate/creatine</td>
<td>≤1.61</td>
<td>100.0</td>
<td>75.0</td>
<td>4.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Choline/creatine</td>
<td>≤0.53</td>
<td>40.0</td>
<td>80.1</td>
<td>2.10</td>
<td>0.74</td>
</tr>
<tr>
<td>myo-Inositol/creatine</td>
<td>&gt;0.61</td>
<td>89.3</td>
<td>36.4</td>
<td>1.40</td>
<td>0.29</td>
</tr>
<tr>
<td>N-Acetylaspartate/myo-inositol</td>
<td>≤2.3</td>
<td>77.8</td>
<td>77.0</td>
<td>3.42</td>
<td>0.29</td>
</tr>
</tbody>
</table>

*a* Memory complaints, normal activities of daily living, normal general cognitive function, and abnormal memory for age but no dementia, with memory inefficiencies corroborated by an informant.

**FIGURE 3. Occipital Cortex Spectrum**

*In the patient who did not go on to develop probable Alzheimer’s disease, the peak amplitude is strikingly higher for N-acetylaspartate than for creatine. For the patient who over time did develop probable Alzheimer’s disease, there is a decrease of N-acetylaspartate peak amplitude in relation to creatine.*

**Discussion**

Memory impairment is a frequent complaint in the elderly, but hitherto we did not have any parameter objective enough to predict conversion to Alzheimer’s disease. Numerous studies have focused on this issue; some of them have used neuropsychological tests as predictors, others have used neuroimaging techniques. Although most studies are interesting, the matter has not been completely resolved because of either insufficient sensitivity or specificity of the tools used.

A cohort of 195 patients with questionable dementia underwent comprehensive neuropsychological examination to predict conversion to dementia, but these predictions achieved at most 64% sensitivity and 76% specificity for verbal recognition and verbal fluency (27). There have been some reports suggesting that hippocampal volume changes in patients with mild cognitive impairment may help predict conversion to Alzheimer’s disease (15, 28, 29). However, the subject of which region of the brain, entorhinal cortex or hippocampus, demonstrates the earliest abnormalities in early Alzheimer’s disease has not yet been elucidated because of the large variability of imaging measures, the intersubject variability of brain anatomy, and the imaging artifacts in the medial temporal region (30). In addition, some degree of hippocampal atrophy may be seen in up to one-third of cognitively normal old people (31). It is clear that we need more powerful objective pa-
rameters, such as magnetic resonance spectroscopy, on which predictions can rely.

Our results are encouraging, but it is surprising to us that hippocampal values did not predict conversion with confidence. It is not consistent with the early involvement of this area in Alzheimer's disease, but there may be some reasons to explain this finding. First, the partial volume effect in the hippocampus and parietal lobe resulting from the presence of CSF in some patients may affect the absolute value of metabolites but exerts little influence on metabolite ratios, since the CSF virtually does not contain the metabolites being assessed (32). Given the large size of the voxel, it is probable that extrahippocampal tissues had been included in the analysis, which may decrease the sensitivity of the technique in this area. Another possible explanation is that converters and nonconverters have equal or minimal neuropathological alterations in the hippocampus. A pathoanatomical study that included 13 nondemented cases, four cases with preclinical Alzheimer's disease, eight cases with very mild symptomatic Alzheimer's disease, and four cases with severe Alzheimer's disease found a significant decrease in neuron number in the entorhinal cortex and hippocampus of patients with very mild Alzheimer's disease but not in healthy and preclinical Alzheimer's disease cases (33). The authors concluded that Alzheimer's disease carries clinical deficits only when it produces significant neuronal loss and recognized that most subjects ascribed to the group of very mild Alzheimer's disease cases were rated as probable Alzheimer's disease according to Consortium to Establish a Registry for Alzheimer's Disease criteria. Theoretically, the lack of significant decrease in neuron number could explain the lack of predictive power of 1H-MRS hippocampal values in our study. Likewise, in a report of 15 patients with mild cognitive impairment who underwent neuropathological examination, all had more neuritic plaques in the hippocampus than did normal subjects, but the plaque density was less than one-third of that found in temporal and insular cortices (34).

On the one hand the useful values found in the occipital cortex look inconsistent at first glance with the hypothesis that the occipital cortex is only affected in the advanced stages of Alzheimer's disease (35). In addition, a cross-sectional study that included 21 patients with mild cognitive impairment, 21 patients with Alzheimer's disease, and 63 elderly comparison subjects showed similar N-acetylaspartate/creatine ratios in the occipital lobe (1.78, 1.8, and 1.82, respectively) (36). However, the authors stated that 14 out of the 21 Alzheimer's disease patients were being treated with donepezil, which may elevate the N-acetylaspartate/creatine ratios.

On the other hand, there have been three reports (37–39) showing low N-acetylaspartate/creatine ratios in the midline occipital cortex of Alzheimer's disease patients (1.14, 1.16, and 1.1, respectively). Furthermore, in one of these studies it was concluded that the occipital cortex myo-inositol/N-acetylaspartate ratio had the best discriminant power because it distinguished Alzheimer's disease from normality at 83% sensitivity and 98% specificity. The authors argued that this cortical area is rich in gray matter and, therefore, ideal for analysis in Alzheimer's disease (38).

Our results are also consistent with the delay of the P2 component of visual evoked potential in Alzheimer's disease patients but not in comparison subjects or patients with other dementing processes (40), which suggests cholinergic defect in visual association areas. The normal latencies of early visual evoked potential components are compatible with spared primary visual cortex. It was corroborated in a histological study showing low neurofibillary tangles in primary projection area 17 but increased 20-fold in adjacent association areas 18 and 20, although no differences were seen with regard to number of neuritic plaques (41).

The exploration of other cortical areas such as the cingulate gyrus also could be of aid in improving predictions. In a cross-sectional study that included 24 patients with mild cognitive impairment and 22 Alzheimer's disease patients, 1H-MRS of the posterior cingulate gyrus discriminated Alzheimer's disease from mild cognitive impairment at 80% specificity and 67% sensitivity by determining the N-acetylaspartate/creatine ratio (42).

In conclusion, 1H-MRS of the occipital cortex may be a valuable tool in predicting conversion from mild cognitive impairment to probable Alzheimer's disease and has proven superior to other predictors including neuropsychological performance. Given the correlational and descriptive nature of the present investigation, our findings should be replicated in a larger investigation. If our results are confirmed, this technique could be used for screening.
or prognostic purposes in patients with mild cognitive impairment. In addition, it could be helpful to determine which patients would benefit from early therapy for Alzheimer’s disease.


From the Neurology Unit, Hospital de Alcañiz, Spain; the Magnetic Resonance Unit, Clínica Quirón de Zaragoza, Spain; and the Neurology Unit, Hospital de Teruel, Spain. Address correspondence and reprint requests to Dr. Modrego, Avda Juan Carlos I, 21, 50009 Zaragoza, Spain; med009626@saludalia.com (e-mail).

This work was conducted with the financial support of Laboratorios Esteve, Spain. The design, writing, and the decision of publication are only of the authors, who have no conflicts of interest to declare.

References

24. González Torres M: Batería 144 para el estudio de las funciones mnésicas y de aprendizaje, in La exploración neuropsicológica. Edited by Peña Casanova J, Barcelona, Spain, MCR, 1988
32. Blüm S: Data acquisition in the research mode, in Clinical MR Spectroscopy: Europrobe Training Course. Edited by Ross BD. Pasadena, Calif, Huntington Medical Research Institutes, 1996, pp 60–84

Does Donepezil Treatment Slow the Progression of Hippocampal Atrophy in Patients With Alzheimer’s Disease?

Mamoru Hashimoto, M.D., Ph.D.
Hiroaki Kazui, M.D., Ph.D.
Keiji Matsumoto, M.D.
Yoko Nakano, M.D., Ph.D.
Minoru Yasuda, M.D., Ph.D.
Etsuro Mori, M.D., Ph.D.

Objective: The only approved pharmacological approach for the symptomatic treatment of Alzheimer’s disease in Japan is the use of a cholinesterase inhibitor, donepezil hydrochloride. Recent in vivo and in vitro studies raise the possibility that cholinesterase inhibitors can slow the progression of Alzheimer’s disease. The purpose of the present study was to determine whether donepezil has a neuroprotective effect in Alzheimer’s disease by using the rate of hippocampal atrophy as a surrogate marker of disease progression.

Method: In a prospective cohort study, 54 patients with Alzheimer’s disease who received donepezil treatment and 93 control patients with Alzheimer’s disease who never received anti-Alzheimer drugs underwent magnetic resonance imaging (MRI) twice at a 1-year interval. The annual rate of hippocampal atrophy of each subject was determined by using an MRI-based volumetric technique. Background characteristics, age, sex, disease duration, education, MRI interval, apolipoprotein E (APOE) genotype, and baseline Alzheimer’s Disease Assessment Scale score were comparable between the treated and control groups.

Results: The mean annual rate of hippocampal volume loss among the treated patients (mean=3.82%, SD=2.84%) was significantly smaller than that among the control patients (mean=5.04%, SD=2.54%). Upon analysis of covariance, where those confounding variables (age, sex, disease duration, education, MRI interval, APOE genotype, and baseline Alzheimer’s Disease Assessment Scale score) were entered into the model as covariates, the effect of donepezil treatment on hippocampal atrophy remained significant.

Conclusions: Donepezil treatment slows the progression of hippocampal atrophy, suggesting a neuroprotective effect of donepezil in Alzheimer’s disease.

At present, the only approved pharmacological approach for the symptomatic treatment of Alzheimer’s disease in Japan is the use of cholinesterase inhibitors. Donepezil hydrochloride has been demonstrated to have significant effects in slowing symptomatic progression in 24-week placebo-controlled trials (1), and some long-term studies have shown that there is no less benefit after 1 year of treatment (2, 3). One observational study (4) showed that cholinesterase inhibitor treatment alters the natural history of Alzheimer’s disease, as indicated by the delay in admission to nursing homes. Furthermore, a longitudinal neuroimaging study using single photon emission computed tomography (SPECT) demonstrated that treatment of patients with Alzheimer’s disease with donepezil for 1 year reduced the decline in regional cerebral blood flow (rCBF) (5), suggesting the preservation of functional brain activity in donepezil-treated patients. Although these studies appear to have demonstrated the efficacy of donepezil in slowing down clinical disease progression, it is not clear whether donepezil treatment slows disease progression in Alzheimer’s disease.

Neuropathologically, Alzheimer’s disease is characterized by the presence of neurofibrillary tangles and senile plaques, impaired synaptic function, and cell loss (6). Although these histological features cannot be examined noninvasively, the cell loss that accompanies them can be seen in vivo as atrophy with magnetic resonance imaging (MRI). Among the characteristic neuropathological changes in Alzheimer’s disease, the most prominent structural changes at the initial stage occur in the hippocampal formation (7, 8). MRI-based volumetry of the medial temporal lobe structures has been proposed as a useful tool for the clinical diagnosis of Alzheimer’s disease (8). Serial MRI studies permit calculation of rates of atrophy over time. It has been proposed that measurement of the rate of atrophy could be used to monitor the effectiveness of antide-mentia drugs for Alzheimer’s disease (9, 10). If an anti-Alzheimer drug can slow down the anatomic progression of Alzheimer’s disease pathology, this should be detectable as a decrease in the rate of hippocampal atrophy in treated patients.
The purpose of the present study was to determine whether a cholinesterase inhibitor, donepezil hydrochloride, has a neuroprotective effect on Alzheimer’s disease. Using an MRI-based volumetric technique, we examined the rates of hippocampal atrophy in donepezil-treated Alzheimer’s disease patients and compared the results with those in control patients.

Method

Subjects and Study Design

In the present study, a prospective cohort was compared with a historical control cohort. All procedures followed the clinical study guidelines of the ethics committee of Hyogo Institute for Aging Brain and Cognitive Disorders, a research-oriented hospital, and were approved by the institutional review board. Written informed consent was obtained from the patients or their families.

The control group was selected among those who participated in the annual follow-up program before donepezil was introduced. Patients with Alzheimer’s disease who were examined at the Hyogo Institute for Aging Brain and Cognitive Disorders were invited to the Hyogo Institute’s Alzheimer’s disease annual follow-up program starting in July 1993. The consent and availability of a reliable caregiver and the absence of severe behavioral problems impelling institutionalization were the requirements for participating in the program. At baseline, the patients were examined comprehensively by both neurologists and psychiatrists under a short-term admission to the infirmary and were given routine laboratory tests, EEGs, and standardized neuropsychological examinations. The patients’ medical history was systematically collected from reliable family members by using a database format. Other imaging studies, such as MRIs of the brain, magnetic resonance angiograms of the head and neck, and cerebral perfusion SPECT were performed at baseline. Neuropsychological tests and MRIs were repeated at 1-year intervals. The clinical and investigational data collected prospectively in a standardized fashion were all added to the database (11).

After donepezil hydrochloride was licensed and marketed in November 1999 in Japan, the annual follow-up program was revised to the donepezil follow-up program because most patients with Alzheimer’s disease were treated with donepezil. The treated group consisted of consecutive patients with mild to moderate Alzheimer’s disease who received donepezil treatment and were enrolled in a prospective longitudinal cohort study between November 1999 and June 2003. Treatment with donepezil was a prerequisite for participating in the revised program. After the same baseline assessments as the annual follow-up program were evaluated, the patients received 3 mg/day of donepezil for 1 or 2 weeks and then 5 mg/day (the approved maximum dose in Japan). Donepezil hydrochloride was prescribed for the entire year after the initial MRI, and compliance was monitored at every visit (every 3 months). Neuropsychological tests and MRIs were repeated at 1-year intervals. The treated patients satisfied six inclusion criteria:

1. Meeting criteria of the National Institute of Neurological Disease and Stroke/Alzheimer’s Disease and Related Disorders Association for probable Alzheimer’s disease (12)
2. Having minimal to moderate functional severity (a score of 15 or more on the Mini-Mental State Examination (MMSE) (13)
3. Having no lifetime history of other neurological or mental disorders
4. Taking no established and documented antidementia drugs other than donepezil (agents with possible antidementia properties, such as nonsteroidal antiinflammatory drugs, vitamins E, ginkgo biloba, and lecithin were allowed)
5. Having no evidence of focal brain lesions on a MRI
6. There being informed consent from patients and their relatives for determination of their apolipoprotein E (APOE) genotype

The control group was selected among those who participated in the annual follow-up program between July 1993 and October 1999. Requirements for inclusion in the present study were the same as those for the donepezil follow-up participants except for the use of donepezil. Those who received anti-Alzheimer drugs other than donepezil were not included in the present study. In this study, only baseline and 1-year follow-up data were used.

Cognitive Functions

The status of global cognitive function was assessed with the MMSE and the Alzheimer’s Disease Assessment Scale (14). The Alzheimer’s Disease Assessment Scale was administrated by psychologists (along with the MRI examination) who were not involved in managing the patients at a 1-year interval (10–14 months).

MRI Volumetry

We directly measured the volume of the hippocampal formation on high spatial resolution three-dimensional spoiled-gradient echo images (15). The images were generated perpendicular to the anterior-posterior commissure plane that covers the whole calvaria with a 1.5-T MRI unit (Sign Advantage 5.x, General Electric Medical Systems, Milwaukee). The operating parameters were as follows: field of view=220 mm, matrix=256×256, 12 k×1.5 mm contiguous sections, TR=14 msec, TE=3 msec, and flip angle=20°.

The detailed hippocampal MRI volumetric procedure is described elsewhere (15). In brief, the MRI data set was transmitted to a computer from the MRI unit, and after an appropriate data conversion, it was analyzed by using the public domain Image version 1.62 program (developed at the National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/nih-image/) with residential macro programs developed in our institute. Hippocampal formation volume in all subjects was measured by a single investigator (M.H.) who was blinded to 1) all clinical information, 2) the order (initial and follow-up) of MRIs, and 3) the time of enrollment. Before starting the measurement, the rater inspected the suitability of MRIs for hippocampal volumetry. If either one of the two MRIs obtained for a subject was unsuitable for volumetry because of motion artifacts or for any other technical reason, the subject was excluded from the study.

For volumetry, we used a combination of semiautomatic segmentation-through-density thresholding and manual tracing, thereby lowering partial voluming and the observer’s bias. The reliability and validity of volumetry have been established and described elsewhere (9, 15). The hippocampal formation was defined to include the pes hippocampi, digitations hippocampi, alveus hippocampi, dentate gyrus, and subiculum (16–18). The boundary of the hippocampal formation was defined from the entire head of the hippocampus to the slice where the crus of the fornix was seen in full profile. The outline of the hippocampal formation together with the surrounding white matter and CSF was first traced with a manually guided mouse cursor, and subsequently, the gray matter of each structure was extracted by density thresholding set at a range between minimum and maximum pixel values. The maximum value was defined as the largest value for any pixel of the gray matter (represented by the caudate head). The minimum value was defined as one-half the sum of the mean pixel value of the gray matter and the mean value of CSF (represented by the lateral ventricle) (16). The slice volume of the hippocampal formation was obtained by automatically counting the number of pixels within the segmented regions and then multi-
TABLE 1. Demographic and Clinical Characteristics of Patients With Alzheimer’s Disease

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Donepezil-Treated Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One or Two Apolipoprotein E (APOE) ε4 Allele(s) Present (N=54)</td>
<td>One or Two Apolipoprotein E (APOE) ε4 Allele Absent (N=43)</td>
</tr>
<tr>
<td></td>
<td>One or Two Apolipoprotein E (APOE) ε4 Allele Absent (N=19)</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Age (years)</td>
<td>69.3</td>
<td>9.0</td>
</tr>
<tr>
<td>Education (years)</td>
<td>9.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Duration of disease (months)</td>
<td>31.8</td>
<td>19.3</td>
</tr>
<tr>
<td>Magnetic resonance imaging interval (days)</td>
<td>389</td>
<td>25</td>
</tr>
<tr>
<td>Baseline score on Mini-Mental State Examination</td>
<td>21.9</td>
<td>3.8</td>
</tr>
</tbody>
</table>

donepezil and hippocampal atrophy

pling the number by the voxel size: 9(220/256)2 × (1.5=1.1078 mm3)). The average volume of the right and left hippocampal formations was used for the analysis. The test-retest reliability (kappa=0.98) was measured as an intraclass correlation coefficient derived from three repeated measurements by a single rater under blinded conditions in 10 randomly selected subjects. Coefficients of variation (standard deviation/mean, where standard deviation indicates square root value of the arithmetic mean of 10 variance estimates) were 2.84% and 2.78% for the right and left hippocampal formations, respectively.

**Determination of APOE Genotype**

The detailed method for APOE genotyping is described elsewhere (19). In brief, genomic DNA was extracted from peripheral blood with a Genomix DNA extraction kit (Talent Corp., Trieste, Italy) according to the manufacturer’s protocol. The APOE genotype was determined by using polymerase chain reaction restriction fragment length polymorphism, according to the procedure described by Wenham and colleagues (20).

**Statistical Analysis**

The annual rate of hippocampal atrophy was defined as the percentage change, which was computed as baseline hippocampal formation volume minus 1-year hippocampal formation volume divided by baseline hippocampal formation volume (×100). The change in cognitive decline was expressed as the difference (points) between the baseline and 1-year Alzheimer’s Disease Assessment Scale scores. Because the APOE ε4 allele is known to be specifically related to hippocampal atrophy in Alzheimer’s disease (9, 21), the effect of donepezil on hippocampal atrophy is possibly influenced by the APOE genotype. Therefore, we used two-way analysis of variance, which included the effects of donepezil treatment (i.e., treated and untreated), APOE type (i.e., presence and absence of the APOE ε4 allele), and their interaction. When the interaction term was not significant, the effects of the donepezil treatment were examined with a t test. When significant effects were found, they were further examined with analysis of covariance (ANCOVA) in which age, sex, disease duration, education, interval between the two MRI studies (days), APOE genotype, and baseline MMSE score were entered into the model as covariates. The level of significance was set at p<0.05 for all statistical analyses.

**Results**

Twelve of 77 patients who were initially enrolled in the donepezil follow-up program were dropped from the study. Seven withdrew their consent; two were institutionalized, and the remaining three withdrew for other reasons. Of the 65 patients who completed the entire procedure of the study, 11 were excluded because the quality of one of their two MRIs was unsuitable for volumetry. Thus, 54 patients remained in the treated group. Two patients did not receive the full, 1-year dosage of donepezil. One patient discontinued donepezil after 3 weeks because of adverse effects, and the other took the drug irregularly (receiving about 60% of the full dose). However, these patients were included in the primary analyses.

Of the 108 patients who completed the entire procedure of the annual follow-up study, 15 patients were excluded because their MRIs were of poor quality. Thus, 93 patients remained in the control group. The background characteristics of both groups are summarized in Table 1. The treated and control groups were not significantly different in age, sex, education, disease duration, intervals between the first and second MRI, or baseline MMSE scores.

Hippocampal volumes and Alzheimer’s Disease Assessment Scale scores are summarized in Table 2. The effect of donepezil treatment on the change in Alzheimer’s Disease Assessment Scale scores was significant (F=5.55, df=143, p<0.02), but neither the effect of the APOE ε4 allele (F=3.64, df=143, p<0.06) nor the interaction term (F=0.32, df=143, p=0.57) was significant. The effects of donepezil treatment and the APOE ε4 allele on the rate of hippocampal atrophy were significant (F=8.19, df=143, p=0.005, and F=5.29, df=143, p<0.03, respectively), but the interaction term was not significant (F=0.08, df=143, p=0.78). In a further analysis, the mean annual rate of hippocampal atrophy in the treated group (mean=3.82%, SD=2.84%) was significantly lower than in the control group (mean=5.04%, SD=2.54%) (t=-2.7, df=145, p=0.008). In a one-way ANCOVA, where
TABLE 2. Annual Change in Alzheimer’s Disease Assessment Scale Score and Rate of Hippocampal Atrophy Among Patients With Alzheimer’s Disease

<table>
<thead>
<tr>
<th>Variable</th>
<th>One or Two APOE ε4 Allele(s) Present (N=35)</th>
<th>One or Two APOE ε4 Allele Absent (N=19)</th>
<th>Total (N=54)</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Alzheimer’s Disease Assessment Scale score</td>
<td>14.0 ± 4.8</td>
<td>16.3 ± 5.6</td>
<td>14.8 ± 5.2</td>
<td>15.9 ± 4.9</td>
</tr>
<tr>
<td>Annual change in Alzheimer’s Disease Assessment Scale</td>
<td>-0.2 ± 3.8</td>
<td>2.5 ± 4.9</td>
<td>0.8 ± 4.4</td>
<td>3.0 ± 6.4</td>
</tr>
<tr>
<td>Baseline hippocampal formation volume (mm³)</td>
<td>2496 ± 428</td>
<td>2514 ± 496</td>
<td>2502 ± 449</td>
<td>2391 ± 461</td>
</tr>
<tr>
<td>Annual hippocampal atrophy rate (%)ab</td>
<td>4.15 ± 2.32</td>
<td>3.21 ± 3.60</td>
<td>3.82 ± 2.84</td>
<td>5.59 ± 2.15</td>
</tr>
</tbody>
</table>

a Significant effect of treatment.

b Significant effect of APOE ε4 allele.

Discussion

In the present study, the mean annual decline in the Alzheimer’s Disease Assessment Scale score in the control group (3.6 points) was significantly larger than in the treated group (0.8 points). Thus, these results are consistent with the results of previous long-term studies of donepezil (2, 3), and suggest that treatment with donepezil for 1 year was associated with a cognitive benefit. Moreover, donepezil seemingly slows the rate of hippocampal atrophy. The decrease in hippocampal volume in the treated patients (3.82%) was about 24% less than in the control patients (5.04%). The present results suggest that donepezil has both symptomatic and disease-progression slowing effects. The mean annual rate of hippocampal atrophy was significantly higher in the patients with the APOE ε4 allele than in those without the APOE ε4 allele, replicating our previous observation that the APOE ε4 allele is specifically related to accelerated hippocampal pathology in Alzheimer’s disease (9). On the other hand, for the rate of hippocampal atrophy and for the change in Alzheimer’s Disease Assessment Scale, no significant interaction was noted between donepezil treatment and the APOE genotype.

In recent studies, long-term donepezil treatment has been demonstrated to slow the decline in cognition and functional activities (2, 3), delay the admission to a nursing home (4), and reduce the decline in rCBF (5). However, these studies do not answer the question of whether the beneficial effects of donepezil are due to symptomatic suppression, which leaves the underlying disease process unaltered, or to neuroprotection modifying the disease process. Recently, one randomized, double-blind, placebo-controlled pilot study in patients with Alzheimer’s disease demonstrated that donepezil-treated patients had significantly smaller mean decreases in total hippocampal volumes compared with placebo-treated patients. Although that study had limitations in that it included a relatively small number of patients and the period of drug treatment was short, our data are consistent with the results (23). Furthermore, two clinical trials, both open-label, placebo-controlled extension studies, have raised the possibility that cholinesterase inhibitors have slowing effects on both symptomatic and disease progression in Alzheimer’s disease patients. In one study (24), the level of cognitive function in patients who were treated with placebo for 3 months, followed by 12 months of donepezil treatment, was lower than in patients who were treated with donepezil for all 15 months. In the other study (25), the level of cognitive function of the patients who were treated with placebo for 26 weeks, followed by 26 weeks of rivastigmine treatment, was lower than that in patients who were treated with rivastigmine for all 52 weeks. The inability to catch up in those in whom the treatment with
cholinesterase inhibitors was postponed reminds us of a potential disease-modifying effect.

A recent long-term randomized, double-blind trial (26) showed that no significant benefits were seen with donepezil compared with placebo in the institutionalization and progression of disability. In the study, there were no significant differences in behavioral symptoms, caregiver well-being, and caregiver time costs. On the other hand, statistically significant cognitive and functional effects were maintained over at least 2 years. Because many of the outcomes are influenced by the interaction of complex biological, social, and environmental factors, they might be inappropriate for assessing the neuroprotective effects of donepezil.

To explain the neuroprotective effect of cholinesterase inhibitors, mechanisms based on beta-amyloid metabolism have been postulated. Accumulation of amyloid is one of the earliest changes in Alzheimer’s disease pathology (27, 28) and may cause neuronal death in the CNS (29, 30). In vitro studies have demonstrated a link between cholinergic activation and beta-amyloid precursor protein metabolism. Wallace et al. (31) found evidence that lesions of the cholinergic nucleus basalis of Meynert increased the synthesis of beta-amyloid precursor protein in the cerebral cortex of rats. Wolf et al. (32), using human CNS neurons, found increased amyloid precursor protein secretion and decreased beta-amyloid protein production with carbachol stimulation of muscarinic receptors (32). These studies support a beneficial alteration in amyloid processing associated with cholinergic stimulation. Kihara et al. (33) examined the effects of nicotinic receptor agonists on amyloid beta cytotoxicity in cultured rat cortical neurons and found that nicotinic receptor stimulation may be able to protect neurons from degeneration induced by amyloid beta. Svensson and Nordberg (34) demonstrated that tacrine and donepezil at clinically relevant concentrations attenuated amyloid beta (25-35)-induced toxicity in rat pheochromocytoma PC12 cells. The neuroprotective effect was blocked by the presence of the nicotinic antagonists mecamylamine and tubocurarine, suggesting an intervention through nicotinic receptors.

Our study has some limitations. First, because it was a comparative study with a historical control group rather than a randomized study, it suffers from several sources of bias. The groups are not comparable because of the selection of subjects who received the intervention (selection bias), the cointerventions and other medical management being received by the two groups were different (performance bias), and the methods of outcome measurement being used in the two groups were different (detection bias). Although there was no intention to select subjects for the control and treatment groups, patients who could not tolerate the initial doses of donepezil were not included in the donepezil follow-up program. This could have been a source of selection bias. However, we are unaware of any evidence that donepezil tolerance predicts disease progression. Because the control group in the present study was not concurrent but historical, there was a generation difference of several years between the groups. Although the generation difference was a possible source of selection and performance bias, the difference in the patients’ generation was not likely to affect the volumetric results, and the cointerventions and other medical management did not change throughout the first and second halves of the study period. Second, it has been recognized that open-label studies are not optimal. The fact that the investigators were not blinded to treatment might increase the donepezil effect. However, neither hippocampal volume nor the Alzheimer’s Disease Assessment Scale score was a subjective outcome measure. Furthermore, in the present study, the Alzheimer’s Disease Assessment Scale was administered by neuropsychologists who were unblinded but not involved in managing the patients, and MRI volumetry was made by an investigator who was blinded to all clinical information. Therefore, it was unlikely that detection bias affected the results of the rate of hippocampal atrophy or the Alzheimer’s Disease Assessment Scale score in favor of the donepezil-treated group. Third, dropouts are a problem for this type of design and a possible source of exclusion bias. Simply ignoring everyone who has withdrawn from a clinical trial may bias the results, usually in favor of the intervention. Therefore, it is standard practice to analyze the results of comparative studies on an intention-to-treat basis. In the present study, two patients who did not take the full 1-year dosage of donepezil because of adverse events or poor compliance were included in the primary analyses.

As a result, the patients’ background characteristics, such as age, sex, education, disease duration, disease severity, and MRI interval, were comparable between the two groups, and even after we controlled for these variables, the effect of donepezil on hippocampal atrophy remained unchanged. In any case, the significant difference of the rate of hippocampal atrophy was likely to be due to the intervention with donepezil. Although our findings should be confirmed by a randomized, controlled long-term trial in patients with Alzheimer’s disease, it would be unethical to conduct such a study to extend the knowledge of donepezil.

The present results suggest that donepezil has not only a symptomatic effect but also a neuroprotective effect. If donepezil does, in fact, influence disease progression, we need to modify our treatment strategies; donepezil is not an optional but rather a mandatory treatment for Alzheimer’s disease and should be started in the prodromal or very early stage of the disease. Mild cognitive impairment is a condition that frequently progresses to Alzheimer’s disease, which requires early diagnostic and therapeutic interventions. Donepezil, through its neuroprotective effect, could possibly inhibit progression from mild cognitive impairment to Alzheimer’s disease. Further studies are needed to determine whether donepezil slows the progres-
sion from mild cognitive impairment to Alzheimer's disease. Hippocampal atrophy could be used as a surrogate marker of disease progression in such studies. Furthermore, the potential neuroprotective mechanism should be refined and exploited to enhance the drug's effectiveness in treating Alzheimer's disease. A better understanding of this mechanism may suggest strategies for designing improved drugs.

Received June 11, 2004; revision received Sept. 1, 2004; accepted Sept. 10, 2004. From the Hyogo Institute for Aging Brain and Cognitive Disorders, Hyogo, Japan; Sawa Hospital at Toyonaka; the Department of Psychiatry and Behavior Science, Osaka University Graduate School of Medicine, Osaka, Japan; the Department of Psychiatry and Neurology, Kobe Graduate University School of Medicine, Hyogo, Japan; and the Department of Behavioral Neurology and Cognitive Neuroscience, Tohoku University Graduate School of Medicine, Sendai, Japan. Address correspondence and reprint requests to Dr. Hashimoto, Sawa Hospital, 1-9-1 Shiroyamacho, Toyonaka, Osaka, 561-8691, Japan; qq257xk9@minos.ocn.ne.jp (e-mail).

References


Effect of Raloxifene on Prevention of Dementia and Cognitive Impairment in Older Women: The Multiple Outcomes of Raloxifene Evaluation (MORE) Randomized Trial

Kristine Yaffe, M.D.
Kathryn Krueger, M.D.
Steven R. Cummings, M.D.
Terri Blackwell, M.A.
Victor W. Henderson, M.D., M.S.
Somnath Sarkar, Ph.D.
Kristine Ensrud, M.D., M.P.H.
Deborah Grady, M.D., M.P.H.

Objective: This investigation examined whether raloxifene, a selective estrogen receptor modulator, affects the risk for Alzheimer’s disease.

Method: The Multiple Outcomes of Raloxifene Evaluation was a randomized, placebo-controlled trial among postmenopausal women with osteoporosis. The effect of raloxifene (60 or 120 mg/day) on vertebral fractures was the primary outcome. Development of mild cognitive impairment and dementia was a secondary outcome. Women were given clinical and cognitive evaluations at baseline and annually. After 3 years, among the 5,386 women enrolled at participating sites, those who had clinical symptoms of dementia or scored in the lowest 10th percentile on cognitive screening were evaluated by a blinded dementia specialist and had brain scans and laboratory tests to evaluate dementia etiology. Dementia was diagnosed by a blinded adjudication committee.

Results: Of the 5,386 women, 5,153 (95.7%) were classified as cognitively normal, 181 (3.4%) had mild cognitive impairment, and 52 (1.0%) had dementia, 36 with Alzheimer’s disease. Compared to those taking placebo, women receiving 120 mg/day of raloxifene had a 33% lower risk of mild cognitive impairment (relative risk, 0.67; 95% confidence interval [CI], 0.46–0.98) and somewhat lower risks of Alzheimer’s disease (relative risk=0.52, 95% CI=0.22–1.21) and any cognitive impairment (relative risk=0.73, 95% CI=0.53–1.01). Risks of mild cognitive impairment, Alzheimer’s disease, and any impairment were not significantly different in the group taking 60 mg/day of raloxifene.

Conclusions: Raloxifene at a dose of 120 mg/day, but not 60 mg/day, resulted in reduced risk of cognitive impairment in postmenopausal women.

Over 33% of women and 20% of men ages 65 and older will develop dementia during their lifetime, and many more will develop a milder form of cognitive impairment (1). With the aging of the population in many Western countries, the incidence and prevalence of Alzheimer’s disease and other forms of dementia are expected to quadruple over the next 50 years, highlighting the importance of preventive interventions (2). To our knowledge, no effective preventive therapies have been identified from randomized trials thus far. There is also growing interest in preventing milder cognitive deficits that are not severe enough to meet criteria for the diagnosis of dementia, as patients with this disorder frequently progress to dementia, especially Alzheimer’s disease (3, 4).

The decline in serum estrogen levels following menopause has been hypothesized to increase the risk of developing cognitive impairment and Alzheimer’s disease in older women. Estrogen has beneficial biological effects in animal and cell culture models that support this hypothesis; these effects include stimulation of choline acetyltransferase activity in ovariectomized rats, neurotrophic effects on cultured neurons, and inhibition of amyloid formation (5). The results of meta-analyses of the findings from observational studies indicated that postmenopausal estrogen therapy reduces the risk of developing dementia by approximately 30% (6, 7). However, in recent large randomized trials, not only did combined conjugated estrogen and progestin therapy (8) or unopposed estrogen (9) fail to prevent all-cause dementia but there was an increase in dementia associated with hormone assignment.

Raloxifene is a selective estrogen receptor modulator used for prevention and treatment of osteoporosis (10, 11). Raloxifene has shown estrogen agonist effects on the CNS in animal (12) and in vitro (13) studies but anti-estrogen effects as well (14). We evaluated the effect of raloxifene treatment among older women with osteoporosis to determine if raloxifene alters the risk of developing Alzheimer’s disease and cognitive impairment.
RALOXIFENE AND DEMENTIA

**FIGURE 1. Enrollment and Assessment of Women in the Dementia Study of the Multiple Outcomes of Raloxifene Evaluation Trial**

Randomly assigned subjects
(N=7,705)

- Women assigned to sites participating in dementia study (N=7,023)
- Women assigned to sites not participating in dementia study (N=682)

- Women screened for dementia (N=5,386)
- Women who left study at or before dementia screening (N=1,637)
- Women who failed dementia screening but were not evaluated further (N=102)
- Women who passed dementia screening (considered cognitively normal) (N=4,540)

Dementia clinical evaluation, phase 1 (N=744)
(assessed presence or absence of dementia)
1. Interviews with patient and caregiver
2. Medical history, physical and neurological examinations
3. Mini-Mental State Examination
4. Geriatric Depression Scale
5. Clinical Dementia Rating scale
6. Hachinski Ischemic Score

- Women who did not finish evaluation (N=2)

Dementia clinical evaluation, phase 2 (N=742)
(assessed type of dementia)
1. Brain CT or MRI scan
2. Dementia laboratory tests

Dementia adjudication committee (N=742)
(final cognitive status assigned after review of results from phase 1 and phase 2 assessments and tests)
- Cognitively normal (N=509)
- Mild cognitive impairment (N=181)
- Dementia (N=52)

**Method**

**Study Subjects**

The Multiple Outcomes of Raloxifene Evaluation (MORE) trial enrolled 7,705 postmenopausal women with osteoporosis. The primary outcomes of the trial were vertebral fractures and bone mineral density. The participants were randomly assigned by site to receive 60 mg/day of raloxifene, 120 mg/day of raloxifene, or identical placebo. All women were also asked to take calcium (500 mg) and vitamin D (400 to 600 IU) daily. Details of the study design and the main results already have been reported (10, 15). The human studies review board at each site approved the protocol; after complete description of the study to the subjects, written informed consent was obtained.

The effect of raloxifene on the occurrence of mild cognitive impairment and dementia was a planned secondary objective of the trial. Of the 180 clinical sites in 25 countries, 19 sites (682 women) were unable or unwilling to participate in the dementia evaluation (Figure 1). The analysis includes the 5,386 women who were still participating in the MORE study at the time of dementia screening (the end of the third year of treatment).

**Measurements**

At baseline we collected information on age, ethnicity, education, smoking, alcohol use, health conditions, prior postmenopausal estrogen use (whether the woman had ever used estrogen, up to 2 months before enrollment), and reproductive history. Height and weight were measured, and the body mass index was calculated. The participants completed the 15-item Geriatric Depression Scale (16), which assesses depressive symptoms over the past week. Total serum estradiol was measured at baseline, after at least a 6-hour fast, in serum shipped the same day to the laboratory (SciCor [now Covance], Central Laboratory Services, Indianapolis). Estradiol concentration was determined by using a double antibody procedure. The intra-assay coefficient of variation is 6.5% (with a standard deviation of 2.1 pmol/liter) at an estradiol concentration of 33 pmol/liter. The detectable limit of the assay used is ≤5 pmol/liter. For our analyses, we dichotomized total estradiol as detectable or nondetectable.

**Cognitive Impairment and Dementia**

As defined at baseline, the primary outcomes of the dementia study were the risks of developing Alzheimer’s disease, mild cognitive impairment, and dementia of any type. The secondary outcomes were the risk of developing vascular dementia and the risk of any cognitive impairment (mild cognitive impairment and dementia combined).

**Screening.** Six cognitive tests, similar to those used by the Consortium to Establish a Registry for Alzheimer’s Disease (17), were administered at baseline and annually in a standard order by trained study personnel who were blinded to treatment assignment. The Short Blessed Test (18) assesses orientation, concentration, and memory. Scores ranged from 0 to 28, with lower scores indicating better performance, and a score of 8 is consistent with probable cognitive impairment. Trail Making Tests A and B measure visuospatial scanning, sequential processing, motor speed, executive function, and attention (19). The Word List Memory and Recall Tests measure learning, immediate memory, and delayed memory (17). The memory test requires immediate recall of 10 standardized words. The Word List Fluency Test measures verbal production, semantic memory, and language by having subjects name as many animals as possible in 60 seconds (17).

We used a staged evaluation similar to that used in clinical practice to diagnose cognitive impairment and dementia. We used the score on the Short Blessed Test to screen for possible cognitive impairment or dementia. At year 3, participants with the worst 10% of scores in each country or having clinical symptoms of cognitive impairment as judged by the investigator were referred for further evaluation. At two sites, the Short Blessed Test was not administered (N=188). At these two sites, women scoring in the worst 10% on the Buschke Selective Reminding Test (a test of verbal memory) (20) and those judged impaired by the site investigator were referred for further evaluation. Women who passed dementia screening at year 3 (N=4,540) were considered cognitively normal and were not evaluated further.

**Diagnosis**

The women who were suspected of having cognitive impairment or dementia were evaluated further by clinical experts to
determine the presence of dementia or cognitive impairment and, among those with dementia, to assess the type of dementia (Figure 1). A blinded clinician who was expert in dementia diagnosis (geriatrician, psychiatrist, or neurologist) 1) interviewed the participant and caregiver for the participant's history of cognitive deficits, functional abilities, and potential precipitating factors, 2) obtained a list of current and recent medications, 3) performed a physical and neurological examination, and 4) administered the Mini-Mental State Examination (MMSE) (21), 30-item Geriatric Depression Scale (22), Clinical Dementia Rating scale (23), and Hachinski Ischemic Score (24). The Clinical Dementia Rating is a semistructured interview of both participant and caregiver that evaluates functional and cognitive status. A summary score (total box score) is calculated from the six individual category scores, and a standardized algorithm is used to assign an overall global Clinical Dementia Rating score. A global score of 1 or higher suggests the diagnosis of dementia, a score of 0.5 suggests mild cognitive impairment, and a score of 0 is consistent with a normal cognitive status (23).

Each participant with evidence of dementia based on the clinical impression of the consulting clinician or with an MMSE score less than 24 was given a brain scan with computerized tomography or magnetic resonance imaging and laboratory tests (fluorescent treponemal antibody, vitamin B12, serum folate, and thyroid-stimulating hormone). All brain scans were read locally for safety purposes but were subsequently read by a neuroradiologist at the University of California, San Francisco. This central reader, blind to treatment groups, evaluated the scans for abnormal findings and determined if these were clinically relevant.

All results of the medical history, physical examination, functional and mental status assessments, cognitive functioning tests, and brain imaging tests were recorded on standardized forms. These forms, along with laboratory test results, were presented to two members of the dementia adjudication committee. This committee comprised four physicians (including V.W.H.) who are recognized experts in the clinical and research diagnosis of cognitive impairment and dementia. The committee members were masked to treatment assignment. Using criteria established at baseline, two committee members independently judged cognitive status as normal cognitive functioning, mild cognitive impairment (3, 25), Alzheimer's disease (based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association [26]), vascular dementia (based on the criteria of the State of California Alzheimer's Disease Diagnostic and Treatment Centers [27]), or other type of dementia. If the committee confirmed dementia but could not assign a type, the diagnosis of "dementia type indeterminate" was assigned.

**Statistical Analyses**

We compared baseline characteristics by treatment group, using chi-square tests for categorical data, analysis of variance (ANOVA) for normally distributed continuous variables, and Kruskal-Wallis tests for skewed continuous variables. Baseline characteristics were analyzed for the 5,386 women who participated in the dementia ancillary study. Among the 742 women who completed phase 1 and phase 2, we also compared baseline characteristics by final diagnosis.

We determined the incidence of each cognitive diagnosis by treatment group and calculated the relative risk and 95% confidence intervals (CIs) by using log-linear regression models for comparisons of the 60- and 120-mg/day doses of raloxifene with placebo (intention-to-treat analyses). All women who passed the dementia screening (N=4,540) were considered to be cognitively normal, and we adjusted the models by country to control for variation in cognitive test scores. The 102 women who should have been clinically evaluated but did not participate (Figure 1) and the two women who did not complete the evaluation were coded as missing. We repeated the analyses after coding them as having any cognitive impairment to determine if our results were sensitive to this categorization. We anticipated having only moderate power to detect differences in dementia outcomes by treatment group, given the low incidence rates of dementia and mild cognitive impairment.

We examined the possibility of pooling the data from the two raloxifene dose groups by performing a chi-square test on the incidence rates of the cognitive outcomes. Since these rates were
statistically different for some diagnoses, we analyzed the data separately by dose of raloxifene.

To examine the effect of baseline serum total estradiol level on the association between raloxifene and cognitive outcome (a post hoc analysis), we stratified the women by whether or not serum estradiol was detectable at baseline. We also tested for the possibility of an interaction between treatment group and estradiol level for all cognitive outcomes.

### Results

There were no significant differences between treatment groups in baseline characteristics (Table 1) or baseline cognitive test scores (Table 2). The mean age of the participants was 66.3 years with a range of 35.7 to 80.9 years. Of the 5,386 participants, 744 women had suspected dementia at screening and were evaluated further (Figure 1). An additional 102 women were suspected of having dementia but were not evaluated further. The 102 women suspected of dementia who did not undergo further evaluation were equally distributed among the three treatment groups (p = 0.64, chi-square analysis). The 744 women who were suspected of having dementia were generally older (p < 0.001, ANOVA), further past menopause, less educated, and more depressed at baseline than were the women not referred for further clinical evaluation (p < 0.04, Kruskal-Wallis test), and they were also more likely to have had a hysterectomy and less likely to drink alcohol at baseline (p < 0.04, chi-square analysis). The mean time between randomization and the completion of the clinical dementia evaluation was 3.8 years and did not differ by treatment group.

We determined if the 5,386 women included in our analyses were similar to the entire 7,705 women who began the MORE trial and found that on almost all of the baseline characteristics, there were no statistically significant differences between these groups. In addition, similar rates of women in the three treatment groups were referred for the clinical dementia evaluation, and there were no statistically significant differences in baseline characteristics between the treatment and placebo groups. Among the 5,386 women who were part of the dementia evaluation study, the rate of compliance (taking 70% or more of the study drug between visits) was 94.2% for placebo, 94.6% for 60 mg/day of raloxifene, and 94.6% for 120 mg/day of raloxifene. These rates were not significantly different (p = 0.54, chi-square analysis).

Of the 744 women referred for further evaluation of dementia, 742 completed the evaluation. Of these, 509 (68.6%) were found to have normal cognitive functioning, 181 (24.4%) had mild cognitive impairment, and 52 (7.0%) met the criteria for dementia (36 with Alzheimer’s disease, one with vascular dementia, and 15 with indeterminate or other type of dementia). As expected, women with mild cognitive impairment or dementia were progressively older, were further past menopause, were less educated, had higher depression scores, and were less likely to use alcohol than those with normal cognitive functioning (Table 3). The percentage of women with a history of estrogen use was nonsignificantly lower for women who developed dementia than for those who developed mild cognitive impairment or who were cognitively normal. Mean scores on the MMSE and Clinical Dementia Rating obtained at the dementia evaluation were worse in women diagnosed with dementia than in those with mild cognitive impairment or normal cognitive functioning, as expected (Table 3).

There was no significant difference between the placebo group and the women receiving 60 mg/day of raloxifene in the risk of developing mild cognitive impairment, Alzheimer’s disease, dementia from any cause, or any cognitive impairment (dementia and mild cognitive impairment combined) (Table 4). Compared with the placebo group, women receiving 120 mg/day of raloxifene had a significantly lower risk of developing mild cognitive impairment (33% lower), a nonsignificantly lower risk of de-

### Table 2. Baseline Cognitive Performance of the 5,386 Women Screened in the Dementia Study of the Multiple Outcomes of Raloxifene Evaluation

<table>
<thead>
<tr>
<th>Baseline Cognitive Test</th>
<th>Placebo (N=1,766)</th>
<th>Raloxifene, 60 mg/day (N=1,792)</th>
<th>Raloxifene, 120 mg/day (N=1,828)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Short Blessed Test (18)a</td>
<td>2.3</td>
<td>3.0</td>
<td>2.3</td>
<td>2.9</td>
</tr>
<tr>
<td>Trail Making Tests (19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>52.8</td>
<td>26.4</td>
<td>53.7</td>
<td>26.4</td>
</tr>
<tr>
<td>B</td>
<td>102.6</td>
<td>40.4</td>
<td>103.9</td>
<td>40.7</td>
</tr>
<tr>
<td>Word List Memory and Recall Tests (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td>7.2</td>
<td>1.3</td>
<td>7.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Recall</td>
<td>7.2</td>
<td>2.0</td>
<td>7.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Fluency</td>
<td>18.8</td>
<td>5.4</td>
<td>18.7</td>
<td>5.3</td>
</tr>
<tr>
<td>Buschke Selective Reminding Test (20)a</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

a At two study sites, the Short Blessed Test was not administered. The 188 women at these sites received the Buschke Selective Reminding Test instead.
veloping Alzheimer’s disease, and a nearly significantly lower risk of developing any cognitive impairment. The risk of dementia from any cause did not differ significantly between the group receiving 120 mg/day of raloxifene and the women taking placebo. When we repeated our analyses by categorizing the 104 women who had a high likelihood of cognitive impairment but did not start or complete the dementia evaluation as having any cognitive impairment, our results were similar.

The effects of raloxifene 60 and 120 mg/day of raloxifene on the risk for mild cognitive impairment and any cognitive impairment were significantly different (p=0.002 for mild cognitive impairment and p=0.007 for any cognitive impairment), indicating that the results of the two dose groups cannot be pooled.

Women taking 120 mg/day of raloxifene who had undetectable serum estradiol levels at baseline had a slightly lower relative risk for each cognitive outcome (compared to the placebo group) than women with detectable estradiol levels. For example, the risk for any cognitive impairment among those who were assigned to 120 mg/day of raloxifene and had undetectable estradiol levels was 0.62 (95% CI=0.37–1.04) and for those with detectable levels the risk was 0.84 (95% CI=0.50–1.39), but these interactions were not statistically significant (p=0.34 for all interactions).

### Discussion

Raloxifene, at a dose of 120 mg/day, reduced the risk of developing mild cognitive impairment by 33% among postmenopausal women with osteoporosis. In the MORE trial, one case of mild cognitive impairment was prevented for every 91 women treated. There was also a somewhat lower risk of developing Alzheimer’s disease among the women treated with the higher dose of raloxifene that did not reach statistical significance. Given that few women developed Alzheimer’s disease, the lower risk observed among women assigned to 120 mg/day of raloxifene may have occurred by chance. However, most people with mild cognitive impairment progress to develop dementia over several years, and the majority of these develop Alzheimer’s disease (3, 4). Thus, a drug that lowers the risk of mild cognitive impairment might also be expected to reduce the risk of developing Alzheimer’s disease. As our results are tentative, trials enrolling women at high risk for cognitive impairment or Alzheimer’s disease, or longer trials, are required to estab-

### Table 3. Characteristics of the 742 Womena Who Completed Phases 1 and 2 of the Dementia Assessment in the Multiple Outcomes of Raloxifene Evaluation, by Final Cognitive Status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Normal (N=509)</th>
<th>Mild Cognitive Impairment (N=181)</th>
<th>Dementia (N=52)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>F (ANOVA) df p</td>
</tr>
<tr>
<td>Years from randomization to dementia evaluation</td>
<td>3.8 0.3</td>
<td>3.7 0.3</td>
<td>3.7 0.3</td>
<td>1.37 2, 739 0.25</td>
</tr>
<tr>
<td>Test scores at dementia evaluation</td>
<td></td>
<td></td>
<td></td>
<td>H (Kruskal-Wallis test) p</td>
</tr>
<tr>
<td>Mini-Mental State Examination</td>
<td>28.6 1.5</td>
<td>26.7 2.4</td>
<td>20.6 4.5</td>
<td>205.39 2 &lt;0.001</td>
</tr>
<tr>
<td>Clinical Dementia Rating</td>
<td>0.11 0.21</td>
<td>0.44 0.16</td>
<td>0.80 0.54</td>
<td>320.26 2 &lt;0.001</td>
</tr>
<tr>
<td>Global</td>
<td>0.20 0.48</td>
<td>0.98 0.89</td>
<td>4.69 3.28</td>
<td>362.47 2 &lt;0.001</td>
</tr>
<tr>
<td>Sum of boxes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline variables</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>F (ANOVA) df p</td>
</tr>
<tr>
<td>Age (years)</td>
<td>67.2 6.7</td>
<td>69.3 6.3</td>
<td>71.5 5.2</td>
<td>15.81 2, 739 &lt;0.001</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>25.5 4.1</td>
<td>25.7 3.8</td>
<td>24.9 3.9</td>
<td>0.66 2, 738 0.52</td>
</tr>
<tr>
<td>Education (years)</td>
<td>10.8 3.7</td>
<td>8.9 3.7</td>
<td>9.3 4.4</td>
<td>31.49 2 &lt;0.001</td>
</tr>
<tr>
<td>Years since menopause</td>
<td>19.2 8.0</td>
<td>22.0 8.2</td>
<td>24.2 6.9</td>
<td>27.73 2 &lt;0.001</td>
</tr>
<tr>
<td>Geriatric Depression Scale score</td>
<td>1.5 1.9</td>
<td>2.2 2.7</td>
<td>3.3 3.8</td>
<td>8.60 2 0.01</td>
</tr>
<tr>
<td>White race</td>
<td>96.1 9.3</td>
<td>90.4 4.8</td>
<td>4.08 2 0.13</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>18.4 16.8</td>
<td>7.7 3.8</td>
<td>3.84 2 0.15</td>
<td></td>
</tr>
<tr>
<td>Alcohol use (&gt;3 drinks/week)</td>
<td>16.0 9.9</td>
<td>5.8 7.11</td>
<td>2.10 2 0.35</td>
<td></td>
</tr>
<tr>
<td>History of hysterectomy</td>
<td>25.3 21.6</td>
<td>30.8 2.35</td>
<td>2.35 2 0.31</td>
<td></td>
</tr>
<tr>
<td>History of myocardial infarction</td>
<td>1.2 3.1</td>
<td>3.9 1.57</td>
<td>2.46 2 0.46</td>
<td></td>
</tr>
<tr>
<td>History of stroke</td>
<td>0.4 3.20</td>
<td>0.0 5.23</td>
<td>2 0.07</td>
<td></td>
</tr>
<tr>
<td>Prior hormone replacement therapy</td>
<td>29.6 32.0</td>
<td>15.7 5.23</td>
<td>2 0.07</td>
<td></td>
</tr>
<tr>
<td>Detectable serum estradiol (≥5 pmol/liter)</td>
<td>49.5 48.3 48.9 0.08 2 0.96</td>
<td>49.5 48.3 48.9 0.08 2 0.96</td>
<td>49.5 48.3 48.9 0.08 2 0.96</td>
<td></td>
</tr>
</tbody>
</table>

*a Two subjects were referred to the dementia evaluation but did not complete it.
RALOXIFENE AND DEMENTIA

TABLE 4. Relative Risk of Cognitive Impairment, Alzheimer’s Disease, or Any Dementia Among the 742 Womena Who Completed Phases 1 and 2 of the Dementia Assessment in the Multiple Outcomes of Raloxifene Evaluation

<table>
<thead>
<tr>
<th>Cognitive Outcome and Treatment Group</th>
<th>Women With Outcome After 3 Years of Treatment</th>
<th>Percent of Total Treatment Groupb</th>
<th>Relative Riskb</th>
<th>95% Cl</th>
<th>pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild cognitive impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>63</td>
<td>3.6</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raloxifene, 60 mg/day</td>
<td>74</td>
<td>4.2</td>
<td>1.18</td>
<td>0.85–1.64</td>
<td>0.32</td>
</tr>
<tr>
<td>Raloxifene, 120 mg/day</td>
<td>44</td>
<td>2.5</td>
<td>0.67</td>
<td>0.46–0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>15</td>
<td>0.9</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raloxifene, 60 mg/day</td>
<td>13</td>
<td>0.7</td>
<td>0.82</td>
<td>0.39–1.71</td>
<td>0.60</td>
</tr>
<tr>
<td>Raloxifene, 120 mg/day</td>
<td>8</td>
<td>0.4</td>
<td>0.52</td>
<td>0.22–1.21</td>
<td>0.12</td>
</tr>
<tr>
<td>Any type of dementia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>18</td>
<td>1.0</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raloxifene, 60 mg/day</td>
<td>17</td>
<td>1.0</td>
<td>0.90</td>
<td>0.47–1.74</td>
<td>0.76</td>
</tr>
<tr>
<td>Raloxifene, 120 mg/day</td>
<td>17</td>
<td>1.0</td>
<td>0.91</td>
<td>0.47–1.76</td>
<td>0.78</td>
</tr>
<tr>
<td>Dementia or mild cognitive impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>81</td>
<td>4.7</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raloxifene, 60 mg/day</td>
<td>91</td>
<td>5.2</td>
<td>1.12</td>
<td>0.84–1.49</td>
<td>0.45</td>
</tr>
<tr>
<td>Raloxifene, 120 mg/day</td>
<td>61</td>
<td>3.4</td>
<td>0.73</td>
<td>0.53–1.01</td>
<td>0.054</td>
</tr>
</tbody>
</table>

a Two women were referred to the dementia evaluation but did not complete it.
b Relative to the placebo group, adjusted for country.

The reduction in risk of cognitive impairment with raloxifene is due to an anti-estrogen effect. Another possible explanation for the differences between our results and those of the Women’s Health Initiative Memory Study may be related to cardiovascular disease. Unlike estrogen (8, 9), raloxifene is not associated with an increase in stroke and other cardiovascular outcomes that may mediate cognitive impairment (35). It is interesting that the women with undetectable baseline estradiol levels had a greater reduction in risk associated with 120 mg/day of raloxifene than did the women with higher baseline estradiol levels, although this interaction was not statistically significant.

We previously studied the effect of 3 years of raloxifene treatment, in both the 60 and 120 mg/day groups, on cognitive functioning in women enrolled in the MORE study and reported that there were no differences on mean 3-year scores between treatment groups (36). However, compared to the women assigned to placebo, those assigned to raloxifene had less risk of developing impairment (as assessed by a cutoff in the change in scores) in verbal memory and attention, cognitive domains affected earliest in mild cognitive impairment and Alzheimer’s disease (37). In addition, among the women who were at risk for cognitive impairment because they were over 70 years of age, those assigned to raloxifene had smaller 3-year declines in memory and attention. Data from a small, randomized trial of 120 mg/day of raloxifene for women with Alzheimer’s disease also indicated improvement in verbal memory in the raloxifene group after 12 weeks of treatment (38).

Our trial has several limitations. Most of the women enrolled in the MORE study were white, and all had osteoporosis. We cannot be sure that our results apply to women without osteoporosis, to nonwhite women, or to men. Even though the trial was large, the rate of develop-
ing dementia is low until advanced old age, and we had limited power to detect differences in the risk of developing Alzheimer’s disease and other types of dementia. Assuming a two-tailed alpha of 0.05 and 80% power, we were able to detect relative risks of 0.22 or less for Alzheimer’s disease, 0.57 or less for mild cognitive impairment, and 0.61 or less for any cognitive impairment. The observed incidence of 1% per year for the development of any cognitive impairment among the MORE participants is at the low end of reported incidence rates for women ages 65–69 years (39). These lower rates are not surprising given that all women who enrolled in the MORE study had high levels of functioning at baseline and volunteered for a long-term trial. Our observed mean MMSE and Clinical Dementia Rating scores for women in each diagnostic group are consistent with scores previously reported for these diagnoses (40, 41), supporting the validity of the criteria used in the dementia evaluation. However, given that some women may have had very mild cognitive deficits at baseline, we cannot determine whether raloxifene (120 mg/day) prevented cognitive impairment or delayed its progression. Finally, some women who were referred for the clinical dementia evaluation did not go or did not complete the evaluation. The number of women in this group did not differ by treatment group, and when we repeated our analyses and coded these women as having cognitive impairment, our results were similar.

Cognitive impairment and especially dementia are devastating conditions that cause severe debility and death among older persons. Thus far, we know of no intervention that has been proven to reduce the risk of developing these conditions directly. Raloxifene at a dose of 120 mg/day (but not 60 mg/day) resulted in a reduced risk of mild cognitive impairment and a nearly significant reduction in any cognitive impairment (mild cognitive impairment or dementia) in postmenopausal osteoporotic women. Additional trials of raloxifene and other selective estrogen receptor modulators for prevention of cognitive impairment, especially in women at high risk, should be conducted to confirm these results.

References

Received March 23, 2004; revision received June 2, 2004; accepted June 14, 2004. From the Departments of Psychiatry, Neurology, Epidemiology, and Medicine, University of California, San Francisco; Lilly Research Laboratories, Eli Lilly and Company, Indianapolis; the Departments of Geriatrics, Neurology, Pharmacology and Toxicology, and Epidemiology, University of Arkansas for Medical Sciences, Little Rock; and the Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center and Division of Epidemiology, University of Minnesota, Minneapolis. Address correspondence and reprint requests to Dr. Yaffe, University of California, San Francisco, Box 181, 4150 Clement St., San Francisco, CA 94121; kyaffe@itsa.ucsf.edu (e-mail).

Funded by Eli Lilly and Company and by support to Dr. Yaffe from the Paul Beeson Faculty Scholars Program, a Mt. Zion/UCSF Women’s Health Grant, and grant RO1 AG-021918 from the National Institute on Aging.

The authors thank the MORE investigators who participated in the Dementia Diagnostic Evaluation ancillary study.
RALOXIFENE AND DEMENTIA


Neuropsychological Differences Between Late-Onset and Recurrent Geriatric Major Depression

Michael A. Rapp, M.D., Ph.D.
Karen Dahlman, Ph.D.
Mary Sano, Ph.D.
Hillel T. Grossman, M.D.
Vahram Haroutunian, Ph.D.
Jack M. Gorman, M.D.

Objective: Executive dysfunction, possibly related to vascular pathology, has been well documented in patients with a first episode of major depressive disorder in later life (late-onset geriatric major depression). However, it is unclear whether the neuropsychological presentation differs in patients with a lifetime history of major depressive disorder (recurrent geriatric major depressive disorder). The purpose of this study was to explore differences in neuropsychological function, symptoms, and cardiovascular comorbidity between patients with late-onset and recurrent geriatric major depression.

Method: The study used a two-by-two factorial design in which one factor was current major depressive disorder (present versus absent) and the second factor was lifetime history of depression (present versus absent). Neuropsychological measures of executive functioning and episodic memory, as well as psychopathological symptoms and comorbid medical illness, were examined in a total of 116 older adults.

Results: Patients with late-onset major depressive disorder showed specific deficits in attention and executive function, whereas patients with recurrent major depressive disorder exhibited deficits in episodic memory. The rates of anhedonia and comorbid cardiovascular illness were higher in patients with late-onset geriatric major depressive disorder.

Conclusions: In contrast to recurrent geriatric major depressive disorder, late-onset major depressive disorder is characterized by specific deficits in tasks of attention and executive function, consistent with increased anhedonia and cardiovascular comorbidity. These findings, if confirmed, suggest that recurrent and late-onset geriatric major depressive disorder may represent distinct phenomenological entities. Such phenomenological differences as a function of lifetime history of major depression can guide research in the neurophysiology, prevention, and treatment of geriatric major depressive disorder.

Major depressive disorder is a highly prevalent disease in old age (1–3). Between 1% and 4% of elderly persons experience major depression, and the prevalence increases to between 6% and 32% in older nursing home residents (4). At the same time, the onset of major depression is variable across the life course. For the majority of persons, the age at onset is in the late 20s, but first episodes are also common after age 40 (5, 6). Epidemiologically, about 40% of cases of major depression in old age represent recurrent depressive episodes, whereas about 30% reflect late-onset depression, while in some cases the distinction between recurrent and late-onset depression cannot be determined reliably (7).

There is evidence to suggest that cerebrovascular disease, especially ischemic small-vessel disease, may be a factor in the pathogenesis of late-onset geriatric major depression. Imaging studies of late-onset geriatric major depressive disorder consistently showed signal hyperintensities in deep white matter (8–10) that may go along with structural brain changes in the frontal lobes (11). These findings have led to the hypothesis that among older persons with major depression, there is a subgroup of individuals with what has been termed “vascular depression” (12–14).

At the same time, there is evidence to suggest that recurrent, early-onset major depressive disorder is associated with significant volume loss in the hippocampus (15, 16). Sheline et al. (15) reported an association of the length of untreated depressive episodes with reductions in hippocampal volume in recurrent geriatric major depressive disorder. Bell-McGinty and colleagues (16) found an inverse correlation between bilateral hippocampal-entorhinal volume and years since onset of depression. These findings have recently been linked to models of decreased hippocampal neurogenesis in major depressive disorder, suggesting that recurrent depressive episodes may lead to persistent neuronal alterations on a molecular level in hippocampal cells (17).

Such findings of structural differences between late-onset and recurrent geriatric major depressive disorder inform clinical hypotheses on differences in the neuropsychological and symptom presentation of geriatric major depression. Vascular depression has been associated with
a loss of motivation or interest and the presence of cognitive decline, especially in attention and executive functioning, consistent with frontal lobe dysfunction (12, 18).

On the other hand, preliminary evidence suggests that impaired short-term memory functioning may be associated with reduced temporal lobe volume in patients with geriatric depression (19) and in middle-aged adults (20). Other reports suggest that depressed patients may have significant deficits primarily in episodic memory, suggesting a more selective dysfunction in mesial temporal lobe function during episodes of depression (21). Memory dysfunction has been shown to be persistent in older depressed patients, even after response to antidepressant treatment (22). Taken together, these findings suggest that recurrent episodes of major depressive disorder may go along with some degree of persistent dysfunction in the mesial temporal lobe, possibly associated with impairment in episodic memory functions.

Based on these considerations, we hypothesized that the neuropsychological profiles of patients with recurrent versus late-onset geriatric major depressive disorder would differ in the degree of dysfunction in tasks of attention and executive function (consistent with fronto-subcortical circuit dysfunction) versus dysfunction in tasks of episodic memory (consistent with temporal lobe dysfunction). Specifically, we hypothesized that patients with late-onset geriatric major depressive disorder would exhibit specific deficits in attention and executive function, whereas patients with recurrent geriatric major depressive disorder would exhibit specific deficits in episodic memory functioning. Furthermore, we hypothesized, consistent with a vascular depression model of late-onset major depressive disorder, that patients with late-onset geriatric major depressive disorder would present with a higher degree with vascular comorbidity.

**Method**

**Subjects**

The study builds on the neuropsychological portion of a prospective, longitudinal study of cognition in old age, the Clinical and Biological Studies of Early Alzheimer’s Disease project, at the Department of Psychiatry, Mount Sinai School of Medicine. For the present analyses, we used baseline neuropsychological data from 299 older nursing home residents from the Jewish Home Nursing Home, Bronx, New York (mean age=83.64 years, SD=4.39). These data represent all nondemented participants who completed baseline neuropsychological and psychiatric assessments. The diagnosis of dementia was made in a research consensus conference according to DSM-III-R or DSM-IV criteria. Ethical approval was obtained from the institutional review boards of the Department of Veterans Affairs, Bronx, N.Y., and the Mount Sinai School of Medicine, New York. Written informed consent to participate in the study was obtained from each participant or, if the participant lacked capacity, from a caregiver.

**Patients With Geriatric Major Depressive Disorder**

The neuropsychological battery was administered, along with a standardized questionnaire assessing psychiatric history and current symptoms, and comprised assessment with the Geriatric Depression Scale (23). Trained research assistants completed a standardized questionnaire assessing the presence or absence of DSM-III-R or DSM-IV symptoms of major depressive disorder. The questionnaire was a modified version of the mood disorders module from the Structured Clinical Interview for DSM-IV Axis I Disorders (24). The presence or absence of a lifetime history of major depressive disorder was extracted from medical information, including charts and information obtained from the treating physician. Both the diagnosis of current major depressive disorder and the diagnosis of a lifetime history of major depressive disorder were reviewed and verified by a research physician with 2 years of specialty training in geriatric psychiatry.

Of the 299 participants, 40 (10.03%) met the criteria for major depressive disorder. Using psychiatric symptom and history data, we defined recurrent geriatric major depressive disorder as present in patients who had at least one episode of major depressive disorder according to psychiatric history, and late-onset geriatric major depressive disorder as present in patients who did not have a history of major depressive disorder. With this classification, 19 patients were defined as having late-onset geriatric major depressive disorder and 21 as having recurrent geriatric major depressive disorder. The external validity of the diagnosis of major depressive disorder was assessed in comparison to scores on the Geriatric Depression Scale by using a cutoff of 11 points. With a standard Geriatric Depression Scale cutoff score of 11 or greater, the diagnosis of major depressive disorder had the following classification accuracy indices (25, 26); hit rate=0.96, sensitivity=0.95, specificity=0.90, false positive rate=0.05, and false negative rate=0.10.

**Nondepressed Older Adults**

In the remaining 259 nondepressed participants, 123 had a history of lifetime depressive disorder and 111 did not (data were missing for 15 participants). We used twofold oversampling in a randomized matching procedure to avoid inflation of type I error, which has been shown to be significant when sample sizes are unequal by a factor of 5 or more (27). Hence, we randomly selected 42 participants without a lifetime diagnosis of depression and 38 with a lifetime diagnosis of major depressive disorder, matched for age and gender. Two nondepressed older adults without and one nondepressed older adult with a lifetime diagnosis of major depressive disorder had incomplete data and were excluded from the sample. The random sample did not differ from the overall group in age (t=0.28, df=257, p=0.78), gender (χ²=1.40, df=1, p=0.24), and general cognitive status as measured by the Mini-Mental State Examination (MMSE) (28) (t=1.02, df=257, p=0.31).

**Assessments**

Medical diagnoses were extracted from medical information, including charts and information obtained from the treating physician, by using a standardized checklist assessing the presence or absence of cardiovascular and cerebrovascular diagnoses. Specifically, during the chart review process, checklists were used to assess the documented presence of cardiovascular and cerebrovascular diagnoses.

Specific symptoms of depression were derived from the questionnaire assessing the presence or absence of DSM-III-R or DSM-IV symptoms of major depressive disorder, as either reported by subjective account or observed by clinicians. Specifically, the symptom “mood” represents “depressed mood most of the day, nearly every day.” “Anhedonia” represents “diminished interest or pleasure in activities of the day.” “Neurovegetative symptoms” represent any one of a group of phenomena, including psychomotor agitation or retardation and changes in sleep, weight, and appetite. The neuropsychological battery was administered by trained research assistants. Scores on all tasks administered were as-
TABLE 1. Demographic Characteristics of Currently Nondepressed Older Adults With Versus Without a Past History of Major Depressive Disorder and Older Adult Patients With Recurrent Versus Late-Onset Major Depressive Disorder

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Currently Nondepressed Older Adult Comparison Subjects</th>
<th>Older Adult Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No History of Major Depressive Disorder (N=39)</td>
<td>Recurrent Major Depressive Disorder (N=21)</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Female sex</td>
<td>21</td>
<td>53.85</td>
</tr>
<tr>
<td>Age (years)</td>
<td>84.05</td>
<td>6.82</td>
</tr>
<tr>
<td>Education (years)</td>
<td>9.84</td>
<td>2.03</td>
</tr>
<tr>
<td>Mini-Mental State Examination score</td>
<td>26.00</td>
<td>2.63</td>
</tr>
<tr>
<td>Geriatric Depression Scale score</td>
<td>6.13</td>
<td>3.93</td>
</tr>
</tbody>
</table>

a Participants with neither a lifetime nor a current diagnosis of major depressive disorder.
b Participants with a lifetime but not a current diagnosis of major depressive disorder.
c Participants with a lifetime and a current diagnosis of major depressive disorder.
d Participants without a lifetime but with a current diagnosis of major depressive disorder.

*Significant difference among groups (F=4.15, df=3, 112, p<0.05), in post hoc comparisons, no history of major depressive disorder > past history of major depressive disorder > recurrent major depressive disorder = late-onset major depressive disorder (t tests).

Results

Demographic and Clinical Characteristics

The overall study group comprised 46 men (39.66%) and 70 women (60.34%), with a mean age of 83.41 years (SD=8.37, range=60–97) and a mean education level of 10.04 years (SD=2.05, range=6–14). There were no statistically significant differences between the four study groups in age (F=0.14, df=3, 115, p=0.93), education (F=0.29, df=3, 115, p=0.83), and gender (χ²=1.40, df=3, p=0.71). The baseline characteristics of the four groups are summarized in Table 1.

Overall cognitive status was assessed with the MMSE. Cognitive status was worse in the currently depressed older adults (patients with late-onset and recurrent major depressive disorder), compared to the nondepressed older adults (F=8.43, df=1, 112, p<0.01). Further testing revealed that in the nondepressed group, cognitive performance was significantly worse in the participants who had a lifetime diagnosis of depression (t=1.99, df=74, p<0.05), whereas overall cognitive status was comparable between the older adults with recurrent and late-onset major depressive disorder (t=0.89, df=38, p=0.38).
Depression severity was measured with the Geriatric Depression Scale. Geriatric Depression Scale scores were higher in patients with current major depressive disorder, compared to the nondepressed participants (F=141.93, df=1, 112, p<0.001), but Geriatric Depression Scale scores did not differ between the two currently depressed groups (t=0.56, df=38, p=0.58) nor between the two currently nondepressed groups (t=0.54, df=74, p=0.59).

**Differences in Neuropsychology**

The eight neuropsychological variables (Table 2) were aggregated in a factor-analytic approach. The factorial solution over both groups generated two factors, accounting for 68.12% of the overall variance in the eight cognitive tasks. Factor loadings on the first factor were high for Trail Making Parts A and B, the digit symbol substitution test, and the verbal fluency test, suggesting a representation of attention/executive function in this factor (“attention/executive function”). The memory measures loaded high on the second factor, suggesting a representation of episodic memory (“episodic memory”). The factor loadings for the factor solution are given in Table 3.

For the attention/executive function factor, ANOVA revealed a main effect for the current presence of major depressive disorder (F=4.25, df=1, 112, p<0.05), a main effect for the presence or absence of a lifetime diagnosis of major depressive disorder (F=5.37, df=1, 112, p<0.05), and an interaction between current and lifetime major depressive disorder (F=6.24, df=1, 112, p<0.01). Post hoc tests revealed that attention/executive function differed between the older adults with and without current major depressive disorder (t=2.51, df=114, p<0.05), but not between the currently nondepressed older adults with and without a lifetime diagnosis of major depressive disorder (t=0.16, df=74, p=0.87). However, there was a significant difference among the currently depressed older adults as a function of lifetime diagnosis of major depressive disorder (t=2.78, df=38, p<0.01), indicating that the patients with late-onset major depressive disorder performed worse in attention/executive function.

For the episodic memory factor, ANOVA revealed a main effect for the current presence of major depressive disorder (F=7.15, df=1, 112, p<0.01) and a main effect for the presence of a lifetime diagnosis of major depressive disorder (F=4.36, df=1, 112, p<0.05). The interaction term, however, was not significant (F=1.95, df=1, 112, p=0.17). Post hoc tests revealed no significant differences on the episodic memory tasks between currently nondepressed older adults with and without a lifetime diagnosis of major depressive disorder (t=0.59, df=74, p=0.55). The patients with recurrent major depressive disorder, however, performed worse on the episodic memory tasks than the patients with late-onset geriatric major depressive disorder (t=2.10, df=38, p<0.05).

As Figure 1 shows, among currently depressed patients only, the mean factor scores for attention/executive function were higher in patients with recurrent major depressive disorder, whereas episodic memory scores were higher in patients with late-onset major depressive disorder. This interaction between cognitive domain and type of depression (recurrent versus late-onset) proved to be statistically significant (F=13.33, df=1, 38, p<0.001) and represented a moderate effect size (η²=0.26). This pattern of results suggests a dissociation in the neuropsychological presen-
versus without a past history of major depressive disorder

a Factor loadings from principal-component analysis with varimax rotation.

b Factor loading greater than 0.5.

ation of late-onset versus recurrent geriatric major depressive disorder, with specific impairment in attention/executive function in late-onset major depressive disorder and specific impairment in episodic memory in recurrent geriatric major depressive disorder.

Furthermore, we were interested in whether the neuropsychological differences between the groups may have been driven by subgroups of patients among the currently depressed older adults. Specifically, we tested whether skewness, a statistical indicator of distribution symmetry (with a value of zero indicating normal distribution [36]), was significantly different from zero for either factor score in either group. This was not the case (all p > 0.11), indicating that distributions were in fact symmetric.

Differences in Depressive Symptoms and Medical Comorbidity

We further explored whether these differences in cognitive performance went along with differences in depressive symptoms and medical comorbidity between patients with late-onset major depressive disorder and those with recurrent geriatric major depressive disorder. Data for medical comorbidity and depressive symptoms are listed in Table 4.

There were no significant differences in depressive symptoms in the currently nondepressed groups (those with and without a lifetime diagnosis of depression) (all p > 0.35). Likewise, both the number of subjects with diabetes (\(\chi^2 = 0.05, df = 1, p = 0.99\)) and the number of subjects with cardiovascular disease (\(\chi^2 = 2.27, df = 1, p = 0.16\)) were comparable in the currently nondepressed groups with and without a lifetime diagnosis of depression.

The analysis of depressive symptoms in the currently depressed patients as a function of lifetime diagnosis (late-onset versus recurrent geriatric major depressive disorder) revealed no significant differences in mood (\(\chi^2 = 0.48, df = 39, p = 0.60\)) and neurovegetative symptoms of depression (\(\chi^2 = 0.42, df = 39, p = 0.55\)). However, patients with late-onset geriatric major depressive disorder were more likely to exhibit anhedonia than patients with recurrent major depressive disorder (\(\chi^2 = 8.39, df = 39, p = 0.01\)). Likewise, patients with late-onset geriatric major depressive disorder were more likely to have comorbid cardiovascular disease than those with recurrent major depressive disorder (\(\chi^2 = 10.51, df = 39, p < 0.001\)), whereas there were no significant differences between these groups in the number of patients with diabetes mellitus (\(\chi^2 = 0.36, df = 39, p = 0.65\)).

Discussion

In line with our hypotheses, we found distinct differences in both the neuropsychological and clinical presentation between late-onset and recurrent geriatric major depression. The principal finding is that recurrent geriatric major depression is characterized by deficits in epi-
sodic memory, whereas late-onset geriatric major depression is characterized by specific deficits in attention/executive function. Neuropsychological performance was compromised in currently depressed older adults, relative to comparison subjects both with and without a lifetime diagnosis of depression. However, within the group of older adults with major depressive disorder, patients with recurrent geriatric major depressive disorder performed worst in tasks of episodic memory, and patients with late-onset geriatric major depressive disorder performed worst in tasks of attention/executive function. In addition, late-onset geriatric major depressive disorder was characterized by the presence of a higher degree of anhedonia and a higher rate of cardiovascular comorbidity.

This study adds to previous studies on geriatric major depressive disorder that have shown executive dysfunction, loss of motivation, an increased number of cardiovascular risk factors, and a higher prevalence of vascular brain changes in patients with late-onset geriatric major depressive disorder (9–15, 37, 38). Beyond those findings, however, the results of this study indicate that recurrent geriatric major depressive disorder may represent a distinct phenomenological subtype within geriatric major depression, in contrast to late-onset geriatric major depressive disorder. The presence of such subtypes is suggested by the dissociation between neuropsychological functions, with specific frontal lobe dysfunction in late-onset geriatric major depressive disorder and specific temporal lobe dysfunction in recurrent geriatric major depressive disorder.

Although these findings are in line with the vascular model of late-onset geriatric major depressive disorder (8–10, 12, 13), as contrasted to temporal lobe dysfunction as a consequence of recurrent depressive episodes and underlying neuronal changes (15–17), the exact mechanisms that may lead to these distinct subtypes in geriatric major depression are not known. Our data suggest that cardiovascular comorbidity may play a role in the development of specific executive dysfunction in late-onset geriatric major depressive disorder. Although the cross-sectional analysis of our data does not permit causal inference, it seems reasonable to assume that vascular changes in the frontal cortex might go along with both anhedonia and executive dysfunction. However, we cannot clearly establish this relationship from our data.

The suggestion that temporal lobe abnormalities are a potential mechanism for episodic memory dysfunction in recurrent geriatric major depressive disorder is supported by both neuropsychological and neuroimaging data (19–22). One hypothesis that can be derived from the episodic memory deficit in recurrent geriatric major depressive disorder found in this study is that recurrent depressive episodes may lead to temporal lobe dysfunction through longstanding effects of decreased hippocampal neurogenesis (17).

The clinical significance of this study is that identification of the neuropsychological profile, clinical presentation, and additional risk factors in subtypes of geriatric major depressive disorder may lead to the development of specific pharmacological or nonpharmacological interventions to address the different subtypes of the disorder. Such specific interventions may be especially needed because of the high rate of treatment nonresponse in geriatric major depressive disorder (39, 40). There are data to indicate that more comprehensive treatment approaches to geriatric major depressive disorder may yield higher response rates (41), and it would be interesting to see whether differential treatment of late-onset and recurrent geriatric major depressive disorder yields similar results.

The prevalences of current major depressive disorder and lifetime history of depression in our study group of very old nursing home residents are comparable to those in representative samples. The reported point prevalence of major depressive disorder in older nursing home residents ranges from 6% to 32% (7) and is thus comparable to the prevalence of about 10% in our study. Furthermore, some studies of young adults have found lifetime prevalences comparable to those in our study (4). To our knowledge, there is only one study of the lifetime prevalence of

### Table 4. Prevalence of Psychopathology and Medical Comorbidity Among Currently Nondepressed Older Adults With Versus Without a Past History of Major Depressive Disorder and Older Adult Patients With Recurrent Versus Late-Onset Major Depressive Disorder

<table>
<thead>
<tr>
<th>Variable</th>
<th>Currently Nondepressed Older Adult</th>
<th></th>
<th>Older Adult Patients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No History of Major Depressive Disorder</td>
<td>Past History of Major Depressive Disorder</td>
<td>Recurrent Major Depressive Disorder</td>
<td>Late-Onset Major Depressive Disorder</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N=39)</td>
<td>(N=37)</td>
<td>(N=21)</td>
<td>(N=19)</td>
<td>(N=21)</td>
</tr>
<tr>
<td>Psychopathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressed mood</td>
<td>1</td>
<td>2.56</td>
<td>3</td>
<td>8.10</td>
<td>20</td>
</tr>
<tr>
<td>Anhedoniaa</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
<td>2.70</td>
<td>7</td>
</tr>
<tr>
<td>Neurovegetative symptoms</td>
<td>2</td>
<td>5.12</td>
<td>1</td>
<td>2.70</td>
<td>11</td>
</tr>
<tr>
<td>Comorbidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular diseaseb</td>
<td>22</td>
<td>56.41</td>
<td>27</td>
<td>72.97</td>
<td>12</td>
</tr>
<tr>
<td>Diabetes</td>
<td>6</td>
<td>15.38</td>
<td>5</td>
<td>13.51</td>
<td>2</td>
</tr>
</tbody>
</table>

a Significant difference between older adult patients with recurrent and with late-onset major depressive disorder ($\chi^2=8.39, df=1, p<0.01$).
b Significant difference between older adult patients with recurrent and with late-onset major depressive disorder ($\chi^2=10.51, df=1, p<0.001$).
major depressive disorder in adults ages 70 years and older (42). This study reported prevalences ranging from 23% to 45% (42), comparable to our finding of a lifetime prevalence of 33.6%. Overall, the subjects in our study represent older adults with a high level of overall cognitive functioning, as measured with the MMSE, who were able to complete a full neuropsychological assessment. Individuals with dementia were excluded in order to rule out effects of cognitive dysfunction due to neurodegenerative disorders.

This study is limited by its cross-sectional design with discrete age groupings, a feature that precludes conclusions about the timing of age effects in patients with major depression. Ideally, studies of geriatric depression would follow older adults with a known presence or absence of past major depressive disorder longitudinally, to identify both age- and disease-related changes over time and their associations with risk factors and indicators of potential underlying mechanisms. Furthermore, we could not extract reliable data on the exact time (i.e., other than before age 65 years) of the onset of the first episode of major depressive disorder in the study group. Prior studies showed that, when informants and structured interviews are used, interrater reliability may exceed 80% (43). Yet, it seems vital to further refine methods to assess lifetime history of major depressive disorder, especially in older adults. Such information would enable correlative studies of neuropsychological performance decrements in recurrent geriatric major depression.

One conceptual problem underlying research on geriatric major depressive disorder as a function of lifetime history of depression is that in fact recurrent or subthreshold depression across the lifespan may in turn increase the risk of vascular pathology. Likewise, recurrent episodes of major depressive disorder across the lifespan may have different underlying etiologies. The inclusion of behavioral, genetic, and cardiovascular variables in such studies would allow for a better understanding of common pathways and differential risk factors.

This study addressed some of the limitations of prior studies by using a two-by-two design, thus disentangling effects of a lifetime history of major depression from effects of the current presence of major depressive disorder. Assessment of depression with a state measure (Geriatric Depression Scale) assured inclusion of depressed subjects with similar levels of current depression severity and made the concurrent validation of depression diagnoses possible. Controlling for age and gender in matched randomly selected groups of subjects helped to prevent the introduction of variance that might obscure group differences.

In conclusion, the contribution of this study is the delineation of specific subtypes within geriatric major depression as a function of the presence or absence of a lifetime diagnosis of depression. Further research is needed to replicate these phenomenological differences. Findings from this study may be used clinically to guide the design of treatment interventions for specific subtypes of geriatric depression and thus may provide clinicians with more satisfying treatment options for this prevalent and disabling disorder.

References


GERIATRIC MAJOR DEPRESSION

34. Schulman AI: Recognition memory for targets from a scanned word list. Br J Psychol 1971; 62:335–346
42. Palsson PS, Ostling S, Skoog I: The incidence of first-onset depression in a population followed from the age of 70 to 85. Psychol Med 2001; 31:1159–1168
MRI Hyperintensities and Depressive Symptoms in a Community Sample of Individuals 60–64 Years Old

Objective: Previous studies have found associations of magnetic resonance imaging (MRI) signal hyperintensities with depression in the elderly. The present study investigates the association in a younger community sample (age 60–64 years) of depressed subjects and comparison groups for potential mediating and confounding variables.

Method: A subsample of 475 persons 60–64 years of age from a larger community survey underwent brain MRI scans. White matter hyperintensities were quantified by using an automated procedure, and basal ganglia hyperintensities were quantified by using semiquantitative visual ratings. The study also assessed depressive symptoms and use of antidepressant medication. Potential mediating or confounding variables assessed included physical disability, hypertension, stroke, diabetes, head injury, cortisol, thyroid-stimulating hormone, cognitive functioning, smoking, and alcohol use.

Results: Depressive symptoms were found to be related to total brain white matter hyperintensities but not to basal ganglia hyperintensities. However, associations disappeared when statistical adjustment was made for physical disability and smoking.

Conclusions: Depressive symptoms are related to white matter hyperintensities in mid-adult life in a community sample. Physical disability appears to play an important role in this association.

Magnetic resonance imaging (MRI) studies of patients with major depression have found a higher prevalence of signal hyperintensities on T2-weighted imaging, particularly in subjects with later depression onset (1). Such findings have influenced the concept of vascular depression, a late-onset subtype of depression that involves increased cardiovascular risk factors and hyperintensities of deep white matter or subcortical gray matter (2, 3).

The early work on hyperintensities and depression was carried out in clinical samples. However, more recently there have been studies in community samples of older people. In a Rotterdam scan study (4), 1,077 persons 60–90 years of age with severe white matter hyperintensities (WMHs) were found to have 3–5 times the risk of depressive symptoms, with the association being stronger for deep than for periventricular WMHs. Similar findings emerged from a scanning study of 92 high-functioning normal elderly volunteers 66–80 years of age (5). Although none of these volunteers had clinical depression, deep WMHs, but not periventricular WMHs, were associated with depressive symptoms. The association was especially strong in individuals carrying the APOE-4 allele. In the Cardiovascular Health Study (6), scans were carried out on 3,236 persons ≥65 years of age, and depressive symptoms were monitored over the subsequent 7 years. Depressive symptoms were associated with small lesions of the basal ganglia, large cortical white matter lesions, and severe subcortical white matter grade.

Most of the research on WMHs and depression has involved elderly samples, and little is known about the association in middle age. Here we report data from a community sample of individuals 60–64 years old that assessed the association between depressive symptoms and white matter and basal ganglia hyperintensities. We also investigated a range of potential mediating or confounding variables that might account for any associations found. These variables were sociodemographic characteristics, physical disability, history of stroke, history of head injury, diabetes, thyroid-stimulating hormone (TSH), cortisol, cognitive functioning, brain atrophy, hypertension, smoking, and alcohol use.

Method

Participants

A community survey was carried out with 2,551 adults 60–64 years of age living in Canberra (Australian Capital Territory) and the neighboring town of Queanbeyan (New South Wales). The participants were part of the PATH Through Life Project, which is a longitudinal study of social, psychological, and biological risk factors for high-prevalence mental health problems. The project involves cohorts initially 20–24, 40–44, and 60–64 years of age, with a plan to conduct follow-up assessments every 4 years for 20 years. Further details on the full PATH Through Life Project have been published previously (7, 8). The present study concerns a
subset of the data from the 60–64-year-old cohort. The participants were recruited randomly from the electoral roll (enrollment to vote is compulsory for Australian citizens), with a response rate of 58.3% for the 60–64 age group. When characteristics of this sample were compared to census data on the population, the sample was found to be better educated but similar in marital status and employment status.

Participants were interviewed by a team of professional survey interviewers experienced in epidemiological and social surveys. Most interviews were carried out in the participant's home, but subjects could also choose to be interviewed at the Centre for Mental Health Research by the same interviewer. Participants were asked to self-complete a questionnaire on a Hewlett-Packard 620LX palmtop personal computer using the Surveycraf software (SPSS Inc., Chicago) for computer-assisted personal interviewing. The interview covered sociodemographic characteristics, anxiety and depression, substance abuse, cognitive function, well-being, physical health, health habits, use of health services, personality, coping styles, early life psychosocial risk factors, and current psychosocial risk factors. The interviewer never saw the answers to these questions presented on the computer. Some basic physical tests were also carried out by the interviewer (e.g., blood pressure, grip strength, visual acuity, lung functioning, memory, reaction time), and the participants were asked to provide a cheek swab from which DNA could be extracted.

During the community survey, respondents were asked whether they would be willing to undergo an MRI scan; 2,076 (81.4%) out of 2,551 said they would. Subsequently, a randomly selected subsample of 622 of the willing 2,076 were invited to undergo a scan, and 478 (76.8%) participated. This subgroup also provided a blood sample. Those who underwent a scan were significantly more likely than the full sample to be of English-speaking background ($\chi^2=5.37, df=1, p=0.02$) and had significantly better physical health per the SF-12 physical component score (t=4.30, df=2539, p<0.001), more years of education (t=2.83, df=2544, p=0.005), and better cognitive functioning per the Mini-Mental State Examination (t=3.76, df=2533, p<0.001), the Spot-the-Word test (t=2.04, df=2483, p=0.05), and a recall task (t=3.42, df=2549, p<0.001). However, there was not a significant difference in depressive symptoms (t=1.54, df=2537, p=0.13).

**Magnetic Resonance Imaging**

All subjects were imaged with a 1.5-Tesla Philips Gyroscan ACS-NT scanner (Philips Medical Systems, Best, the Netherlands) for $T_1$-weighted three-dimensional structural and $T_2$-weighted fluid attenuated inversion recovery (FLAIR) sequence MRI. A scout mid-sagittal image was first acquired to locate anterior to the posterior commissure plane. The $T_1$-weighted MRI was acquired in coronal orientation using a $T_1$-FFE sequence with the following parameters: repetition time (TR)/echo time (TE)=28.05/2.64 msec; flip angle=30°; matrix size=256×256; field of view=260×260 mm; slice thickness=2.0 mm; and mid-slice to mid-slice distance=1.0 mm, yielding overcontiguous coronal slices. The FLAIR sequence was acquired in coronal orientation with TR/TE/TI=11000/140/2600 msec; matrix size=256×256; field of view=230×230 mm; slice thickness=4.0 mm, with no gap between slices.

**Quantification of Hyperintensities**

MRI scans were transferred to an independent Windows NT workstation and visualized by using the software package ANALYZE (Mayo Foundation, Rochester, Minn.). A special computer program written by one of us (W.W.) automatically delineated WMHs in both the periventricular and deep regions, and this was validated against visual ratings. The major steps of the WMH volumetric estimation are as follows: 1) making an age-specific FLAIR template; 2) coregistration of FLAIR and $T_1$ images of the same subject, using FLAIR as the target and $T_1$ as the source; 3) spatial normalization of the coregistered FLAIR and $T_1$ images into Talairach space; 4) detecting and grading WMHs from each FLAIR image, with $T_1$ as a reference, for the removal of false WMH detection caused by artifacts such as partial-volume effect; 5) visual inspection of each WMH map resulting from the computer algorithm, and manual removal of any wrongly classified WMHs from the map; and 6) computing the total WMH volume in each brain and the separate WMH volumes in each lobe and arterial territory using corresponding masks. The details of the process are described elsewhere (9). The absolute volume of total white matter and WMHs were determined, and the percentage of white matter with a hyperintense signal was calculated for each subject. The results of the automated process were examined by an experienced operator for face validity. Twenty scans were processed twice to determine the restest reliability of the procedure, and 100% correspondence was noted. For concurrent validity, the scans were visually rated by two independent clinicians experienced in examining MRI scans on a modified Fazekas scale (10). One hundred scans were rated visually, and the intraclass correlations (ICC) between the automated measures and the visual ratings were modest (whole brain WMHs: ICC=0.43 [F=1.76, df=1, 99, p=0.003]; deep WMHs: ICC=0.63 [F=2.74, df=1, 99, p<0.001]; periventricular WMHs: ICC=0.59 [F=2.44, df=1, 99, p<0.001]). Pearson correlations were also calculated (whole brain: r=0.79, df=98, p<0.001; deep: r=0.72, df=98, p<0.001; periventricular: r=0.72, df=98, p<0.001). The automated process did not identify hyperintensities in gray matter (e.g., basal ganglia), for which visual ratings on the modified Fazekas scale were used. The interrater reliability (weighted kappa) for these ratings, established on 57 cases, was 1.00 for right basal ganglia extent, 0.89 for left basal ganglia extent, 0.90 for right basal ganglia intensity, and 0.80 for left basal ganglia intensity.

**Assessment of Depressive Symptoms and Syndrome**

As part of the community survey, participants completed the depression section of the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (11), which asks about DSM-IV major depressive symptoms in the last 2 weeks. This questionnaire was scored to give a continuous scale from 0 to 27 and also to classify participants as having a major or minor depressive syndrome. To be classified with major depressive syndrome, a participant had to have five or more symptoms, including “little interest or pleasure in doing things” or “feeling down, depressed, or hopeless.” For minor depressive syndrome, a person had to have two to four symptoms, including the two specified. Participants were also asked “In the last month, have you taken or used any medications (including herbal remedies) for depression?” In some of the analyses reported below, those who responded “yes” were grouped together with those classified as having a depressive syndrome on the Patient Health Questionnaire.

**Assessment of Potential Mediators or Confounders**

Age and gender were recorded during the community survey. Participants were asked a series of questions about educational achievements, and these were used to calculate years of education. Physical disability was assessed with the physical component summary of the 12-item Short Form Health Survey (12). History of stroke was assessed by asking “Have you ever suffered a stroke, ministroke or TIA (transient ischemic attack)?” History of head injury was assessed by asking “Have you ever had a serious head injury where you became unconscious for more than 15 minutes?” Diabetes was defined as following a diabetic diet, using oral hypoglycemics or insulin, or having a fasting blood glucose of ≥10. Other blood tests examined were TSH and cortisol levels. Cognitive functioning was assessed with the Mini-Mental State Examination (13). Brain atrophy was measured by dividing cerebrospinal fluid volume by intracranial volume. Blood pressure...
TABLE 1. Regional White Matter Hyperintensity Density in a Community Sample of Individuals 60–64 Years of Age, by Depression Classification

<table>
<thead>
<tr>
<th>Region</th>
<th>Depression Syndrome Subjects (N=17)</th>
<th>Nondepressed Subjects Taking Depression Medication (N=21)</th>
<th>Nondepressed Subjects Not Taking Depression Medication (N=437)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Frontal</td>
<td>1.11 ± 0.22</td>
<td>0.87 ± 0.22</td>
<td>0.34 ± 0.24</td>
</tr>
<tr>
<td>Temporal</td>
<td>0.27 ± 0.16</td>
<td>0.16 ± 0.06</td>
<td>0.17 ± 0.15</td>
</tr>
<tr>
<td>Parietal</td>
<td>0.84 ± 0.44</td>
<td>0.78 ± 0.42</td>
<td>0.36 ± 0.29</td>
</tr>
<tr>
<td>Occipital</td>
<td>1.02 ± 0.66</td>
<td>0.99 ± 0.67</td>
<td>0.73 ± 0.66</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>0.01 ± 0.00</td>
<td>0.01 ± 0.00</td>
<td>0.01 ± 0.00</td>
</tr>
<tr>
<td>Anterior horn</td>
<td>9.47 ± 7.16</td>
<td>7.25 ± 5.16</td>
<td>6.43 ± 5.98</td>
</tr>
<tr>
<td>Posterior horn</td>
<td>3.87 ± 2.80</td>
<td>2.16 ± 1.19</td>
<td>1.91 ± 1.70</td>
</tr>
<tr>
<td>Periventricular body</td>
<td>7.62 ± 6.08</td>
<td>5.63 ± 4.25</td>
<td>5.44 ± 5.13</td>
</tr>
<tr>
<td>Total brain white matter</td>
<td>1.48 ± 1.08</td>
<td>1.15 ± 0.79</td>
<td>0.81 ± 0.73</td>
</tr>
</tbody>
</table>

Abbreviations:
a Significant difference among groups.
b Adjusted for gender.
c Subjects classified as having major depressive syndrome (five or more symptoms from the depression section of the Patient Health Questionnaire, including “little interest or pleasure in doing things” and “feeling down, depressed, or hopeless”) or minor depressive syndrome (no more than four symptoms, including the two specified).

Statistical Analysis

Initial descriptive analyses involved comparing mean hyperintensities for depression syndrome subjects, nondepressed subjects taking depression medication, and nondepressed subjects not taking depression medication. This comparison was carried out by using estimated marginal means (95% confidence intervals) after we adjusted for gender.

Predictors of depression caseness were evaluated by using logistic regression, with the odds ratio used as the index of effect size. Predictors of depression symptom scores were evaluated by using negative binomial regression, which was used in preference to linear regression because of the extreme skew of the score distribution. For the negative binomial regressions, effect size was measured by incidence rate ratios, which give the rate of increase in depression symptoms for each unit increase in the predictor variable (17).

Variables were evaluated as potential mediators or confounders by assessing whether they had a significant Pearson correlation with both depression symptom score and total brain WMHs. If a variable fulfilled both of these conditions, it was entered as a covariate in the logistic and negative binomial regressions predicting depression from hyperintensities. The significance level was set at p<0.05.

Results

There were 475 participants with data for both MRI hyperintensities and depressive symptoms. Given the small number of cases, those with major and minor depressive syndrome were combined into one depression syndrome group. There were 17 participants in the depression syndrome group (five with major and 12 with minor depressive syndrome; mean Patient Health Questionnaire score=10.88 [SD=3.41]). There were 21 participants who were taking depression medication (prescription [N=17] or nonprescription [N=4]) but who were not currently syndromal (mean Patient Health Questionnaire score=5.38 [SD=4.02]), and 437 nondepressed participants not taking depression medication (mean Patient Health Questionnaire score=1.81 [SD=2.12]). English was not the first language of 9.5% of the participants, but this did not differ significantly between the depression groups.

We present the results for WMHs first, then the basal ganglia hyperintensities. Associations with all WMHs and severe WMHs were generally similar, so only the former are presented here. Table 1 shows the mean regional WMH density for the three groups. It can be seen that the depression syndrome subjects had a higher density of WMHs in most regions, while the nondepressed subjects taking depression medication tended to be intermediate between the other two groups. Infarct-like lesions were present in six subjects (1.2%), four subcortical and two cortical. None of these participants was rated as depressed.

An examination of correlations among the regional WMHs found that they were generally high across regions, justifying the use of a total brain WMH score as a global summary measure. The exception was the cerebellum, which was uncorrelated with the other regions.

Evaluation of potential mediators or confounders (Table 2) found only three variables that fulfilled the condition of being correlated with both total brain WMHs and depression score. These were physical disability, being a current...
TABLE 2. Correlations of Potential Mediators or Confounders With Depressive Symptom Score and Total Brain White Matter Hyperintensities (WMHs)

<table>
<thead>
<tr>
<th>Potential Mediator or Confounder</th>
<th>Correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depressive Symptom Score</td>
</tr>
<tr>
<td>Age</td>
<td>0.02</td>
</tr>
<tr>
<td>Female gender</td>
<td>0.12*</td>
</tr>
<tr>
<td>Years of education</td>
<td>-0.10*</td>
</tr>
<tr>
<td>Physical disability</td>
<td>-0.18*</td>
</tr>
<tr>
<td>History of stroke</td>
<td>0.08</td>
</tr>
<tr>
<td>History of head injury</td>
<td>-0.04</td>
</tr>
<tr>
<td>Diabetes</td>
<td>-0.01</td>
</tr>
<tr>
<td>TSH</td>
<td>0.02</td>
</tr>
<tr>
<td>Cortisol</td>
<td>0.04</td>
</tr>
<tr>
<td>Mini-Mental State Examination</td>
<td>0.00</td>
</tr>
<tr>
<td>Brain atrophy</td>
<td>0.16*</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.09</td>
</tr>
<tr>
<td>Current smoker</td>
<td>0.16*</td>
</tr>
<tr>
<td>Past smoker</td>
<td>0.04</td>
</tr>
<tr>
<td>Hazardous/harmful alcohol user</td>
<td>0.02</td>
</tr>
<tr>
<td>Abstainer or occasional alcohol user</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*p<0.05.

smoker, and female gender. Mean physical disability scores (with lower scores indicating greater disability) were 44.32 (SD=14.85) for the depression syndrome group, 45.65 (SD=11.56) for the nondepressed subjects taking depression medication, and 50.30 (SD=8.13) for the nondepressed subjects not taking depression medication. The prevalence of smoking in the three groups was 23.5% for the depression syndrome group, 14.3% for the nondepressed subjects taking depression medication, and 6.6% for the nondepressed subjects not taking depression medication. Women comprised 65% of the depression syndrome group, 57.6% of the nondepressed subjects taking depression medication, and 46% of the nondepressed subjects not taking depression medication. These three variables (gender, smoking, and physical disability) were therefore included as covariates in regression analyses predicting depression from WMHs. Table 3 shows the results adjusted for gender only, and Table 4 shows the results adjusted for gender, smoking, and physical disability. It can be seen that most of the significant associations disappear after adjustment for all three covariates. Only two associations remain significant (out of 27 regression analyses), and one association (cerebellar hyperintensities with depression symptom score) is actually in the opposite direction to prediction. In the regression analyses summarized in Table 4, physical disability was the only variable to significantly predict depression in all 27 analyses. Smoking was significant in 12 of the 18 logistic regression analyses, but in none of the negative binomial regressions. Gender was not significant in any analysis.

Analysis of manual ratings of basal ganglia hyperintensities showed them to be uncommon. For the right basal ganglia, 97% of the sample had no hyperintensities. For the left basal ganglia, the figure was 98%. Regression analyses predicting depression from hyperintensity ratings showed no significant associations.

Discussion

The present study of a community sample of individuals 60–64 years old replicates earlier findings from both clinical and community samples of an association between white matter hyperintensities and depression. While some earlier studies reported associations specifically with deep WMHs (4, 18), the present study found evidence of associations in both frontoparietal and periventricular white matter. This is not surprising given the generally high correlations found between WMHs across these regions. We did not find a specific relationship between depressive symptoms and hyperintensities in the frontal white matter. This is not inconsistent with the literature, since only a few authors have reported such specificity in the association (19).

Contrary to the findings of the Cardiovascular Health Study (6), no associations were found between basal ganglia hyperintensities and depressive symptoms. However, the prevalence of basal ganglia hyperintensities in the present sample was low (only 2%–3%, compared with >25% in the Cardiovascular Health Study), which limits the ability to detect any association. The basal ganglia have received much attention in the depression literature, with many studies having found decreased basal ganglia volumes or hyperintense lesions in these structures in late-onset depression patients (20). It is noteworthy, however, that basal ganglia volumes were found to be normal in depressed subjects who were physically healthy in one study (21). While the focus of our report is hyperintensities rather than the volume of brain structures, this finding raises issues about the relationship among depression, the brain, and physical health, which we will subsequently discuss.

The dominant interpretation of the association between WMHs and depression is that these lesions disrupt frontal-subcortical circuits controlling mood and thereby produce “vascular depression” (3). However, we found that the association between WMHs and depressive symptoms disappeared after statistical adjustment for physical disability and current smoking. In regression analyses in which WMHs, physical disability, and smoking were simultaneously entered as predictors, only physical disability had a significant association with depressive symptoms. This finding is consistent with results from the Cardiovascular Health Study that found MRI infarcts and white matter grade were no longer associated with depressive symptoms after adjustment for physical disability and cognitive impairment (18, 22). The Cardiovascular Health Study also found that physical disability continued to be a strong predictor of depressive symptoms independent of MRI infarcts and white matter grade.
TABLE 3. Association of Depression With Regional White Matter Hyperintensity Density, Adjusting for Gender

<table>
<thead>
<tr>
<th>Region of WMH Density</th>
<th>Depression Syndromea</th>
<th>Depression Syndrome or Medication for Depression</th>
<th>Depression Symptom Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>95% CI</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Frontal</td>
<td>1.28</td>
<td>1.03–1.61</td>
<td>1.34</td>
</tr>
<tr>
<td>Temporal</td>
<td>2.55</td>
<td>0.80–8.07</td>
<td>1.7</td>
</tr>
<tr>
<td>Parietal</td>
<td>1.37</td>
<td>0.99–1.91</td>
<td>1.42*</td>
</tr>
<tr>
<td>Occipital</td>
<td>1.44</td>
<td>0.87–2.40</td>
<td>1.46*</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>0.82</td>
<td>0.00–1095.45</td>
<td>0.52</td>
</tr>
<tr>
<td>Anterior horn</td>
<td>1.11*</td>
<td>1.02–1.20</td>
<td>1.07*</td>
</tr>
<tr>
<td>Posterior horn</td>
<td>1.21*</td>
<td>1.06–1.37</td>
<td>1.14*</td>
</tr>
<tr>
<td>Parietals body</td>
<td>1.16*</td>
<td>1.03–1.30</td>
<td>1.09</td>
</tr>
<tr>
<td>Total brain white matter</td>
<td>1.49*</td>
<td>1.09–2.04</td>
<td>1.47*</td>
</tr>
</tbody>
</table>

* Excludes participants taking medication who are not current depression syndrome cases.
* p<0.05.

TABLE 4. Association of Depression With Regional White Matter Hyperintensity Density, Adjusting for Gender, Physical Disability, and Smoking

<table>
<thead>
<tr>
<th>Region of WMH Density</th>
<th>Depression Syndromea</th>
<th>Depression Syndrome or Medication for Depression</th>
<th>Depression Symptom Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>95% CI</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Frontal</td>
<td>1.18</td>
<td>0.88–1.57</td>
<td>1.19</td>
</tr>
<tr>
<td>Temporal</td>
<td>1.52</td>
<td>0.43–5.16</td>
<td>1.10</td>
</tr>
<tr>
<td>Parietals</td>
<td>1.26</td>
<td>0.86–1.86</td>
<td>1.24</td>
</tr>
<tr>
<td>Occipital</td>
<td>1.31</td>
<td>0.77–2.22</td>
<td>1.35</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>0.61</td>
<td>0.00–454.39</td>
<td>0.41</td>
</tr>
<tr>
<td>Anterior horn</td>
<td>1.09</td>
<td>0.99–1.19</td>
<td>1.04</td>
</tr>
<tr>
<td>Posterior horn</td>
<td>1.21*</td>
<td>1.03–1.42</td>
<td>1.11</td>
</tr>
<tr>
<td>Parietals body</td>
<td>1.13</td>
<td>1.00–1.28</td>
<td>1.05</td>
</tr>
<tr>
<td>Total brain white matter</td>
<td>1.39</td>
<td>0.93–2.08</td>
<td>1.28</td>
</tr>
</tbody>
</table>

* Excludes participants taking medication who are not current depression syndrome cases.
* p<0.05.

There are a number of possible interpretations of the role of physical disability. The first is that WMHs cause physical disability, which in turn causes depressive symptoms. In other words, physical disability is a mediator or intervening variable in the causal path (23). Consistent with this possibility, several other studies have found that WMHs are associated with physical disability (24), motor impairments (25), and impaired balance (26), and there is evidence that physical disability is a risk factor for depression in older people (27). Another possibility is that WMHs independently cause both depressive symptoms and physical disability; this is less likely, however, since WMHs predict disability after depressive symptoms are controlled but do not predict depressive symptoms after disability is controlled. However, because our data are cross-sectional, there is a limitation on causal inference. A better test of the causal paths would be provided by longitudinal data. The present sample is due for follow-up after 4 years, which will provide a stronger basis for causal inference. Whatever the causal mechanism, the presence of WMHs associated with depressive symptoms in midlife justifies efforts at preventing this source of depression at the population level (28).

Finally, a number of limitations of the study must be acknowledged. The prevalence of depression was low, and the diagnosis was not made by a clinician but with a self-completed interview. Our subjects were in mid-adult life, while the majority of the vascular depression literature concerns elderly individuals over the age of 65. It is possible that the association between depressive symptoms and cerebrovascular disease becomes stronger in the elderly. Depression, when present in our subjects, was mild in intensity, and it is possible that the clinical picture of vascular depression is one of severe depression, often resistant to treatment and needing hospitalization. Furthermore, those who accepted to have a scan tended to be healthier, with better cognitive functioning and higher education and who spoke English as their first language. On the other hand, the study has considerable strengths, with its relatively large community sample covering a narrow age range and the automated assessment of regional WMHs. It therefore offered us the opportunity to examine the relationship between WMHs and depressive symptoms at the mild end of the spectrum, in which confounding factors of other illnesses and cognitive impairment were less likely to have an impact.
HYPERINTENSITIES AND DEPRESSIVE SYMPTOMS

Dr. Jorm, Centre for Mental Health Research, Australian National University, Canberra 0200, Australia; Anthony.Jorm@anu.edu.au (e-mail).

This study was supported by project and program grants from the National Health and Medical Research Council.

The authors thank the staff of the PATH Through Life Project for their contributions.

References

7. Jorm AF, Dear KBG, Rodgers B, Christensen H: Interaction between mother's and father's affection as a risk factor for anxiety and depression symptoms: evidence for increased risk in adults who rate their father as having been more affectionate than their mother. Soc Psychiatry Psychiatr Epidemiol 2003; 38:173–179

Predictors of Antidepressant Use Among Older Adults: Have They Changed Over Time?

Dan G. Blazer, M.D., Ph.D.
Celia F. Hybels, Ph.D.
Gerda G. Fillenbaum, Ph.D.
Carl F. Pieper, Dr.P.H.

Objective: Antidepressant use increased substantially among older adults with the introduction of the new-generation medications such as the selective serotonin reuptake inhibitors. The authors analyzed data from two follow-up intervals—1986–1987 to 1989–1990 (interval 1) and 1992–1993 to 1996–1997 (interval 2)—from a community-based cohort of 4,162 older adults to determine predictors of future antidepressant use.

Method: Information on antidepressant use, demographic and health characteristics, and categories of depressive symptoms—positive affect, negative affect, somatic complaints, and interpersonal problems—were obtained. Logistic regression was used to control simultaneously for multiple variables predicting antidepressant use during the two intervals. Repeated-measures logistic regression (with generalized estimating equations) was employed to model the probability of antidepressant use, with adjustment for the effect of time.

Results: Prior antidepressant use and white race were strong predictors of future use during both intervals. Negative affect was the only additional significant predictor of use during interval 1. In contrast, low positive affect scores, cognitive impairment, and poorer health were additional significant predictors during interval 2. In a repeated-measures model, race, prior antidepressant use, poor health, low positive affect scores, and somatic complaints varied as predictors over time. Negative affect and cognitive impairment were consistent predictors over time.

Conclusions: The predictors of antidepressant use by older adults changed over time, with health-related measures of quality of life, such as positive affect, health status, and somatic complaints, becoming more prominent as predictors of use.

TABLE 1. Demographic and Health-Related Variables at the Beginning of Interval 1 (1986–1987) and Interval 2 (1992–1993) for Older Adults With Antidepressant Use Data at the Beginning and End of the Intervala

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interval 1 Subjects (N=3,017)</th>
<th>Interval 2 Subjects (N=1,540)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years, as of 1986–1987)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–74</td>
<td>1,949</td>
<td>1,147</td>
</tr>
<tr>
<td>≥75</td>
<td>1,068</td>
<td>393</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,003</td>
<td>470</td>
</tr>
<tr>
<td>Female</td>
<td>2,014</td>
<td>1,070</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White or other</td>
<td>1,361</td>
<td>697</td>
</tr>
<tr>
<td>African American</td>
<td>1,656</td>
<td>843</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than ninth grade education</td>
<td>1,576</td>
<td>737</td>
</tr>
<tr>
<td>Ninth grade education or more</td>
<td>1,441</td>
<td>803</td>
</tr>
<tr>
<td>Health index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1,190</td>
<td>330</td>
</tr>
<tr>
<td>Medium</td>
<td>824</td>
<td>602</td>
</tr>
<tr>
<td>High</td>
<td>985</td>
<td>589</td>
</tr>
<tr>
<td>Cognitive status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not impaired</td>
<td>2,674</td>
<td>1,319</td>
</tr>
<tr>
<td>Impaired</td>
<td>313</td>
<td>213</td>
</tr>
<tr>
<td>Mean</td>
<td>89.5</td>
<td>139</td>
</tr>
<tr>
<td>SD</td>
<td>10.5</td>
<td>13.9</td>
</tr>
</tbody>
</table>

Center for Epidemiologic Studies Depression Scale (modified) score

<table>
<thead>
<tr>
<th>Measure</th>
<th>Interval 1</th>
<th>Interval 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3.11</td>
<td>2.73</td>
</tr>
<tr>
<td>Positive affect</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>Negative affect</td>
<td>1.17</td>
<td>1.00</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>1.39</td>
<td>1.32</td>
</tr>
<tr>
<td>Interpersonal problems</td>
<td>0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>

a Some totals do not add up to 3,017 or 1,540 because of missing data.

Depression was assessed by using a modified version of the CES-D Scale, and identified four factors: 1) positive affect—seven items (felt as good as other people, felt hopeful about the future, felt happy, and enjoyed life), 2) negative affect—seven items (could not shake the blues, felt depressed, thought life was a failure, felt fearful, felt lonely, had crying spells, and felt sad), 3) somatic complaints—seven items (bothered by things, appetite poor, trouble concentrating, felt everything an effort, sleep restless, talked less than usual, and could not get going), and 4) interpersonal problems—two items (people were unfriendly and felt people disliked me). This factor structure was confirmed in the Duke Established Populations for Epidemiologic Studies of the Elderly sample (14). The ranges of scores for the four subscales were 0–4 for positive

import covariates, such as prior use of antidepressants, race, cognitive impairment, and education (variables that we previously found to be associated with antidepressant use in cross-sectional analysis) (5). We also hypothesized that, with the introduction of the new-generation antidepressants, some of the predictors of future antidepressant use would change over time, specifically that factors other than the core symptoms of depression would be predictive of future use during interval 2.

Method

Participants

Data for this study are from the Duke Established Populations for Epidemiologic Studies of the Elderly (10, 11). The Duke sample consisted of community residents selected from five contiguous Piedmont counties in North Carolina, one of which was predominantly urban and the other four predominantly rural. The Duke Established Populations for Epidemiologic Studies of the Elderly was a 10-year prospective cohort study. The sampling design has been described in detail previously (10). Briefly, the study used a four-stage probability sample of 4,162 people age ≥65 years, 54% of whom were African American. (With the exception of 26 subjects, all designated their race as either white or African American. Those who designated their race as “other” were categorized with the white subjects in this analysis.) A baseline interview was conducted in 1986–1987, and three additional in-person interviews were conducted in 1989–1990, 1992–1993, and 1996–1997. All subjects (or their designated proxy respondents) signed a written consent form approved by the Duke Institutional Review Board.

Measures

A comprehensive demographic section of the interview assessed age, sex, race, and education. The demographic variables at baseline were dichotomized and coded as follows: age (1=65–74 years, 2=≥75 years), self-designated race (1=white/other, 2=African American), and education (1=<9 years, 2=≥9 years). Health status was measured with a health index that assessed the weighted number of chronic illnesses self-reported by the subject and was dichotomized, with lower values indicating better medical status (12). Cognitive status was assessed with the 10-item Short Portable Mental Status Questionnaire (13). The scale score was dichotomized (1=three or more errors, 0=less than three errors) (13).

Depression

Depression was assessed by using a modified version of the CES-D Scale (6, 11). The modified version included all 20 items from the original scale and is highly comparable to the original version (11). Subjects answered the items by responding either yes or no, rather than using the four options from the original scale, for a range of scores from 0 to 20. Radloff (6) and Weissman et al. (9) explored the factor structure of the original 20-item scale and identified four factors: 1) positive affect—four items (felt as good as other people, felt hopeful about the future, felt happy, and enjoyed life), 2) negative affect—seven items (could not shake the blues, felt depressed, thought life was a failure, felt fearful, felt lonely, had crying spells, and felt sad), 3) somatic complaints—seven items (bothered by things, appetite poor, trouble concentrating, felt everything an effort, sleep restless, talked less than usual, and could not get going), and 4) interpersonal problems—two items (people were unfriendly and felt people disliked me). This factor structure was confirmed in the Duke Established Populations for Epidemiologic Studies of the Elderly sample (14).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Interval 1</th>
<th></th>
<th>p</th>
<th></th>
<th>Interval 2</th>
<th></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≥75 years</td>
<td>1.09</td>
<td>0.76–1.56</td>
<td>0.64</td>
<td></td>
<td>0.92</td>
<td>0.60–1.40</td>
<td>0.68</td>
</tr>
<tr>
<td>Male sex</td>
<td>0.85</td>
<td>0.58–1.24</td>
<td>0.39</td>
<td></td>
<td>0.80</td>
<td>0.53–1.21</td>
<td>0.30</td>
</tr>
<tr>
<td>African American race</td>
<td>0.44</td>
<td>0.31–0.64</td>
<td>&lt;0.001</td>
<td></td>
<td>0.29</td>
<td>0.20–0.44</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Less than ninth grade education</td>
<td>1.10</td>
<td>0.78–1.56</td>
<td>0.60</td>
<td></td>
<td>0.66</td>
<td>0.45–0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>Cognitive impairment</td>
<td>1.65</td>
<td>1.02–2.66</td>
<td>0.04</td>
<td></td>
<td>1.62</td>
<td>1.02–2.58</td>
<td>0.04</td>
</tr>
<tr>
<td>Health index</td>
<td>1.17</td>
<td>0.95–1.43</td>
<td>0.14</td>
<td></td>
<td>1.31</td>
<td>1.02–1.69</td>
<td>0.04</td>
</tr>
<tr>
<td>Positive affect score</td>
<td>1.19</td>
<td>0.96–1.49</td>
<td>0.12</td>
<td></td>
<td>1.67</td>
<td>1.30–2.13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative affect score</td>
<td>1.25</td>
<td>1.14–1.37</td>
<td>&lt;0.001</td>
<td></td>
<td>1.15</td>
<td>1.04–1.27</td>
<td>0.01</td>
</tr>
<tr>
<td>Somatic complaints score</td>
<td>1.21</td>
<td>1.10–1.33</td>
<td>&lt;0.001</td>
<td></td>
<td>1.07</td>
<td>0.96–1.19</td>
<td>0.20</td>
</tr>
<tr>
<td>Interpersonal problems score</td>
<td>1.15</td>
<td>0.79–1.69</td>
<td>0.46</td>
<td></td>
<td>0.71</td>
<td>0.38–1.33</td>
<td>0.29</td>
</tr>
<tr>
<td>Overall Center for Epidemiologic Studies Depression Scale Score</td>
<td>1.10</td>
<td>1.06–1.56</td>
<td>&lt;0.0001</td>
<td></td>
<td>1.06</td>
<td>1.01–1.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Took antidepressants at index episode</td>
<td>44.37</td>
<td>27.50–71.59</td>
<td>&lt;0.0001</td>
<td></td>
<td>13.02</td>
<td>7.69–22.05</td>
<td>&lt;0.00</td>
</tr>
</tbody>
</table>

*a All tests are two-tailed. One degree of freedom for each test.

centages for all covariates for interval 1 and interval 2. Unadjusted odds ratios for antidepressant use at follow-up were then calculated for all covariates. Next, logistic regression was used to estimate the direct effect of each of the independent variables on antidepressant use at follow-up for each of the two intervals. All variables were entered in the regression analysis simultaneously. Finally, a repeated-measures logistic regression model (with generalized estimation equations or generalized estimating equations under PROC GENMOD in SAS [20]) was employed.

**Results**

Among participants for whom antidepressant data were available (that is, subjects who reported use/no use) during interval 1, 86 (2.9%) reported antidepressant use in 1986–1987 and 134 (4.5%) reported such use in 1989–1990. Among participants in interval 2, 66 (4.3%) reported antidepressant use in 1992–1993 and 125 (8.1%) reported such use in 1996–1997. Among subjects taking antidepressants at 1986–1987, 57% were also taking them 3 years later. Among subjects taking antidepressants in 1992–1993, 47% were taking them 4 years later.

The demographic characteristics of the samples are presented in Table 1. Data are presented for those subjects for whom antidepressant data were available for both interviews during interval 1 or both interviews during interval 2. As would be expected, subjects in interval 2 were younger and more likely to be female than the subjects in interval 1, reflecting the survival characteristics of the cohort. With aging, the cohort was more likely to report poorer self-rated health and more likely to be cognitively impaired. Scores on the entire CES-D Scale and the four subscales were slightly better at the beginning of interval 2 than at the beginning of interval 1.

As Table 2 shows, bivariate predictors from logistic regression analysis of antidepressant use during both intervals included white race (African American race was “protective” against use), cognitive impairment, negative affect, and antidepressant use at the beginning of the interval (by far the strongest predictor). The overall CES-D Scale score (entered as a continuous variable) and the somatic complaint scale score were significant predictors for antidepressant use.
of future use during interval 1 and interval 2. In contrast, higher education (lower education was “protective” against use), poorer health index, and low positive affect scores were significant predictors of future use during the interval 2.

The full logistic models shown in Table 3 demonstrate a clearly different pattern of predictors for interval 1, compared to interval 2. Prior antidepressant use and race remain strong predictors during both intervals. Negative affect was the only additional significant predictor of future use during interval 1. In contrast, negative affect was not a significant predictor for the later time period. Poor health index, cognitive impairment, and low positive affect scores were significant predictors. In a separate analysis, the composite CES-D Scale was a significant predictor in the full model during interval 1 but not interval 2.

To accommodate a repeated-measures design, we used a generalized estimating equations model in which we entered the CES-D Scale subscale scores, cognitive status measures, the health index, and data on use of antidepressant medications at the index wave as time-varying covariates along with demographic variables. A term classified as “interval” was then entered as a separate variable (coded 1 for the first interval and 2 for the second) to control for the effect of time. In the initial full repeated-measures model, interval (time) was a highly significant predictor of future use (p<0.0001). In addition, race (p<0.0001), use of an antidepressant at the index wave (p<0.0001), cognitive impairment (p<0.03), and negative affect (p=0.008) were significant predictors of future use. To determine if these effects were stable over time, we entered 11 interaction terms simultaneously (one term for each covariate by interval) to determine if, as a set, the interaction terms were significant additions to the model (21). We conducted a log-likelihood ratio test with 11 degrees of freedom; the result was significant, indicating that one or more of the interaction terms was significant on a multiplicative scale. We next did a series of generalized estimating equations models and removed nonsignificant interaction terms, one step at a time, until all the remaining terms in the model were significant (21). The final model is presented in Table 4. We found that cognitive impairment and negative affect were significant as main effects in the model. The interaction terms for interval-by-race, taking antidepressants at the beginning of the interval, the health index, low positive affect scores, and high somatic complaint scores were each significant, suggesting that the effects of these factors varied over time. The effect of race, poor health, and low positive affect scores increased over time, while the effects of taking an antidepressant at the beginning of the interval and of high somatic complaint scores decreased over time.

### Discussion

The findings of this study expand our earlier findings (5) by documenting that white race was not only a correlate but also a strong predictor of antidepressant use and that the predictive strength of that variable increased from interval 1 (1986–1987 to 1989–1990) to interval 2 (1992–1993 to 1996–1997). In addition, previous antidepressant use predicted future use, although the predictive strength of that variable decreased from interval 1 to interval 2. We also confirmed our first hypothesis: the negative affect scale score predicted future antidepressant use regardless of interval. In contrast, the somatic complaints scale score was more predictive during interval 1 than during interval 2. Cognitive function was a predictor of future antidepressant use, perhaps reflecting the relatively high frequency of clinically significant depressive symptoms as cognitive dysfunction increases (22, 23).

The positive affect scale score and chronic illness became significant predictors of antidepressant use during interval 2, with increased use of antidepressants in this cohort. These findings follow the trend of using the new-generation antidepressant medications for a wider range of problems. The greater safety of these agents may have increased use in people with chronic illnesses. Greater safety also may lead to greater use among persons with subthreshold depression (24, 25).
The positive affect scale score may reflect overall physical and psychological well-being. Scoring low on these items may therefore suggest either a reaction to a less than optimal psychosocial environment or a perception that “things are just not as they should be.”

The strengths of this study include good response rates in a cohort of older adults living in one geographic area throughout the duration of the study. We used two analytic approaches to facilitate understanding of the data. The first permitted “eyeball” comparison of change over time across the two intervals; the second, a more sophisticated longitudinal analytic approach, took into account in a single analysis all changes that occurred since the start of the study.

There are a number of problems in the study design that could bias the results. First, follow-up occurred over 3–4 years, and many factors can intervene during these intervals to either increase or decrease the likelihood that the subjects will take antidepressants in the future, such as the occurrence of a first-time episode of major depression with no prior known risk. The subjects who reported use of antidepressants during interval 2, compared to those who participated during interval 1, were older, as survivors were more healthy than nonsurvivors. They were less healthy and slightly more cognitively impaired than subjects who reported use of antidepressants during interval 1, as would be expected. They also scored slightly lower on the four CES-D Scale subscales. In our previous study (5) we did not find age to be a significant predictor of antidepressant use when other factors were controlled. We do not know how long participants used antidepressants. Some subjects may have taken the drugs for the entire 3–4 years, others for the short periods that were captured in the follow-up interviews, and still others might have stopped and started use during the study. In addition, we have no data on the efficacy of the medications in symptom reduction. We only have data on medication use for the past 24 hours before the interview. The overall number of participants who took antidepressants was small; the width of the confidence interval for the use of antidepressants is noteworthy. Finally, the study subjects consisted of a group of survivors.

Nevertheless, we were able to capitalize on a fortuitous circumstance. New-generation antidepressants were introduced and became widely prescribed during the 10 years of the study (2). In consequence, we have been able to examine how characteristics of users have changed with the availability of these new medications.

We found that the basic characteristics of users (white race, prior use of antidepressants, negative symptoms indicative of depression) persisted across the two intervals and that the discrepancy in use between whites and African Americans appears to be increasing. We previously reported that African Americans in North Carolina were less likely to use over-the-counter medications and total medications, with the greatest differences found for psychotropic drugs and nutritional supplements (16). These medications perhaps are viewed by both doctors and their patients as less imperative to be prescribed than, for example, medications for hypertension (for which use does not vary by race). One may speculate that the dramatic increase in use of antidepressants reflects physicians’ greater comfort in prescribing these medications for patients who are less critically ill. If so, then the increasing disparity in the use of antidepressants by race suggests that the need for antidepressants among African Americans may be viewed as less critical to their health care than their need for other medications. Given that the frequency of depression does not vary by race (14), then community-dwelling older African Americans may be significantly undertreated.

Other characteristics of the user base have broadened to include physical health, cognitive status, and a broader array of depressive symptoms. Although the indications for


<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>SE</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>–3.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older age</td>
<td>–0.04</td>
<td>0.17</td>
<td>0.96</td>
<td>0.69–1.34</td>
<td>–0.24</td>
<td>0.81</td>
</tr>
<tr>
<td>Male sex</td>
<td>–0.06</td>
<td>0.16</td>
<td>0.94</td>
<td>0.69–1.30</td>
<td>–0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>African American race</td>
<td>1.29</td>
<td>0.23</td>
<td>3.64</td>
<td>2.31–5.73</td>
<td>5.57</td>
<td>0.00</td>
</tr>
<tr>
<td>Less than ninth grade education</td>
<td>–0.08</td>
<td>0.16</td>
<td>0.92</td>
<td>0.68–1.26</td>
<td>–0.49</td>
<td>0.62</td>
</tr>
<tr>
<td>Took antidepressant at index</td>
<td>1.97</td>
<td>0.30</td>
<td>7.21</td>
<td>4.00–12.97</td>
<td>6.58</td>
<td>0.00</td>
</tr>
<tr>
<td>Cognitive impairment</td>
<td>0.42</td>
<td>0.21</td>
<td>1.53</td>
<td>1.00–2.32</td>
<td>1.97</td>
<td>0.05</td>
</tr>
<tr>
<td>Health index</td>
<td>0.37</td>
<td>0.14</td>
<td>1.44</td>
<td>1.09–1.93</td>
<td>2.55</td>
<td>0.01</td>
</tr>
<tr>
<td>Interval</td>
<td>–0.16</td>
<td>0.36</td>
<td>0.86</td>
<td>0.42–1.72</td>
<td>–0.44</td>
<td>0.66</td>
</tr>
<tr>
<td>Interpersonal problems</td>
<td>–0.30</td>
<td>0.21</td>
<td>0.74</td>
<td>0.49–1.13</td>
<td>–1.39</td>
<td>0.16</td>
</tr>
<tr>
<td>Positive affect</td>
<td>0.38</td>
<td>0.14</td>
<td>1.46</td>
<td>1.10–1.93</td>
<td>2.62</td>
<td>0.01</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>–0.16</td>
<td>0.08</td>
<td>0.85</td>
<td>0.73–0.98</td>
<td>–2.14</td>
<td>0.03</td>
</tr>
<tr>
<td>Negative affect</td>
<td>0.19</td>
<td>0.55</td>
<td>1.186</td>
<td>1.07–1.32</td>
<td>3.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval by race</td>
<td>–0.66</td>
<td>–0.29</td>
<td>0.52</td>
<td>0.29–0.91</td>
<td>–2.28</td>
<td>0.02</td>
</tr>
<tr>
<td>Interval by antidepressant use at index</td>
<td>1.62</td>
<td>0.39</td>
<td>5.05</td>
<td>2.33–10.94</td>
<td>4.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Interval by health index</td>
<td>–0.39</td>
<td>0.19</td>
<td>0.68</td>
<td>0.47–0.97</td>
<td>–2.10</td>
<td>0.04</td>
</tr>
<tr>
<td>Interval by positive affect</td>
<td>–0.38</td>
<td>0.19</td>
<td>0.68</td>
<td>0.47–0.99</td>
<td>–1.99</td>
<td>0.05</td>
</tr>
<tr>
<td>Interval by somatic complaints</td>
<td>0.22</td>
<td>0.09</td>
<td>1.24</td>
<td>1.04–1.50</td>
<td>2.39</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\(^a\) Generalized estimating equations model (df=4, 332) was used to account for repeated measures.
use of the new-generation antidepressant agents is widening, these findings suggest that these agents are being used for an even wider range of symptoms, perhaps for behavioral control of dementing disorders and mild psychiatric symptoms associated with physical illness. In addition, these agents are clearly being used for depressive symptoms that do not meet the criteria for major depressive disorder, given the frequency of use compared to the frequency of major depressive disorder in late life among community-dwelling older adults, which rarely exceeds 5% in survey data (26, 27).

References

Objective: The authors examine national patterns in psychotherapy for older adults with a diagnosis of depression and analyze correlates of psychotherapy use that is consistent with Agency for Health Care Policy and Research guidelines for duration of treatment.

Method: Linked Medicare claims and survey data from the 1992–1999 Medicare Current Beneficiary Survey were used. The data were merged with the Area Resource File to assess the effect of provider-supply influences on psychotherapy treatment. An episode-of-care framework approach was used to analyze psychotherapy use and treatment duration. Multiple logistic regression analysis was used to predict psychotherapy use and its consistency.

Results: The authors identified 2,025 episodes of depression treatment between 1992 and 1999. Overall, psychotherapy was used in 25% (N=474) of the episodes, with 68% of episodes with psychotherapy involving services received only from psychiatrists. (Percentages were weighted for the complex design of the Medicare Current Beneficiary Survey.) Use of psychotherapy was correlated with younger patient age, higher patient educational attainment, and availability of local psychotherapy providers. Among episodes in which psychotherapy was used, only a minority (33%, N=141) involved patients who remained in consistent treatment, defined as extending for at least two-thirds of the episode of depression. Availability of local providers was positively correlated with consistent psychotherapy use. In analyses with adjustment for provider-related factors, patients’ socioeconomic and demographic characteristics did not affect the odds of receiving consistent psychotherapy.

Conclusions: Use of psychotherapy remains uncommon among depressed older adults despite its widely acknowledged efficacy. Some of the disparities in psychotherapy utilization suggest supply-side barriers. Increasing the geographic availability of mental health care providers may be one way of increasing access to psychotherapy for depressed older adults.

There is widespread consensus that geriatric depression is highly prevalent (1–3), adversely affects daily function and quality of life, and contributes to increased health care costs (4), physical decline (5), and even death (6).

Several types of psychotherapy are effective for depression in older adults (7, 8). In a meta-analysis of psychosocial treatments for geriatric depression, the overall effect size for treatment versus no treatment or placebo control was 0.78 (9). Although antidepressant medications are a mainstay of medical treatment for depression and have been shown to be safe and efficacious in older adults (10), patients often prefer psychotherapy (11, 12). Specifically, concern over medication side effects may be heightened when geriatric depression occurs in context of diabetes mellitus (13), cardiac disease (14), and other general medical illnesses. Among patients who are treated with antidepressants, high rates of medication nonadherence constrain clinical effectiveness (15). Older adults may face several barriers to psychotherapy treatment. Although nearly all older adults in the United States have basic health coverage through Medicare, Medicare fee-for-service reimbursement currently covers only 50% of mental health care costs. High copayments may put psychotherapy out of the reach of those in greatest need (16, 17). Often, depressed older adults are unwilling to seek treatment for their symptoms when they have to pay out of pocket for services (11). Primary care physicians may also not refer patients to psychotherapy because of a lack of awareness of its clinical efficacy (18).

Little information exists concerning patterns of psychotherapy use for depression in older adults. Most clinical trials involve participants who are highly selected from clinical settings and typically receive care free of charge. Given the results from patient preference studies (12, 19), psychotherapy remains an important but poorly understood option in the community treatment of depression in older adults. Although the Agency for Health Care Policy and Research guidelines (20) recommend treatment for 4–9 months to reduce the risk of recurrence, the frequency with which depressed older adults receive psychotherapy of this duration remains unknown.

In the study reported here, we examined national patterns in psychotherapy treatment of depression among older adults. The study is based on linked Medicare claims and survey data from the 1992–1999 Medicare Current Beneficiary Survey cost and use files.
Method

Data Sources

The Medicare Current Beneficiary Survey is a continuous survey of a nationally representative sample of aged, disabled, and institutionalized Medicare beneficiaries with information on health status, health care use and expenditures, health insurance coverage, and socioeconomic and demographic characteristics. It has a complex sample design in which sampling is conducted first at the level of primary sampling units (geographically based clusters that represent cities or county groups) and then at the person level. Weights are provided to generalize to the total population enrolled in Medicare for a given year; because data for some respondents are weighted more heavily than those for others to achieve national representativeness, weighted percentages differ somewhat from those based on raw cell sizes. Data from interviews and Medicare claims were linked to examine depression diagnosis and use of psychotherapy among adults ages 65 years and older. The data were augmented through merging with the Area Resource File to provide county-level information (21).

Study Sample

The study sample (N=1,542) was restricted to community-dwelling Medicare beneficiaries who were ages 65 and older, were enrolled in fee-for-service Medicare, and had at least one claim with a diagnosis of depression between 1992 and 1999. Sixty-eight percent of the final sample lived in metropolitan areas, and 73% were female.

Episode of Care

We used an episode-of-care framework to analyze psychotherapy treatment for depression. Agency for Health Care Policy and Research guidelines (20) recommend a minimum of 4–9 months of continuous treatment after a diagnosis of depression. Based on depression diagnosis dates, we defined each episode to be at least 6 months from the date of the initial depression diagnosis. To allow for varying length of individual treatment episodes, after the minimum period of 6 months, lack of a subsequent claim with a depression diagnosis for at least 8 weeks indicated the end of an episode. Length of episodes varied from 6 months to 35 months. This approach is consistent with the identification of fixed and varying length episodes in previous research (21–23). Episodes observed for less than 6 months were excluded from the analysis. Excluded episodes consisted of those in which the individuals died (N=20) or had their initial depression diagnosis within 6 months before the end of the study period (N=496).

Measures

Depression diagnosis was identified through ICD-9-CM/DSM-IV diagnosis codes (296.2, 296.3, 300.4, and 311) recorded in the Medicare claim files. On the basis of our episode-of-care approach, the sampling criteria captured 2,025 episodes.

Use of psychotherapy was identified by using physician’s current procedure terminology (Physician’s Current Procedural Terminology, 4th ed. [CPT-4]) and Health Care Financing Administration Common Procedure Coding System codes in the Medicare physician/supply claims. Psychotherapy was defined broadly, by using codes in the range from 90841 to 90857 for 1992 through 1999. In addition, codes ranging from H5010 to H5025 were included for 1992–1997 data; codes 90875, 90867, and 90880 were for 1993–1999 data; and codes ranging from G0071 to G0094 were included for 1997–1999 data.

Guidelines recommend that treatment for major depression continue for 4–9 months (20). Therefore, we defined consistent psychotherapy use as extending for at least two-thirds of the total months of an episode.

Demographic characteristics included gender, age, race, and living arrangement. Patients were categorized into two groups on the basis of age: 65–74 years, and 75 years and older. Race was classified as white versus nonwhite. Living arrangements were categorized as living alone, living with spouse, and living with others.

Economic characteristics included education, poverty status, pharmacy coverage, and supplemental insurance coverage. On the basis of completed years of schooling, education was categorized as no college or college. “Low income” was defined as the respondent’s personal income, or joint income if married, below 200% of the family-size-adjusted federal poverty level (24). Type of prescription drug insurance coverage was derived from survey responses about monthly drug coverage and included five possible types of private plans, Medicaid, and other public plans. A supplemental insurance variable was created to indicate whether the individual was covered for at least 1 month under any of the following insurance plans: private plans, other public plans, health maintenance organization (HMO) coverage, Medicaid, and Qualified Medicare Beneficiary/Specified Low-Income Medicare Beneficiary coverage.

Environmental variables included characteristics of the health care delivery system, external environment, and community. Person-level area of residence was classified as living in or near “metropolitan” areas versus “nonmetropolitan” areas. Two binary variables were created from Area Resource File data to measure availability of a psychiatrist and a mental health center in the county where the patient lived.

Health care professionals who treated patients during the psychotherapy treatment episodes were categorized by clinical specialty: only psychiatrists, psychiatrists and others, and only others. Other health care professionals included psychologists, social workers, and general practitioners.

Previous studies have suggested that treatment with antidepressant medications increases responsiveness, motivation, and accessibility to psychotherapy (25, 26). We defined medication management visits (CPT-4 code 90862) as those including no more than minimal medical psychotherapy.

Health status variables included self-perceived health and number of comorbid conditions. Both are ordinal variables, and higher values imply lower health status. Health perceptions were classified as 1) excellent or very good, 2) good, and 3) fair or poor. The number of comorbid conditions was measured by respondents’ self-reports about whether they had ever been told by a doctor that they had heart disease, diabetes, cancer, stroke, arthritis, hypertension, emphysema, osteoporosis, or Alzheimer’s disease. The number of comorbid conditions was categorized into three groups: 1) none to one, 2) two to four, and 3) five or more.

Year of diagnosis was included to account for changes in available treatments and practice patterns, with 1992 as the reference year.

Statistical Methods

The chi-square statistic was used to test unadjusted group differences in rates of psychotherapy and their consistency. Multiple logistic regressions were performed to determine the effect of each covariate on the odds of receiving psychotherapy during an episode of depression treatment and, if psychotherapy was used, the odds of receiving consistent psychotherapy.

Given the design of the Medicare Current Beneficiary Survey data and our definition of episode, an individual could be followed for a maximum of 4 years and contribute up to six episodes. Because the Medicare Current Beneficiary Survey assigns weights by calendar year, for episodes that spanned multiple years we used only the baseline weights. In the statistical analysis, we estimated standard errors by using linearization methods that ac-
college (odds ratio=2.72, 95% confidence interval [CI]=
entered. College-educated older adults were more than
income effect became insignificant once education was
County of residence was excluded from the model be-
coverage.
available were more likely to use psychotherapy, as
those living alone or with others, and in metropolitan
tion of the episodes occurred among female beneficiaries,
no college. Overall 25% (N= 474) of the episodes had any use of psychotherapy for de-
(The percentage was weighted for the complex
design of the Medicare Current Beneficiary Survey.) No
significant difference in rate of psychotherapy across years
was found. This finding held when the year of treatment
was entered as a continuous variable (results not shown),
indicating the absence of a time trend.

There were significant differences in rates of psycho-
therapy by age, education, income, living arrangement,
availability of local providers, medication management,
and prescription coverage. A higher proportion of psy-
chotherapy use was found in the younger elderly group
than in the older group, beneficiaries with college than in
those with no college, in higher-income than in low-in-
come beneficiaries, in those living with spouse than in
those living alone or with others, and in metropolitan
than in nonmetropolitan residents. Older adults living in
a county where a psychiatrist or a mental health center
was available were more likely to use psychotherapy, as
were those who ever had medication management visits
during the episodes and those who had prescription drug
coverage.

Findings from the multiple logistic regression con-
firmed the results from the chi-square tests (Table 1).
County of residence was excluded from the model be-
cause of its colinearity with local provider supplies. The
income effect became insignificant once education was
entered. College-educated older adults were more than
twice as likely to receive psychotherapy than those with no
college (odds ratio=2.72, 95% confidence interval [CI]=
1.67–4.43). Older adults who received medication man-
agement during the episode were more than twice as likely
to receive psychotherapy as those who did not (odds ra-
tio=2.65, 95% CI=1.79–3.93). Availability of a psychiatrist or
a mental health center in the county where the patient
lived also increased the odds of receiving psychotherapy
about twofold (availability of a psychiatrist: odds ratio=
2.19, 95% CI=1.79–3.93; availability of a mental health cen-
ter: odds ratio=1.50, 95% CI=1.03–2.21). Prescription drug
coverage also did not significantly affect the likelihood of
using psychotherapy, when other factors were taken into
account.

Consistency of Psychotherapy

Table 2 shows the results for consistency of psychother-
apy. Overall, 33% of the psychotherapy treatment episodes
were consistent with guideline recommendations. Chi-
square tests showed significant differences in consistency
by education attainment, income, county of residence,
provider discipline, and local provider supplies. Older
adults with no college were less likely to receive consistent
psychotherapy, as were nonmetropolitan residents and
low-income persons. Psychiatrists alone provided psycho-
therapy in almost 70% of the total episodes, but the pa-
tients they treated were significantly less likely to receive
consistent psychotherapy than those who received at least
some psychotherapy from other health care professionals.
Beneficiaries living in a county with a mental health center
were significantly more likely to receive consistent psy-
chotherapy than those who did not.

There were few differences in findings between the chi-
square tests and the multiple logistic regressions (Table 2).
With adjustment for other factors, users’ socioeconomic
and demographic characteristics no longer affected the
likelihood of receiving consistent psychotherapy. In con-
trast, provider-supply effects were persistent. Older pa-
tients treated only by psychiatrists were significantly less
likely to receive consistent psychotherapy than those
who were treated by other professionals (odds ratio=0.38, 95% CI=
0.21–0.67), and a combination of psychiatrists and other
professionals increased the likelihood of receiving consis-
tent psychotherapy (odds ratio=2.41, 95% CI=0.99–5.90).
Availability of a mental health center in the county where
the patient lived increased the odds of receiving consistent
psychotherapy about twofold (odds ratio=2.24, 95% CI=
1.30–3.87).

Discussion

Despite substantial empirical evidence supporting the
efficacy of psychotherapy for depression in older adults,
our study showed that only a minority of depressed elderly
patients received psychotherapy. This finding is consistent
with findings in previous studies (28). Overall, psychother-
apy use remained stable at this low level between 1992 and
1999, although antidepressant use in this population in-
creased during this period (29).
### TABLE 1. Characteristics Associated With Treatment Episodes for Elderly Medicare Beneficiaries With a Diagnosis of Depression, 1992–1999a

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All Treatment Episodes With a Diagnosis of Depression (N=2,025)</th>
<th>Treatment Episodes With Use of Psychotherapy (N=474) [25.2%]</th>
<th>Multiple Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td><strong>Beneficiary characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1,499</td>
<td>73.3</td>
<td>341</td>
</tr>
<tr>
<td>Male</td>
<td>526</td>
<td>26.7</td>
<td>133</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1,777</td>
<td>89.6</td>
<td>418</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>248</td>
<td>10.4</td>
<td>56</td>
</tr>
<tr>
<td>Age (years)b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–74</td>
<td>842</td>
<td>49.4</td>
<td>243</td>
</tr>
<tr>
<td>≥75</td>
<td>1,183</td>
<td>50.6</td>
<td>231</td>
</tr>
<tr>
<td>Educationb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>166</td>
<td>9.1</td>
<td>68</td>
</tr>
<tr>
<td>No college</td>
<td>1,845</td>
<td>90.9</td>
<td>402</td>
</tr>
<tr>
<td>Income-to-needs ratiob</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;200%</td>
<td>1,252</td>
<td>57.9</td>
<td>256</td>
</tr>
<tr>
<td>≥200%</td>
<td>773</td>
<td>42.1</td>
<td>218</td>
</tr>
<tr>
<td>Living arrangementb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>779</td>
<td>36.6</td>
<td>151</td>
</tr>
<tr>
<td>With spouse</td>
<td>849</td>
<td>45.2</td>
<td>235</td>
</tr>
<tr>
<td>With other(s)</td>
<td>397</td>
<td>18.2</td>
<td>88</td>
</tr>
<tr>
<td>County of residenceb,e</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>1,405</td>
<td>71.8</td>
<td>396</td>
</tr>
<tr>
<td>Nonmetropolitan</td>
<td>620</td>
<td>28.2</td>
<td>78</td>
</tr>
<tr>
<td>Supplemental insurance coveragef</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1,930</td>
<td>95.5</td>
<td>457</td>
</tr>
<tr>
<td>No</td>
<td>95</td>
<td>4.5</td>
<td>17</td>
</tr>
<tr>
<td>Prescription coverageb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1,455</td>
<td>72.1</td>
<td>364</td>
</tr>
<tr>
<td>No</td>
<td>570</td>
<td>27.9</td>
<td>110</td>
</tr>
<tr>
<td>Received a medication management visitb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>195</td>
<td>9.9</td>
<td>93</td>
</tr>
<tr>
<td>No</td>
<td>1,830</td>
<td>90.1</td>
<td>381</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very good/excellent</td>
<td>517</td>
<td>26.1</td>
<td>149</td>
</tr>
<tr>
<td>Good</td>
<td>598</td>
<td>31.0</td>
<td>135</td>
</tr>
<tr>
<td>Fair/poor</td>
<td>901</td>
<td>42.9</td>
<td>185</td>
</tr>
<tr>
<td>Number of comorbid conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 and 1</td>
<td>350</td>
<td>18.7</td>
<td>96</td>
</tr>
<tr>
<td>≥2–4</td>
<td>1,344</td>
<td>65.8</td>
<td>304</td>
</tr>
<tr>
<td>≥5</td>
<td>330</td>
<td>15.5</td>
<td>73</td>
</tr>
<tr>
<td>Year of diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>194</td>
<td>9.5</td>
<td>48</td>
</tr>
<tr>
<td>1993</td>
<td>255</td>
<td>12.6</td>
<td>65</td>
</tr>
<tr>
<td>1994</td>
<td>231</td>
<td>10.9</td>
<td>53</td>
</tr>
<tr>
<td>1995</td>
<td>205</td>
<td>10.4</td>
<td>51</td>
</tr>
<tr>
<td>1996</td>
<td>278</td>
<td>14.3</td>
<td>65</td>
</tr>
<tr>
<td>1997</td>
<td>292</td>
<td>14.9</td>
<td>59</td>
</tr>
<tr>
<td>1998</td>
<td>343</td>
<td>16.4</td>
<td>82</td>
</tr>
<tr>
<td>1999</td>
<td>227</td>
<td>10.6</td>
<td>51</td>
</tr>
<tr>
<td>Availability of psychotherapy providers in the beneficiary's county of residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychiatristb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1,615</td>
<td>81.4</td>
<td>428</td>
</tr>
<tr>
<td>No</td>
<td>410</td>
<td>18.6</td>
<td>46</td>
</tr>
<tr>
<td>Mental health centerb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1,010</td>
<td>51.1</td>
<td>296</td>
</tr>
<tr>
<td>No</td>
<td>1,015</td>
<td>48.9</td>
<td>178</td>
</tr>
</tbody>
</table>

a Data were based on episodes of depression constructed from the claims of a nationally representative sample of elderly Medicare beneficiaries with a diagnosis of depression between calendar years 1992 and 1999 who were age 65 years and older, enrolled in fee-for-service Medicare throughout the episode, and living in the community. Percentages are weighted for the complex design of the Medicare Current Beneficiary Survey.

b Significant difference in rate of psychotherapy treatment (p<0.05, chi-square test).

c Significant effect (p<0.05) on psychotherapy treatment in the multiple logistic regression.

d Significant effect (p<0.10) on psychotherapy treatment in the multiple logistic regression.

e Not included in multiple logistic regression analysis because of colinearity with availability of a psychiatrist and mental health center in the county of residence.

f Significant difference in rate of psychotherapy treatment (p<0.10, chi-square test).
Although education and income are indicators of socioeconomic status, in the multiple logistic regression only education was related to psychotherapy use. College-educated older adults were more likely to use psychotherapy than were those with no college. Previous research with groups of nonelderly subjects has found that acceptance of psychotherapy is directly related to level of education (30). Psychiatrists initiating treatment of depression for older patients should bear in mind that their patients who have not attended college may tend to be less accepting of psychotherapy than those with higher educational attainment. The selection of treatment modality should be influenced by the patient’s preferences as well as clinical considerations such as symptom severity, the presence of interpersonal difficulties, or a comorbid axis II disorder (31). With older patients who have less formal education, it may be especially impor-
tiant for psychiatrists to discuss how psychotherapy works and remain responsive to patients’ concerns or misconceptions regarding this treatment.

In our study, environmental factors significantly affected use and consistency of psychotherapy. Our findings suggest that older Americans face significant barriers in access to psychotherapy because of limited local availability of qualified providers (21).

Provider discipline played an important role in affecting the likelihood of receiving consistent psychotherapy. Patients who received psychotherapy only from psychiatrists were less likely to receive psychotherapy for the period of time recommended by Agency for Health Care Policy and Research guidelines. Because a great majority of the medication management visits were provided by psychiatrists, psychotherapy may be used by psychiatrists to supplement pharmacotherapy, rather than as a primary treatment. Costs may also contribute to the observed psychotherapy utilization pattern. Given that a majority of the patients had prescription drug coverage and that Medicare covers 80% of medication management visit payments (32), patients who are treated with antidepressant medications incur lower copayments than those who receive psychotherapy. Psychologists and social workers also have substantially lower fee schedules than psychiatrists (33).

The risk of early termination may be greater for patients under the care of psychiatrists in solo practice than for patients treated by psychiatrists in mental health centers or multidisciplinary groups. Psychiatrists who have institutionalized referral relationships with nonpsychiatrist psychotherapists may have an advantage in maintaining their older patients through the continuation and maintenance phases of psychotherapy for depression. Therefore, higher financial barriers to treatment, differences in rates of antidepressant treatment, and variation in psychotherapy orientations may help to explain the low rates of consistent psychotherapy among patients who received psychotherapy only from psychiatrists.

The study has some important limitations. Coding biases are likely to affect the sensitivity, more than the specificity, of diagnosis (34–36). Therefore, our study may have underestimated the overall number of depressive episodes. The sample includes only non-HMO and non-institution-dwelling Medicare beneficiaries, and so the findings cannot be generalized to HMO and institution-dwelling populations. The Medicare Current Beneficiary Survey data provide only event-level information on prescription drug use, and thus we were unable to determine precise dates of antidepressant use. Most important, Medicare Current Beneficiary Survey data provide no information on the types of psychotherapy being provided (e.g., cognitive behavior, interpersonal, nonspecific supportive) or on their clinical effectiveness.

Despite these limitations, our study provides important information on use of psychotherapy among elderly patients with a diagnosis of depression. The findings confirm that use of psychotherapy is influenced by a host of factors at the patient, provider, and health care system levels.

RECEIVED March 19, 2003; revision received Feb 10, 2004; accepted June 1, 2004. From the Institute for Health, Health Care Policy, and Aging Research, Rutgers University; and the Department of Psychiatry, Columbia University, New York State Psychiatric Institute. Address correspondence and reprint requests to Dr. Wei, Institute for Health, Health Care Policy, and Aging Research, Rutgers University, 30 College Ave., New Brunswick, NJ 08901; wwh@rci.rutgers.edu (e-mail).

Supported by NIMH grants MH-60831 and MH-43450, grants HS-11825 and HS-09566 from the Agency for Healthcare Research and Quality, and grant AG-15166 from the National Institute on Aging. The findings and opinions reported here are those of the authors and do not necessarily represent the views of any other individuals or organizations.
The authors thank Rizie Kumar, M.A., and Vatsala Karwe, Ph.D., for programming assistance, Michelle Kennedy for research assistance, and the Centers for Medicaid and Medicare Services for making the data available.

References


http://ajp.psychiatryonline.org 717

Am J Psychiatry 162:4, April 2005
Remission in Depressed Geriatric Primary Care Patients: A Report From the PROSPECT Study

George S. Alexopoulos, M.D.
Ira R. Katz, M.D.
Martha L. Bruce, Ph.D.
Moonseong Heo, Ph.D.
Thomas Ten Have, Ph.D.
Patrick Raue, Ph.D.
Hillary R. Bogner, M.D.
Herbert C. Schulberg, Ph.D.
Benoit H. Mulsant, M.D.
Charles F. Reynolds III, M.D.
The PROSPECT Group

Objective: This study compared time to first remission for elderly depressed patients in primary care for practices that implemented a care management model versus those providing usual care. In addition, it sought to identify risk factors for nonremission that could guide treatment planning and referral to care managers or specialists.

Method: Prevention of Suicide in Primary Care Elderly: Collaborative Trial (PROSPECT) data were analyzed. Participants were older patients (≥60 years) selected following screening of 9,072 randomly identified primary care patients. The present analysis examined patients with major depression and a 24-item Hamilton Depression Rating Scale score of 18 or greater who were followed for at least 4 months (N=215). Primary care practices were randomly assigned to offer the PROSPECT intervention or usual care. The intervention consisted of services of trained care managers, who offered algorithm-based recommendations to physicians and helped patients with treatment adherence over 18 months.

Results: First remission occurred earlier and was more common among patients receiving the intervention than among those receiving usual care. For all patients, limitations in physical and emotional functions predicted poor remission rate. Patients experiencing hopelessness were more likely to achieve remission if treated in intervention practices. Similarly, the intervention was more effective in patients with low baseline anxiety.

Conclusions: Longitudinal assessment of depression, hopelessness, anxiety, and physical and emotional functional limitations in depressed older primary care patients is critical. Patients with prominent symptoms or impairment in these areas may be candidates for care management or mental health care, since they are at risk for remaining depressed and disabled.

A critical goal in the care of depression is the attainment of remission, defined as an almost asymptomatic state. Patients displaying residual depressive symptoms have functional impairment, compromised quality of life, and high utilization of health care services (1). Moreover, remission is a stable state with a lower risk for relapse than depression improvement, which leaves the patient with residual symptoms (2). While remission is desirable, clinical trials have shown that only a little over one-third of patients treated with antidepressants achieve this clinical state (3). Time to achieve remission has been the focus of studies of psychiatric populations (4), since persisting depression increases suffering, disability, and suicide risk. Remission may be more difficult to achieve in depressed older patients in primary care because of the clinical complexity of this population and the limitations of the primary care setting.

Most depressed older adults are treated by primary care physicians (5). Over 80% of depressed primary care patients prefer to be treated by their primary care physician (6). These preferences might be even stronger among older patients concerned about stigma (7).

Geriatis depression frequently remains inadequately treated in primary care settings. Previous studies have documented that approximately 41% of depressed primary care patients received no antidepressant treatment regardless of age and medical comorbidity (8). While antidepressant prescriptions are rising in primary care practices (9), antidepressants continue to be used at insufficient dosages and for an inadequate length of time (10). Poor treatment adherence by patients further compromises the care of depressed primary care patients (11).

Several health services models have sought to improve the treatment of depression in primary care settings. Training primary care physicians (12), introducing computer-driven decision support (13), and integrating the management of depression with the care of other medical illnesses (14, 15) have had varying success. However, collaborative care of primary care physicians with on-site mental health specialists has enhanced quality of care and improved the outcomes of depression (16).

Recently, the Prevention of Suicide in Primary Care Elderly: Collaborative Trial (PROSPECT) compared a primary care-based intervention with usual care in subjects with geriatric depression (17). The intervention relied on
Practices and Participants

Method

The PROSPECT study compared the outcomes of depressed elderly primary care patients in practices that implemented an intervention based on care management versus practices that offered usual care.

Practices and Participants

Twenty practices participated from three regions: greater New York City, Philadelphia, and Pittsburgh. They were paired within each region by setting (urban, suburban, rural), academic affiliation, size, and racial distribution of patients. Within pairs, practices were randomly assigned to provide either the PROSPECT intervention or usual care. Randomization by practice was chosen in order to reduce “contamination” of usual care by that offered by care managers.

A two-stage sampling generated a representative sample of primary care patients with DSM-IV major or minor depression persisting for at least 1 month. The subject flow has been reported elsewhere (17). Briefly, an age-stratified (60–74 years, ≥75 years) random sample was screened by telephone for depressive symptoms with the Center for Epidemiologic Studies Depression Scale (CES-D Scale) (20) following oral consent. All patients with a CES-D Scale score above 20, those with a history of depression, and a 5% random sample of patients with lower scores were invited to participate. After signing institutional review board-approved consent, patients were interviewed in person according to a protocol (21).

Remission was one of the primary outcomes of the PROSPECT study and was defined as the first occurrence of achieving a Hamilton depression scale score <10. This definition is often used in studies focusing on time to remission of geriatric depression (4) and reflects the concern that elderly patients who no longer have depressive symptoms may still receive Hamilton depression scale score points because of somatic symptoms originating from medical illnesses. Secondary analysis used a second, stricter definition (Hamilton depression scale score <7).

PROSPECT Intervention

The intervention (described elsewhere [17]) was implemented by 15 “care managers,” who used operationalized guidelines (18) to provide “on-time and on-target” recommendations to primary care physicians and help patients with treatment adherence. The PROSPECT initial report demonstrated that patients receiving the care management intervention had less severe depressive symptoms and greater remission rates at 4, 8, and 12 months than patients receiving usual care. The initial report included patients with either major or minor depression and a wide range of depression severity (i.e., score >10 on the 24-item Hamilton Depression Rating Scale [19]). The report also did not distinguish patients who achieved first remission from those with a fluctuating course of depression.

The present analysis focuses on the time to first remission for primary care patients experiencing major depression and significant depressive symptom severity. Two hypotheses were tested. The first postulated that patients in primary care practices implementing the PROSPECT intervention would achieve first remission more rapidly than would patients in practices providing usual care. The second hypothesis was that patients with a complex clinical picture, i.e., those with hopelessness and disability, would be less likely to achieve remission. Finally, this analysis sought to identify patients with risk factors for nonremission for whom the intervention was more effective than usual care.

Method

The PROSPECT study compared the outcomes of depressed elderly primary care patients in practices that implemented an intervention based on care management versus practices that offered usual care.

Practices and Participants

Twenty practices participated from three regions: greater New York City, Philadelphia, and Pittsburgh. They were paired within each region by setting (urban, suburban, rural), academic affiliation, size, and racial distribution of patients. Within pairs, practices were randomly assigned to provide either the PROSPECT intervention or usual care. Randomization by practice was chosen in order to reduce “contamination” of usual care by that offered by care managers.

A two-stage sampling generated a representative sample of primary care patients with DSM-IV major or minor depression persisting for at least 1 month. The subject flow has been reported elsewhere (17). Briefly, an age-stratified (60–74 years, ≥75 years) random sample was screened by telephone for depressive symptoms with the Center for Epidemiologic Studies Depression Scale (CES-D Scale) (20) following oral consent. All patients with a CES-D Scale score above 20, those with a history of depression, and a 5% random sample of patients with lower scores were invited to participate. After signing institutional review board-approved consent, patients were interviewed in person according to a protocol (21).

Remission was one of the primary outcomes of the PROSPECT study and was defined as the first occurrence of achieving a Hamilton depression scale score <10. This definition is often used in studies focusing on time to remission of geriatric depression (4) and reflects the concern that elderly patients who no longer have depressive symptoms may still receive Hamilton depression scale score points because of somatic symptoms originating from medical illnesses. Secondary analysis used a second, stricter definition (Hamilton depression scale score <7).

**Table 1. Remission Probability Among Elderly Depressed Patients in Primary Care Practices Implementing a Care Management Intervention Versus Practices Providing Usual Care**

<table>
<thead>
<tr>
<th>Remission Criterion and Follow-Up Assessment</th>
<th>Intervention(^a)</th>
<th>Usual Care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients Achieving Remission</td>
<td>Cumulative Probability of Remission</td>
</tr>
<tr>
<td>Hamilton depression scale score &lt;10 (^b)</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4 months</td>
<td>119</td>
<td>39</td>
</tr>
<tr>
<td>8 months</td>
<td>64</td>
<td>10</td>
</tr>
<tr>
<td>12 months</td>
<td>43</td>
<td>8</td>
</tr>
<tr>
<td>18 months</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>Hamilton depression scale score &lt;7 (^d)</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4 months</td>
<td>119</td>
<td>25</td>
</tr>
<tr>
<td>8 months</td>
<td>76</td>
<td>13</td>
</tr>
<tr>
<td>12 months</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>18 months</td>
<td>39</td>
<td>5</td>
</tr>
</tbody>
</table>

\(^a\) Trained care managers offering algorithm-based recommendations to primary care physicians and helping patients with treatment adherence.

\(^b\) Likelihood of remission among those who were depressed at the previous follow-up assessment and thus candidates for remission.

\(^c\) Log-rank \(\chi^2=4.06, \text{df}=1, p<0.05\).

\(^d\) Log-rank \(\chi^2=1.07, \text{df}=1, p=0.30\).
Depression diagnoses were assigned by trained research assistants after administration of the Structured Clinical Interview for DSM-IV (24) and review of ratings by study psychiatrists. Severity of depression was assessed with the 24-item Hamilton depression scale (19). Suicide ideation was rated with the Scale for Suicide Ideation (25). Interrater reliability and rater drift were monitored throughout the study and have been reported elsewhere (17). Anxiety was quantified with the Clinical Anxiety Scale (26). Hopelessness was assessed with the Beck Hopelessness Scale (27). Cognitive impairment was rated with the Mini-Mental State Examination (28). The Medical Outcomes Study 12-item Short-Form Health Survey (29) was used to assess limitations in functioning due to physical problems (physical component summary) and emotional difficulties (emotional component summary). The intensity of antidepressant pharmacotherapy at entry was quantified by using the Composite Antidepressant Treatment Intensity Scale (30), a 5-point scale modified to include recently introduced antidepressants.

Participants were followed for 18 months; they had telephone assessments at 4, 8, and 18 months and an in-person interview 12 months after entry. Although desirable, it was not feasible to blind the research assistants to the treatment assignments (intervention versus usual care) of the practices.

**Data Analysis**

Baseline demographic and clinical characteristics were compared by using t tests for continuous variables and chi-square analyses for binary variables. Kaplan-Meier survival analysis was employed to test the significance of differences in the study outcome, time until first remission of depressive symptoms. For the identification of predictors to remission we used mixed effects logistic regression models to fit the outcome with adjustment for antidepressants or psychotherapy. Covering the cost of citalopram and interpersonal psychotherapy limited the study of cost as a barrier to treatment.

**Usual Care**

Physicians of "usual care" practices were notified in writing of the patients' depression diagnosis and contacted by the investigators when the study's Risk Management Guideline (23) indicated suicide risk in individual patients. Physicians received a videotape and printed material on geriatric depression and treatment guidelines. These measures reduced barriers to recognition of depression.

**Systematic Assessment**

Depression diagnoses were assigned by trained research assistants after administration of the Structured Clinical Interview for DSM-IV (24) and review of ratings by study psychiatrists. Severity of depression was assessed with the 24-item Hamilton depression scale (19). Suicide ideation was rated with the Scale for Suicide Ideation (25). Interrater reliability and rater drift were monitored throughout the study and have been reported elsewhere (17). Anxiety was quantified with the Clinical Anxiety Scale (26). Hopelessness was assessed with the Beck Hopelessness Scale (27). Cognitive impairment was rated with the Mini-Mental State Examination (28). The Medical Outcomes Study 12-item Short-Form Health Survey (29) was used to assess limitations in functioning due to physical problems (physical component summary) and emotional difficulties (emotional component summary). The intensity of antidepressant pharmacotherapy at entry was quantified by using the Composite Antidepressant Treatment Intensity Scale (30), a 5-point scale modified to include recently introduced antidepressants.

Participants were followed for 18 months; they had telephone assessments at 4, 8, and 18 months and an in-person interview 12 months after entry. Although desirable, it was not feasible to blind the research assistants to the treatment assignments (intervention versus usual care) of the practices.
were excluded because of missed follow-up assessments. Among patients included in this analysis and those who ability, or intensity of antidepressant treatment at baseline severity of depression, anxiety, cognitive impairment, dis- significant differences in age, gender, race, education, se-

When remission was defined as Hamilton depression scale score $<10$, patients treated by practices implement-
ing the care management intervention had a higher cu-
mulative probability of remission (probability of achieving remission at each follow-up point) than practices offering usual care (Table 1, Figure 1). Similarly, the hazard to remission (likelihood of remission among those were de-

The main effect of the symptom was not significant.

The flow of participants has been reported elsewhere (17). Briefly, 16,708 older patients were sampled, and 9,072 were screened for depression with the CES-D Scale. Of the screened patients, 1,888 were invited to enroll in the study, and 1,238 agreed to a baseline interview. Among them, 267 met criteria for major depression and had a Hamilton depression scale score $\geq18$ at entry. This analysis included only participants (N=215) who at least had been evaluated at the 4-month follow-up visit. Outcomes during the follow-up period were either 1) first instance of remission identified before any missed assessment session; or 2) no instance of remission identified before the first missed appointment. This strategy enabled determination of the most accurate time to remission within the constraints of infrequent follow-up assessments. Time to remission was defined as the time from entry into the study until initially meeting criteria for remission. There were no statistically significant differences in age, gender, race, education, severity of depression, anxiety, cognitive impairment, dis-

The most parsimonious multivariate model indicated that patients with limitations in physical and emotional aspects of functioning, hopelessness, and anxiety were less likely to achieve remission regardless of treatment as-

The occurrence of remission differed across the three regions (F=7.03, df=2, 471, p<0.01), but there was no site-
experiment effect was principally related to the differential efficacy of
The principal finding of this analysis is that the PROSPECT intervention was more effective than usual care in promoting remission of depression in elderly primary care patients, especially when remission was defined as a Hamilton depression scale score lower than 10. Differences in remission rates between intervention and usual care were most pronounced among depressed elders experiencing hopelessness. When remission was defined according to stricter criteria, i.e., Hamilton depression scale score less than 7, the intervention group had numerically higher rates of remission during the early phases of treatment but, when the whole 18-month period was taken into consideration, differences in remission rates did not reach significance. Regardless of the remission criterion, depressed patients with comorbid anxiety disorders, hopelessness, and limitations in physical and emotional functioning were associated with low remission rates in primary care elderly patients receiving either the intervention or usual care.

A strength of this study is the use of random sampling and screening that may have resulted in a representative sample of depressed older primary care patients. Therefore, its findings may be relevant to clinical practice. Limitations of the study include the lack of blinding of raters, the infrequent follow-up, the lack of information on discrete medical problems of participants, and the lack of information about specific antidepressant treatments received by each group during the 18 months of the study. While more frequent follow-up assessments would have been desirable, concerns about participant burden and cost led us to select few, yet clinically meaningful, follow-up times. This analysis used disability resulting from physical problems as a proxy of medical burden (the Short-Form Health Survey physical component summary). Data on specific medical disorders and their impact on remission need further attention. Finally, treatment changed frequently both in the intervention and the usual care practices. Future analyses may compare the impact on remission of selected treatments at specific times. Nonetheless, focusing on predictors of the first remission occurrence serves to identify the clinical profile of elderly primary care patients likely to remain depressed and in need of close follow-up and perhaps referral to mental health specialists.

The favorable remission rates of the PROSPECT intervention on depression is consistent with evidence that interventions aimed at changing primary care practice can improve the quality of depression care in mixed-age (14) and elderly (15) patients. Despite differences in design, measurements, and type of intervention, both the PROSPECT and the Improving Mood-Promoting Access to Collaborative Treatment studies (the latter being another study of depressed primary care elders) demonstrated differences in remission rates favoring the intervention over usual care (15). These findings underscore the value of such interventions for resolving late-life depression in primary care patients.

Remission occurred earlier for patients in practices providing the intervention than for those in practices providing usual care. By 8 months, 43% of patients receiving the care management intervention had achieved remission (Hamilton depression scale score <10) compared with 28% of patients receiving usual care. The remission rate with usual care eventually reached the level achieved with the
intervention, perhaps reflecting the fact that some depressive episodes eventually subside and those that persist or worsen receive additional attention. These observations suggest that in at least some patients, the use of trained care managers can accelerate remission and perhaps reduce suffering, disability, and family disruption several months earlier than usual care.

Compromised physical and emotional function predicted low remission rates in patients treated by either the intervention or usual care. There is evidence, however, that antidepressant treatment and reduction of depressive symptoms can improve the functional status of elderly patients (35). These observations suggest that older primary care patients with major depression and physical and emotional function limitations need aggressive antidepressant treatment and perhaps referral to mental health professionals if their symptoms persist.

Depressed elderly primary care patients experiencing hopelessness were more likely to benefit from the PROSPECT intervention than usual care. Hopelessness is a set of beliefs that influence how a person interprets information and behaves. Hopeless thoughts can be chronic and persistent in some individuals and activated during depression in others (36). Hopelessness has a strong association with suicidal ideation and behavior in younger adults (37). A similar relationship between hopelessness and suicidal ideation was demonstrated in institutionalized elderly patients and was dependent upon the level of depression (38). Hopelessness was associated with suicidal ideation in patients with severe depression, but there was no significant relationship between hopelessness and suicidal ideation in patients with mild depression. These observations suggest that elderly primary care patients experiencing hopelessness require special clinical attention—and perhaps a referral to care managers—since these patients have a low likelihood for remission under usual care and may even be at increased risk for suicide, especially in the presence of a severe depression. The provision of care management is particularly feasible in large medical practices such as HMOs, which often include behavioral specialists on their staff.

Anxiety adversely influenced the rate of remission in patients of the intervention practices, whereas it had a non-significant effect in patients receiving usual care. Specifically, the intervention was more effective than usual care in patients with low anxiety but added little benefit for patients with higher anxiety severity. Patients with comorbid anxiety and depression frequently exhibit more severe symptoms overall, have a more protracted course of illness, and experience less positive treatment outcomes (39). Therefore, care management may be insufficient treatment for some anxious depressed elderly primary care patients, and their physicians should recognize this clinical pattern as one of several warranting referral to a psychiatrist.

Primary care occupies a strategic position in the management of late-life depression. This study has documented that remission of late-life depression occurs in a large percentage of primary care patients. Collaborative care by trained care managers and primary care physicians leads to earlier remission than usual care and can reduce suffering and disability. As many of the care managers’ services are reimbursable under the existing Medicare codes, referral to appropriately trained care managers is feasible. Priority for such services may be given to depressed elders experiencing hopelessness, since usual care is less likely to be helpful. Longitudinal assessment of anxiety and limitations in physical and emotional functioning of depressed older primary care patients is critical. Some patients with prominent impairment in these areas may not achieve remission when treated in primary care settings and may require mental health consultation and care.

References


Received June 2, 2004; revision received July 15, 2004; accepted July 23, 2004. From the Department of Psychiatry, Weill Medical College of Cornell University, New York; the Departments of Psychiatry and Medicine, University of Pennsylvania, Philadelphia; the Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh; and the Geriatric Research, Education and Clinical Center, VA Pittsburgh Health Care System, Pittsburgh. Address correspondence and reprint requests to Dr. Alexopoulos, Cornell Medical Center-Westchester Division, 21 Bloomingdale Rd., White Plains, NY 10605; gsalex@med.cornell.edu (e-mail).

This work was supported by grants from NIMH (R01 MH-59366, R01 MH-059380, R037 MH-51842, R01 MH-59318, P30 MH-68638, P30 MH-52247, and P30 MH-866270), the Hartford Foundation, and Forest Pharmaceuticals, Inc.
REMISSION IN DEPRESSED ELDERLY


Critical Periods of Suicide Risk in Huntington’s Disease

Jane S. Paulsen, Ph.D.
Karin Ferneyhough Hoth, M.A.
Carissa Nehl, B.S.
Laura Stierman, B.S.
The Huntington Study Group

Contrary to popular belief, receiving a diagnosis of a devastating fatal disease does not exacerbate, and may even alleviate, the risk of suicide. Suicidal ideation was examined in 4,171 individuals in the Huntington Study Group database. Participants were grouped according to a standardized neurological examination from 0 (i.e., normal examination) to 3 (definite Huntington’s disease). Patients with an unequivocal diagnosis of Huntington’s disease were further divided by stage of disease, from stage 1 (early) to stage 5 (end stage). Findings showed that the frequency of suicidal ideation doubled from 9.1% in at-risk persons with a normal neurological examination to 19.8% in at-risk persons with soft neurological signs and increased to 23.5% in persons with “possible Huntington’s disease.” In persons with a diagnosis of Huntington’s disease, 16.7% had suicidal ideation in stage 1, and 21.6% had suicidal ideation in stage 2, whereas the proportion of Huntington’s disease patients with suicidal ideation diminished thereafter. Findings suggest two critical periods for increased risk of suicide in Huntington’s disease. The first critical period is immediately before receiving a formal diagnosis of Huntington’s disease, and the second is in stage 2 of the disease, when independence diminishes. Although the underlying mechanisms of suicidal ideation in Huntington’s disease are poorly understood, it is critical for health care providers to be aware of periods during which patients may be at an increased risk.

Huntington’s disease is an autosomal-dominant neurodegenerative disease that is characterized by a triad of symptoms, including cognitive disturbance, motor abnormalities, and psychiatric features. The disease results from a trinucleotide CAG expansion on chromosome 4. The prevalence of Huntington’s disease is approximately 7–10 per 100,000 individuals (1, 2). Disease onset typically occurs between the ages of 35 and 44 years, with an average survival of 20 years. There is currently no cure for the disease, although treatments to slow the progression of symptoms are currently being developed.

It is well established that psychiatric symptoms are common in Huntington’s disease (3–5). In fact, completed suicide has been reported to be as high as 13% in Huntington’s disease (3), reflecting a seven- to 12-fold increase from the rate in the general population (6, 7). Comparison with other research suggests that suicide rates in Huntington’s disease remain higher than those found in other medical and neurodegenerative diseases (8, 9).

Although it is considered mandatory to inquire about suicidal ideation in psychiatric consultations, it is seldom part of the traditional medical or neurological assessment. Most research on suicide risk in Huntington’s disease has focused on persons undergoing genetic testing (7, 10, 11). Less attention has been devoted to suicidal ideation over the stages of presymptomatic and diagnosed Huntington’s disease. In developing suicide prevention practice guidelines for Huntington’s disease, it is imperative to know what factors may involve an increased suicide risk and, if possible, when in the course of the disease the risk is greatest. In the studies published to date, controversy exists about whether suicide risk is elevated in the earliest (12) or the later stages of Huntington’s disease (7).

The purpose of the present study was to examine suicidal ideation in individuals at risk for and diagnosed with Huntington’s disease. A large number of participants from the Huntington Study Group were evaluated with the Unified Huntington’s Disease Rating Scale (13) and grouped according to ratings obtained from its standardized neurological examination and total functional capacity rating scales. The proportion of persons in each group who endorsed suicidal ideation was compared among groups. The informed consent from the Unified Huntington’s Disease Rating Scale was approved by each institution’s ethics review board.

Method

Procedure

The participants were evaluated with the Unified Huntington’s Disease Rating Scale (13) by Huntington Study Group members from 43 sites in North America, Australia, and Europe. All participants were grouped according to a standardized neurological examination administered by an experienced movement disorder neurologist. The diagnosis of Huntington’s disease was based upon the findings obtained from the motor section of the Unified Huntington’s Disease Rating Scale, and diagnostic confidence was reported based upon the following scale: 0=nonspecific motor abnormalities or soft signs, 1=spastic motor abnormalities that indicated possible Hunting-
Huntington’s disease, and 3-unequivocal motor abnormalities that indicated definite Huntington’s disease. Good reliability has been demonstrated for the diagnosis of Huntington’s disease with the Unified Huntington’s Disease Rating Scale (14). Since the participant group with a definite diagnosis of Huntington’s disease was large (i.e., 2,688 individuals) and Huntington’s disease is often characterized by stage of illness, this group was further grouped by stage of disease (stages 1 to 5) with a measure of total functional capacity (15) from the Unified Huntington’s Disease Rating Scale. The percentages of individuals in each group who acknowledged suicidal ideation upon questioning were compared.

**Participants**

The total study group consisted of 4,171 individuals. There were 1,483 persons not diagnosed with Huntington’s disease but considered “at risk” for the disease by virtue of having a parent with Huntington’s disease. Of these, 712 were classified as having no motor symptoms of Huntington’s disease (i.e., a score of 0 for motor evaluation on the Unified Huntington’s Disease Rating Scale). Three hundred sixty-three people showed minor soft signs (score of 1), and 408 showed motor symptoms that indicated possible Huntington’s disease (score of 2). Two thousand, six hundred eighty-eight people received a score of 3 on the motor evaluation, indicating that they had been diagnosed with definite Huntington’s disease. Of these individuals, 2,637 (98%) also had complete data regarding stage of Huntington’s disease based upon functional capacity scores (15). Six hundred eleven (23%) of these individuals were in stage 1, 923 (34%) were in stage 2, 724 (27%) were in stage 3, 277 (10%) were in stage 4, and 102 (4%) were in the final stage (stage 5) of the disease.

Demographic characteristics for the participant groups are presented in Table 1. Briefly, the participants’ ages ranged from 20 to 90, with an average age of 47 years (SD=13). Not surprisingly, increasing age was associated with greater deficits on the neurological examination. As shown in Table 1, the participants with definite Huntington’s disease were, on average, older than the at-risk participants. The average level of education of the study group was approximately 13 years (SD=3). In general, the study group was more than 90% Caucasian, approximately 46% were men, and the majority of the participants were right-handed.

**Measures and Group Division**

The Unified Huntington’s Disease Rating Scale is a standardized clinical rating scale that assesses four components of Huntington’s disease: motor function, functional capacity, cognitive ability, and psychiatric symptoms. The motor section of the Unified Huntington’s Disease Rating Scale assesses motor features of Huntington’s disease with standardized ratings of oculomotor function, dysarthria, chorea, dystonia, gait, and postural stability.

The Unified Huntington’s Disease Rating Scale’s functional assessment consists of the Huntington’s disease functional capacity scale, the independence scale, and the functional checklist (15). Higher scores on the functional capacity scales are indicative of better functioning.

The cognitive section of the Unified Huntington’s Disease Rating Scale includes the letter fluency test (16), the Symbol Digit Modalities Test (17), and the Stroop Color and Word Test (18). Higher scores on these measures are indicative of better performance.

The behavioral section of the Unified Huntington’s Disease Rating Scale includes standardized interview questions regarding the frequency and severity of psychiatric symptoms (i.e., depressed mood, low self-esteem, anxiety, suicidal thoughts, disruptive or aggressive behavior, irritable behavior, perseverative/obsessional thinking, compulsive behavior, delusions, hallucinations, and apathy). During the administration of the Unified Huntington’s Disease Rating Scale, the frequency of a behavioral symptom is initially rated; if the symptom is present, a rating of the severity of the symptom is made. In the current study, endorsement of suicidal ideation (presence or absence) was used to compare the proportion of each subgroup with suicidal thoughts. Next, the severity of suicidal thoughts (a rating of 1–4) was used to quantify the severity of suicidal thoughts in the persons endorsing these. Other Unified Huntington’s Disease Rating Scale measures used included diagnosis based upon the motor section of the Unified Huntington’s Disease Rating Scale and total score on the Huntington’s disease functional capacity scale.

**Data Analysis**

Study sites sent the Unified Huntington’s Disease Rating Scale forms to the coordination center at the University of Rochester, where data were entered. Data were cleaned and summarized at the University of Iowa. The percentage of participants with suicidal ideation endorsed was calculated within each group. The percentage of individuals experiencing suicidal thoughts was compared across groups. A chi-square test of independence was initially carried out to determine whether differences in the percentage of individuals with suicidal ideation across groups were large enough to warrant further comparisons. Significant findings were followed up with post hoc chi-square tests of independence between groups. After we compared the proportion of parti-
FIGURE 1. Percentage of 4,171 Subjects From the Huntington Study Group Database Experiencing Suicidal Ideation by Neurological Examination Scores and Huntington’s Disease Stage

Measure From Unified Huntington’s Disease Rating Scale

<table>
<thead>
<tr>
<th>Score on Neurological Examination</th>
<th>Subjects With Suicidal Ideation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (normal)</td>
<td>0</td>
</tr>
<tr>
<td>1 (soft signs)</td>
<td>1</td>
</tr>
<tr>
<td>2 (possible Huntington’s disease)</td>
<td>2</td>
</tr>
<tr>
<td>3 (definite Huntington’s disease)</td>
<td>3</td>
</tr>
</tbody>
</table>

Measure From Total Functional Capacity Scale

<table>
<thead>
<tr>
<th>Stage of Disease Based on Score on Total Functional Capacity Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
</tr>
<tr>
<td>0 (normal)</td>
</tr>
<tr>
<td>1 (soft signs)</td>
</tr>
<tr>
<td>2 (possible Huntington’s disease)</td>
</tr>
<tr>
<td>3 (definite Huntington’s disease)</td>
</tr>
</tbody>
</table>

* Participants who were considered by a movement disorder specialist to have a normal neurological examination were scored 0. Participants with “soft signs” that were considered nonspecific motor abnormalities were scored 1. Neurological examination findings that demonstrate motor abnormalities that may be signs of Huntington’s disease are considered to be “possible” Huntington’s disease and were scored 2. Motor abnormalities that are unequivocal signs and were diagnosed as Huntington’s disease were scored 3. Disease stage is only determined in participants with a diagnosis of definite Huntington’s disease. Huntington’s disease stage 1 was a total functional capacity score of 11 to 13, Huntington’s disease stage 2 was a total functional capacity score of 7 to 10, Huntington’s disease stage 3 was a total functional capacity score of 3 to 6, Huntington’s disease stage 4 was a total functional capacity score of 1 to 2, and Huntington’s disease stage 5 was a total functional capacity score of 0.

Results

The frequency distribution of suicidal ideation for the entire study group is presented in Table 2. Briefly, 82.5% did not endorse any suicidal thoughts. Mild suicidal ideation (defined as “slight” in severity) was endorsed by 7.2%. Moderate suicidal ideation (defined as “mild to moderate” in severity) was endorsed by 9.0%. Severe suicidal ideation (defined as “severe suicidal thoughts with intent and plan”) was endorsed by 1.3%.

The proportion of individuals in each diagnostic group with suicidal ideation is presented in Figure 1 (neurological examination score: 0=9.1%, 1=19.8%, 2=23.5%, and 3=18.5%). These proportions were found to be significantly different with the chi-square test ($\chi^2=48.1, df=3, p<0.001$). Follow-up chi-square tests among participants grouped by neurological examination score revealed a significant increase in the rate of suicidal ideation among individuals with an examination score of 1 or 2 ($\chi^2=16.1, df=2, p<0.001$) when they were compared with the participants with an examination score of 3, suggesting a diagnosis of Huntington’s disease. Additionally, the percentage of individuals with suicidal ideation significantly decreased from examination score 2 to examination score 3 ($\chi^2=5.7, df=1, p<0.05$).

The proportion of participants with definite Huntington’s disease who endorsed suicidal ideation is also presented in Figure 1, with patients grouped by stage of illness (stage 1=16.7%, 2=21.6%, 3=19.5%, 4=14.1%, 5=9.8%). A chi-square test of independence across Huntington’s disease stage in the definite Huntington’s disease group revealed a significant difference in the percentage of individuals with suicidal ideation among stages ($\chi^2=16.1, df=4, p<0.01$). Several statistically significant post hoc chi-square comparisons among stages of Huntington’s disease were observed. A statistically significant increase in ideation was observed between stages 1 and 2 ($\chi^2=5.5, df=1, p<0.05$). The proportion of individuals with suicidal ideation subsequently decreased with each stage beyond stage 2 ($\chi^2=13.8, df=3, p<0.01$).

The average severity of suicidal ideation for each group is shown in Figure 2. Although differences among groups based on neurological examination score were not significant, greater severity of ideation existed for diagnosed persons in later stages of disease based on functional capacity.

Conclusions

The current study confirms research suggesting elevated rates of suicidal ideation in persons at risk for and diagnosed with Huntington’s disease. It has been suggested that suicide in patients with Huntington’s disease occurs at a rate between seven and 200 times more often than in the general population (7, 19). Studies of suicide risk and actual completed suicide in Huntington’s disease are encumbered with methodological diversity, however,
making comparisons difficult (see Stenger and Stenger [20] for a review). Autopsy studies have reported suicide rates up to 13% (12), although the most frequently cited, and average, percentage is 5.7% (7). In one of the largest studies to date, Almqvist and her colleagues (21) reported completed suicide rates in a group of 1,817 persons recruited from 100 genetic testing centers from 21 countries. Bird (22) fluently put these findings into perspective when he compared the suicide rate in Huntington’s disease of 138 of 100,000 persons per year with that of the U.S. population (i.e., 12 to 13 per 100,000 persons [23]). Not surprisingly, studies assessing suicide risk, rather than completion, report much higher rates and have indicated that suicidal ideation occurs in up to 50% of people with Huntington’s disease (11, 24–26). Although suicidal ideation is commonly considered one index of suicidal risk, few studies have demonstrated the weight of the various predictors of suicide. Indeed, our current ability to predict actual suicide is poor. Future research should attempt to validate whether suicidal ideation in Huntington’s disease patients is predictive of actual suicide attempts.

Findings from the current study are consistent with, and extend confidence in, the existent research literature. This study used the largest group size ever studied with regard to suicidal ideation in Huntington’s disease. Unlike other large studies, with survey data obtained from family members, our study directly interviewed each individual about current suicidal ideation. There are other aspects of the current study that increase our confidence in the findings. Most notably, this study involves the collaboration of 43 sites in five countries with a standardized interview measure to assess suicidal ideation, a standardized rating scale for clinical features of Huntington’s disease, and a uniform method of disease diagnosis. All study investigators were required to complete training to ensure interrater reliability of the data collection. Finally, the sample studied involves a large number of geographically diverse individuals.

Using suicidal ideation as an index of suicide risk, findings suggest two “critical” periods for increased risk of suicide in Huntington’s disease. First, the proportion of individuals with suicidal ideation more than doubled, from 9.1% in persons at risk for Huntington’s disease with a normal neurological examination to 19.8% in at-risk persons with soft neurological signs. These findings highlight the vulnerability of this particular group of genetically at-risk individuals who begin manifesting signs and experiencing symptoms of Huntington’s disease. The time around the onset of Huntington’s disease has been previously recognized as a time of high risk for suicide. For example, Di Maio and colleagues (27) compared suicide age and age of disease onset among patients who committed suicide and concluded that suicide may occur at the first appearance of Huntington’s disease symptoms. Similarly, Schoenfeld and colleagues (12) commented that the prevalence of suicide appears four times higher among suspected Huntington’s disease patients than among those who are diagnosed with definite Huntington’s disease. During the initial onset of symptoms, persons are aware of the course of the illness.
and are still able to plan and implement a suicide strategy. Further research is needed to characterize the personality, environmental, and biological factors associated with suicidal ideation in presymptomatic Huntington's disease.

The clinical implications of these findings are great and suggest a major paradigm shift in clinical protocol. There is a widely held belief among clinicians that being given a diagnosis of Huntington's disease (or other devastating diseases) will worsen depression, instill hopelessness, and increase suicidal ideation. It has been argued that delaying diagnosis of terminal, fatal, and/or devastating disease somehow protects the patient (albeit temporarily) from the trauma of the disease. Our findings suggest that these not-so-uncommon views may be false and that, indeed, the opposite may be true, where the diagnosis reduces the risk of suicide. Contrary to clinical lore, receiving the diagnosis does not seem to be associated with more suicidal ideation. Our findings show that suicidal ideation actually goes down immediately after the diagnosis of Huntington's disease, suggesting that elevated suicide risk may be associated with the time period before diagnosis, when participants have greater uncertainty about whether they have Huntington's disease. Perhaps the frequency of suicide may be reduced by the expeditious diagnosis of Huntington's disease combined with appropriate treatment of depression. These findings mandate further discussion and exploration to better guide current clinical practice.

A second critical period for suicide risk in Huntington's disease occurs in the second stage of disease after diagnosis. Nearly 17% of persons in stage 1 endorsed thoughts of suicide, whereas over 21% indicated having suicidal ideation in stage 2. The second stage of Huntington's disease is often a difficult time period, where activities may be restricted (i.e., driving, managing finances) and dependence on others for activities of daily living increases. One study examining the proportion of deaths attributed to suicide among individuals diagnosed with Huntington's disease found that more than half of the suicides occurred in individuals showing early signs of the disease (12). It can be argued that intentions to terminate one's life increase when one's perception of independence dissipates.

The results suggest that certain periods during the progression of disease are associated with increases in the frequency of suicidal ideation in Huntington's disease. These critical periods include the development of soft neurological signs before a neurological diagnosis and, after diagnosis, the initial progression of the disease, including functional decline. Although the underlying mechanisms of suicide risk in Huntington's disease are poorly understood, it is beneficial for health care providers to be aware of periods during which patients may experience increased suicidal thoughts. Given the elevated rates of suicidal ideation in Huntington's disease, it is essential that health care providers regularly screen for suicidal ideation during clinical assessment.

Although our findings suggest that the proportion of persons having suicide ideation diminish with advancing disease, it is striking to note that the severity of suicidal thoughts increased over this same time period. It is difficult to interpret these findings in the context of the current study. One possibility for the findings of decreased proportion with advanced stage of illness may be that people who have already committed suicide are unavailable for study. It is also possible that persons in the later stages of disease are less often queried with regard to suicidal ideation or that they experience decreased verbal fluency, making the assessment of ideation difficult to ascertain. Further research in Huntington's disease as well as other terminal diseases may offer insight into suicide, its risk factors, and variation with disease course.

The current study has some limitations. Longitudinal study is required to better understand the relationship between suicidal ideation, illness course, and actual suicide attempts. Additionally, future studies should examine the correlates of suicidal ideation in Huntington's disease so that possible underlying mechanisms for increased suicidal ideation might be established. For instance, it is unknown whether the presence of a major depressive disorder is associated with suicide risk in Huntington's disease. Although the high prevalence of depression in Huntington's disease has been well established, the nature of depression and suicidality in Huntington's disease are poorly understood. It is unknown what proportion of persons with Huntington's disease experience symptoms of depression secondary to biological changes in the basal ganglia and what proportion of persons are experiencing increased distress secondary to anticipation of a pending event. The interdependence of depression due to neurotransmitter regulation in the brain and depression secondary to life stressors is complex and requires further elaboration. Huntington's disease may be a good model in which to further study the underlying mechanisms of depression and suicidality. Further research may also require efforts to better separate suicide associated with major depressive disorder from rational suicide in persons with a known terminal illness.

Received March 1, 2004; revisions received April 30 and June 17, 2004; accepted July 16, 2004. From the Departments of Psychiatry, Neurology, and Psychology, University of Iowa. Address correspondence and reprint requests to Dr. Paulsen, University of Iowa, Roy J. and Lucille A. Carver College of Medicine, 1-305 Medical Education Bldg., Iowa City, IA 52241-1000; jane-paulsen@uiowa.edu (e-mail).

Funded by grant 40068 from the National Institute of Neurological Disorders and Stroke, NIMH grant 01579 to Jane S. Paulsen, the Huntington's Disease Society of America, the Huntington's Society of Canada, Hereditary Disease Foundation grants to the Huntington Study Group, grant AG-00214 awarded to Karin Ferneyhough Hoth through the University of Iowa Center on Aging from the National Institute on Aging National Research Service Award Interdisciplinary Research Training Program on Aging, and an NIH National Research Service Award (MH-065830) to Carissa Nehl.
Acknowledgments

Participating investigators and coordinators in the Huntington Study Group were as follows: Phillipa Hedges, Elizabeth McCusker, M.D., Samantha Pearce, and Ronald Trent, Ph.D., Westmead Hospital, Sydney, New South Wales, Australia; David Abwender, Ph.D., Peter Como, Ph.D., Irenita Gardiner, R.N., Charlyne Hickey, R.N., Elise Kayson, R.N., Karl Kieburz, M.D., Frederick Marshall, M.D., Nancy Pearson, R.N., Ira Shoulson, M.D., and Carol Zimmerman, R.N., University of Rochester, N.Y.; Elan Louis, M.D., Karen Marder, M.D., Carol Moskowitz, R.N., Carmen Polanco, B.A., Stuart Taylor, M.D., and Naomi Zulin, B.A., Columbia Presbyterian Medical Center, New York; Catherine Brown, R.N., Jill Burkeholder, Mark Guttman, M.D., Sandra Russell, Dwight Stewart, M.D., and Jackie Thomson, R.N., Markham Health Center, Toronto; Daniel S. Sax, M.D., and Marie Saint-Hilaire, M.D., Boston University, Boston; Jackie Gray, Cindy Hunter, M.S., Nanette Mercado, Ph.D., Eric Siemers, M.D., and Joanne Wojcieszek, M.D., Indiana University School of Medicine, Indianapolis; Ted Dawson, M.D., Elizabeth Leritz, B.S., Adam Rosenblatt, M.D., Meia Sherr, R.N., and Candace Young, R.N., Johns Hopkins University, Baltimore; Tetuio Asashizawa, M.D., Jenny Beach, R.N., and Joseph Jankovic, M.D., Baylor College of Medicine, Houston; Jeana Jaglin, R.N., and Kathleen Shannon, M.D., Rush Presbyterian/St. Luke’s Medical Center, Chicago; Anders Lundin, M.D., Karolinska Hospital, Stockholm; Kathleen Francis, M.D., and Kim Lane, University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School, Camden; Alexander Auchus, M.D., J. Timothy Greenamyre, M.D., Steven Hersch, M.D., Randi Jones, Ph.D., and David Olson, M.D., Emory University, Atlanta; Jang-Ho John Cha, M.D., Merit Cudkowicz, M.D., Walter Koroshetz, M.D., Greg Rudolf, Paula Sexton, M.A., and Anne B. Young, M.D., Massachusetts General Hospital, Boston; Roger Albin, M.D., and Kristine Wernette, R.N., University of Michigan, Ann Arbor; Donald S. Higgins, M.D., and Carson Reiser, M.D., Ohio State University, Columbus; Vicki Hunt, R.N., and Francis Walker, M.D., Bowman Gray School of Medicine, Winston-Salem, N.C.; Robert Hauser, M.D., Juan Sanchez-Ramos, M.D., and Audrey Walker, R.N., University of South Florida, Tampa; Martha Nance, M.D., Minneapolis Veterans Administration Medical Center, Minneapolis; Susan Cleary, R.N., Gina Rohs, and Oksana Chowersky, M.D., University of Calgary Medical Center, Calgary, Alta., Canada; Kerry Duncan and Lauren Seeberger, M.D., Colorado Neurologic Institute, Englewood; Jody Corey-Bloom, M.D., Michael Swenson, M.D., and Neal S. Fruin, M.D., University of California, San Diego; Henry Paulson, M.D., Ph.D., Robert Rodnitzky, M.D., Lynn Vining, R.N., and Jane Paulsen, Ph.D., University of Iowa, Iowa City; Wayne Martin, M.D., and Marguerite Wieler, B.Sc., P.T., University of Alberta, Edmonton, Alta., Canada; Alicia Facca, M.D., Gustavo Rey, Ph.D., and William Weiner, M.D., University of Miami, Miami; Charles Adler, M.D., John Caviness, M.D., Cindy Lied, R.N., and Stephanie Newman, R.N., Mayo Clinic, Scottsdale, Ariz.; Andrew Feigin, M.D., and Jennifer Mazurkiewicz, B.A., North Shore University Hospital, Manhasset, N.Y.; Karen Caplan, M.S.W., Janet Cellar, R.N., and Kenneth Marek, M.D., Yale University School of Medicine, New Haven, Conn.; Michael Hayden, M.D., Lynn Raymond, M.D., and Gina Rohs, University of British Columbia, Vancouver, B.C., Canada; Leon S. Dure, M.D., and Jane Lane, R.N., Children’s Hospital of Alabama, Birmingham; Diane Brown, R.N., Stewart Factor, D.O., and Eric Molho, M.D., Albany Medical College, Albany, N.Y.; Madeline Harrison, M.D., Carol Manning, Ph.D., and Elke Rost-Ruffner, R.N., University of Virginia, Charlottesville; Jonelle Adams, Ruth Cummings, R.N., and Vicki Wheelock, M.D., University of California, Davis; Richard Dubinsky, M.D., and Carolyn Gray, R.N., University of Kansas Medical Center, Kansas City; Ann Catherine Bachoud-Levi, M.D., Hôpital Henri Mondor, Creteil, France; Hartmut Meierkord, M.D., Universitätsklinikum Charité, Berlin; Joseph Friedman, M.D., and Margaret Lannon, R.N., Memorial Hospital of Rhode Island, Pawtucket; Joan Lawrence, M.D., Royal Brisbane Hospital, Brisbane, Queensland, Australia; and Allen Rubin, M.D., and Rose Schwarz, R.N., Allegheny University, Philadelphia. The biostatistics and coordination center staff were as follows: Alicia Brocht, B.A., Kathi Claude, M.S., Joshua Goldstein, Michael McDermott, Ph.D., and David Oakes, Ph.D., University of Rochester, Rochester, N.Y.

References

16. Benton AL, Hamsher KD: Multilingual Aphasia Examination. Iowa City, University of Iowa, 1976
19. Hayden MR, Berkowicz AL, Beighton PH, Yiptong C: Hunting-
21. Almqvist EW, Bloch M, Brinkman R, Craufurd D, Hayden MR: A worldwide assessment of the frequency of suicide, suicide at-
ttempts, or psychiatric hospitalization after predictive testing for Huntington disease. Am J Hum Genet 1999; 64:1293–1304
22. Bird TD: Outrageous fortune: the risk of suicide in genetic test-
24. Bloch M, Fahy M, Fox S, Hayden MR: Predictive testing for Hun-
tington disease, II: demographic characteristics, life-style pat-
terns, attitudes, and psychosocial assessments of the first fifty-
26. Mastromauro C, Myers RH, Berkman B: Attitudes toward pre-
Symptom Fluctuation in Eating Disorders: Correlates of Diagnostic Crossover

Federica Tozzi, M.D.
Laura M. Thornton, Ph.D.
Kelly L. Klump, Ph.D.
Manfred M. Fichter, M.D.
Katherine A. Halmi, M.D.
Allan S. Kaplan, M.D.
Michael Strober, Ph.D.
D. Blake Woodside, M.D.
Scott Crow, M.D.
James Mitchell, M.D.
Alessandro Rotondo, M.D.
Mauro Mauri, M.D.
Giovanni Cassano, M.D.
Pamela Keel, Ph.D.
Katherine H. Plotnicov, Ph.D.
Christine Pollice, M.P.H.
Lisa R. Lilenfeld, Ph.D.
Wade H. Berrettini, M.D.
Cynthia M. Bulik, Ph.D.
Walter H. Kaye, M.D.

Objective: The course of anorexia nervosa often includes the emergence of bulimic symptoms and a crossover to the full syndrome of bulimia nervosa. However, clinicians’ ability to predict who will develop bulimia nervosa is limited. The converse phenomenon, crossover from bulimia nervosa to anorexia nervosa, has not been investigated as thoroughly. The authors identified factors that are associated with crossover from anorexia nervosa to bulimia nervosa and from bulimia nervosa to anorexia nervosa.

Method: All participants were from the International Price Foundation Genetic Study. Two groups were studied. The first comprised 88 individuals with an initial diagnosis of anorexia nervosa, of whom 32 developed bulimia nervosa. The second included 350 individuals with bulimia nervosa, of whom 93 developed anorexia nervosa. Several variables, including DSM-IV axis I and II disorders and personality traits, were evaluated as potential predictors of crossover.

Results: For the majority of affected individuals, crossover occurred by the fifth year of illness. A low level of self-directedness was associated with crossover in both directions. Other factors differed by diagnosis: high parental criticism was associated with crossover from anorexia nervosa to bulimia nervosa, whereas alcohol abuse/dependence and a low level of novelty seeking were associated with crossover from bulimia nervosa to anorexia nervosa.

Conclusions: Low self-directedness may be associated with diagnostic instability in general, whereas other specific factors are related to the direction of diagnostic crossover. These results indicate that personality and family characteristics may influence the course of eating disorders and may be informative for planning interventions.

of the emergence of the symptom of binge eating and not predictors of the emergence of frank bulimia nervosa. However, of the 95 patients who entered the treatment program for restricting-type anorexia nervosa, 29% developed binge eating during the follow-up, of whom about 65% developed a full bulimia nervosa syndrome. Studies of crossover from bulimia nervosa to anorexia nervosa are rarer, have generally included small numbers of participants (mean N=77, range=30–196), and suggest a lower rate of occurrence (0%–4%) than has been observed for crossover from anorexia nervosa to bulimia nervosa (13, 19–24). To our knowledge, no data exist regarding predictors of crossover from bulimia nervosa to anorexia nervosa.

Although few studies have examined diagnostic crossover, several studies have characterized the psychopathological features of anorexia nervosa and bulimia nervosa (25, 26). These studies are relevant because features that are shared by both disorders may predict diagnostic stability, whereas features that are distinct may influence diagnostic fluidity. Individuals with anorexia nervosa and bulimia nervosa share high levels of neuroticism, obsessiosity, and perfectionism (27–29) and low levels of self-directedness (30, 31).

Nonetheless, there are several differences in personality associated with the disorders that might influence crossover. For example, individuals with restricting-type anorexia nervosa tend toward rigidity and overcontrol, whereas individuals with bulimia nervosa tend toward impulsivity and affective dysregulation (27, 28, 32, 33). Personality profiles of individuals with the binge-purge subtype of anorexia nervosa tend to be intermediate between those of individuals with bulimia nervosa and with restricting-type anorexia nervosa, especially regarding impulsivity (30, 34).

To our knowledge, the present study represents the first attempt to study crossover both from anorexia nervosa to bulimia nervosa and from bulimia nervosa to anorexia nervosa. We explored variables previously shown either to be associated with crossover from anorexia nervosa to bulimia nervosa or to differentiate women with anorexia nervosa from those with bulimia nervosa.

Screening and Diagnostic Procedures

If—in an initial screening—a proband met the criteria for bulimia nervosa, no exclusion criteria (see next paragraph), and reported a possible history of an eating disorder in a nonparent, nonchild, or a nonmonozygotic twin blood relative, the proband was asked to discuss the study with the affected relative and obtain permission to be contacted by study personnel. If both the proband and affected relative met the inclusion criteria, informed consent was obtained and both were scheduled to complete assessments. If participants lived near one of the study sites, interviews were completed in person. Others were interviewed by phone.

The participants for this study included both probands and affected relatives. Probands met the following criteria: 1) modified DSM-IV lifetime diagnosis of bulimia nervosa, purging type (purging must have included regular vomiting [with other means of purging allowed], and bingeing and vomiting must have occurred at least twice a week for a duration of at least 6 months) and 2) ages between 13 and 65 years. The exclusion criteria for probands were mental retardation (IQ <70); dementia; organic brain syndromes; psychotic disorders, including schizophrenia, schizophreniform disorder, delusional disorder, and schizoaffective disorder; Turner’s syndrome; any medical condition that could affect appetite, body weight, or eating (e.g., individuals with diabetes and thyroid conditions were excluded if the onset of the disease preceded the onset of the eating disorder); bipolar I disorder or bipolar II disorder only if symptoms of bulimia nervosa occurred exclusively during manic or hypomanic episodes; and neurological problems, except for participants with a diagnosis of a seizure disorder resulting from trauma after the onset of the eating disorder. Probands whose premorbid weight exceeded the body mass index for the 95th percentile for gender and age on the Hebebrand index (36) or whose high lifetime body mass index was greater than 35 kg/m² were also excluded.

Affected relatives were biological family members who 1) were between ages 13 and 65 years and 2) met the criteria for lifetime bulimia nervosa, purging type or anorexia nervosa, restricting type. The eating disorder diagnoses for the relatives were defined as follows: 1) DSM-IV bulimia nervosa, purging type, and 2) modified DSM-IV anorexia nervosa (i.e., criterion D not required). To be defined as having persistent restricting-type anorexia nervosa, participants had to have never binged or purged except experimentally (defined as less than five times ever). The exclusion criteria for affected relatives included those for probands plus 1) being a monozygotic twin of the proband; 2) being a biological parent with an eating disorder, unless there was another affected family member with whom the parent could be paired; and 3) having a diagnosis of binge-eating disorder as the only lifetime eating disorder diagnosis.

We conducted two series of analyses. First, to explore crossover from anorexia nervosa to bulimia nervosa, we compared 1) individuals with persistent restricting-type anorexia nervosa (N=56) and 2) individuals with an initial diagnosis of restricting-type anorexia nervosa who developed bulimia nervosa (N=32). Second, to explore crossover from bulimia nervosa to anorexia nervosa, we compared individuals who met the criteria for a diagnosis of 1) persistent normal weight bulimia nervosa, purging type (N=257) and 2) individuals who initially had normal weight bulimia nervosa followed by the development of anorexia nervosa, binge-eating/purging type (N=93). As a result of the recruitment procedure (i.e., probands had to have bulimia nervosa), all individuals with a diagnosis of anorexia nervosa were affected relatives, and all individuals with bulimia nervosa or both diagnoses were either probands or affected relatives.

Male participants were excluded from analyses, as the number of male participants with these diagnoses was too small for meaningful comparison. We were unable to examine other types of cross-disorder.
We evaluated self-reported current, crossover types. Anorexia Nervosa to Bulimia Nervosa and From Bulimia Nervosa to Anorexia Nervosa because of the low frequency of these crossover (e.g., normal weight bulimia nervosa to restricting-type anorexia nervosa) because of the low frequency of these crossover types.

**Measures**

**Physical characteristics.** We evaluated self-reported current, past minimum, and past maximum body mass index (kg/m²) and age at menarche.

**Eating disorder psychopathology.** The Structured Interview for Anorexic and Bulimic Disorders for DSM-IV and ICD-10 (37) was used to assess specific and general psychopathology of eating disorders and lifetime history of eating disorders among probands and relatives.

**Psychiatric diagnoses.** The Structured Clinical Interview for DSM-IV (SCID) Axis I and Axis II Disorders (38, 39) was used to assess lifetime axis I and axis II diagnoses. Among axis I diagnoses, we excluded bipolar I disorder and bipolar II disorder, because the frequencies were too low among the study participants. Impulse-control disorders (intermittent explosive disorder, pathological gambling, pyromania, trichotillomania, compulsive shopping, and kleptomania) were grouped together as an umbrella impulse-control disorder diagnosis. Among axis II diagnoses, we included only those disorders with adequate frequency, including avoidant, obsessive-compulsive, and borderline personality disorder. Lifetime diagnoses of eating disorders were determined by using information from both the Structured Interview for Anorexic and Bulimic Disorders and the SCID.

**Perfectionism.** The Multidimensional Perfectionism Scale (40) is designed to assess multidimensional facets of perfectionism. Although the instrument includes six subscales, we included only those subscales that have been shown to be most relevant to eating disorders, including concern over mistakes, doubts about actions, personal standards, parental criticism, and parental expectation.

**Temperament and character.** Temperament, as measured by the Temperament and Character Inventory (41), is postulated to be largely biological in origin (42, 43) and to include the dimensions of novelty seeking, harm avoidance, reward dependence, and persistence. Character dimensions on the Temperament and Character Inventory are thought to develop with experience (41) and include self-directedness, cooperativeness, and self-transcendence.

**Anxiety.** Although the State-Trait Anxiety Inventory—Form Y (44) measures two aspects of anxiety (i.e., state and trait anxiety), we included only the trait anxiety scale in the present analyses.

**Neuroticism.** The present analysis included data from the neuroticism scale of the Neuroticism, Extraversion, and Openness to Experience Inventory (45), as neuroticism has been consistently associated with eating disorders (46, 47), particularly a bulimic symptom presentation (48, 49).

**Statistical Analyses**

Logistic regressions with generalized estimating equation (50–52) corrections for dependent data were used to predict crossover. As status of illness (i.e., acutely ill or recovered) can influence scores on personality measures (53), this variable was included in the models as a covariate. Women were considered recovered if they had an absence of eating disorder behaviors (e.g., binge eating, purging, dieting, nonpurging behaviors) for a period of at least 1 year, as reported on an expanded version of Module H of the SCID (38). In addition, women with anorexia nervosa had to have their body mass index above the fifth percentile of their ideal body mass index for their age for at least the past year.

Variables that emerged in the initial logistic regressions as significantly associated with crossover were then entered into a stepwise logistic regression. All continuous measures were standardized before the analyses to have a mean of zero and unit variance. Given that we performed multiple statistical tests, we set the significance level to $p < 0.01$, representing a compromise between the more stringent Bonferroni correction and the exploratory nature of this study.

**Results**

**Time to Crossover**

Of the participants whose initial diagnosis was restricting-type anorexia nervosa (N=88), 36% (N=32) developed bulimia nervosa. For the majority (91%, N=29), crossover occurred within the first 5 years of illness. Of the 350 individuals who first had bulimia nervosa, 27% (N=93) developed anorexia nervosa, binge-eating/purging type. For the majority (77%, N=72), crossover occurred within the first 5 years. See Figure 1 for the survival density function of both types of crossovers.

**Correlates of Crossover**

**Crossover from anorexia nervosa to bulimia nervosa.** Logistic regression indicated that low self-directedness and high parental criticism, neuroticism, and trait anxiety were significantly associated with crossover from anorexia nervosa to bulimia nervosa (Table 1). Among DSM diagnoses, only depression and substance use disorder were associated with crossover (Table 2). All of these variables were entered into a stepwise logistic regression to identify the factors that had the strongest association with crossover. Body mass index variables, although significant, were not included in further analyses, as body mass index is inherently confounded because it is a defining criterion for anorexia nervosa. Depression, substance use disorder, and neuroticism were eliminated from the initial model,
TABLE 1. Characteristics of Women Study Participants With Diagnostic Crossover From Anorexia Nervosa to Bulimia Nervosa and With Anorexia Nervosa Only

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants With Crossover From Anorexia Nervosa to Bulimia Nervosa (N=32)a</th>
<th>Participants With Anorexia Nervosa Only (N=56)a</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Mean 28.28, SD 9.60</td>
<td>Mean 25.21, SD 8.38</td>
<td>Odds Ratio 1.40</td>
</tr>
<tr>
<td>Menarche</td>
<td>Mean 15.16, SD 4.32</td>
<td>Mean 17.38, SD 3.98</td>
<td>Odds Ratio 0.82</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>Mean 13.44, SD 1.50</td>
<td>Mean 13.39, SD 1.73</td>
<td>Odds Ratio 1.03</td>
</tr>
<tr>
<td>At interview</td>
<td>Mean 21.15, SD 1.65</td>
<td>Mean 18.84, SD 2.07</td>
<td>Odds Ratio 5.17</td>
</tr>
<tr>
<td>Maximum</td>
<td>Mean 23.48, SD 2.88</td>
<td>Mean 21.30, SD 2.68</td>
<td>Odds Ratio 2.34</td>
</tr>
<tr>
<td>Minimum</td>
<td>Mean 14.62, SD 1.92</td>
<td>Mean 15.11, SD 1.65</td>
<td>Odds Ratio 0.75</td>
</tr>
</tbody>
</table>

a Data are missing for some participants for some characteristics.
b Chi-square values test for significant differences between groups from logistic regressions that were conducted with generalized estimating equation corrections. To correct for multiple testing, significance is indicated if the p value is ≤0.01.

and only self-directedness (χ²=7.30, df=1, p=0.007, odds ratio=0.45, 95% confidence interval [CI]=0.26–0.81) and parental criticism (χ²=5.00, df=1, p=0.03, odds ratio=1.83, 95% CI=1.08–3.12) remained significant.

**Crossover from bulimia nervosa to anorexia nervosa.**
Low novelty seeking and self-directedness, high harm avoidance, alcohol abuse/dependence, and avoidant personality disorder were associated with crossover from bulimia nervosa to anorexia nervosa in the univariate analyses (Table 2 and Table 3). As in the earlier analyses, body mass index variables were not included in the stepwise analysis. Variables that remained significant in the stepwise logistic models included novelty seeking (χ²=6.11, df=1, p=0.02, odds ratio=0.72, 95% CI=0.56–0.94), self-directedness (χ²=10.81, df=1, p=0.001, odds ratio=0.63, 95% CI=0.48–0.83), and alcohol abuse/dependence (χ²=7.75, df=1, p=0.005, odds ratio=0.46, 95% CI=0.27–0.80).

**Discussion**
This study is, to our knowledge, the first to examine crossover both from anorexia nervosa to bulimia nervosa and from bulimia nervosa to anorexia nervosa. Supporting previous research, we observed that for most individu-
other factors characteristic of individuals with eating disorders with fluctuating clinical features. For example, low self-directedness has been associated with greater comorbidity (54, 55) and a poorer outcome (9). Consequently, it may be that individuals with diagnostic crossover represent cases with greater comorbidity in which additional eating disorder diagnoses simply reflect greater psychopathology in general.

Whereas low self-directedness appeared to be a common factor characterizing both types of crossovers, other factors were specifically related to each crossover type. Family factors, namely perceived parental criticism, were particularly salient for individuals with anorexia nervosa who crossed over to having bulimia nervosa. This finding is noteworthy given early observations of family environment across eating disorder subtypes, suggesting that families of individuals with bulimia tend to exhibit greater conflict and disorganization and less cohesion than families of those with anorexia nervosa (56), that mothers of individuals with bulimia nervosa were reported to be more domineering and have higher expectations of their daughters than were control individuals (57), and that women who crossed over to having bulimia nervosa reported maternal deficits in nurturance and empathy (58). Our findings also corroborate those of Strober et al. (4), who found low levels of parental empathy and affection to be significant predictors of the onset of binge eating in women with anorexia nervosa. These observations are particularly noteworthy, given the finding in expressed emotion research that maternal critical comments are strongly predictive of treatment outcome for adolescents with eating disorders (59).

Low scores on impulse-related personality traits (e.g., novelty seeking) and the presence of certain behaviors (e.g., alcohol abuse/dependence) were important in crossover from bulimia nervosa to anorexia nervosa. Lower levels of impulsivity may enable the maintenance of rigid dietary regimes of sufficient duration to lose the amount of weight necessary for an anorexia nervosa diagnosis.

We did not find anxiety, neurotic personality traits, or depression to be significantly associated with crossover in the hierarchical models, although they were often significant predictors in univariate analyses. These findings are somewhat contradictory to previous observations of anxiety as an important predictor of crossover from anorexia nervosa to bulimia nervosa (1). However, earlier studies used different instruments and methods and are not entirely comparable to the current study.

Although our study had many strengths (e.g., carefully defined diagnostic phenotypes, examination of a range of variables), several limitations should be noted. First, we were unable to examine all possible diagnostic crossovers (e.g., purging-type bulimia nervosa to restricting-type anorexia nervosa) because some were rare. Future research should examine whether similar types of predictors also characterize these diagnostic changes. Second, participants may not have reached the “steady state” of their illness. Approximately 57% of the anorexia nervosa participants and 38% of the bulimia nervosa participants were interviewed before the fifth year of illness. Therefore, additional crossovers that were not captured in the present analyses could still occur.

TABLE 2. Axis I and Axis II Disorders in Women Study Participants With Diagnostic Crossover From Anorexia Nervosa to Bulimia Nervosa, Anorexia Nervosa Only, Diagnostic Crossover From Bulimia Nervosa to Anorexia Nervosa, and Bulimia Nervosa Only

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Participants With Crossover From Anorexia Nervosa to Bulimia Nervosa</th>
<th>Participants With Anorexia Nervosa Only</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Impulse-control disorderb</td>
<td>4</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>15</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>Social phobia</td>
<td>10</td>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td>Specific phobia</td>
<td>4</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Alcohol abuse/dependency</td>
<td>11</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>Posttraumatic stress disorder</td>
<td>5</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Depression</td>
<td>25</td>
<td>78</td>
<td>27</td>
</tr>
<tr>
<td>Avoidant personality disorder</td>
<td>8</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Obsessive-compulsive personality disorder</td>
<td>7</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Borderline personality disorder</td>
<td>5</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>3</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Agoraphobia</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Any anxiety disorderc</td>
<td>19</td>
<td>59</td>
<td>21</td>
</tr>
<tr>
<td>Psychoactive substance use disorder, excluding alcohol use disorders</td>
<td>11</td>
<td>34</td>
<td>6</td>
</tr>
</tbody>
</table>

a Chi-square values test for significant differences between groups from logistic regressions, with generalized estimating equation correction.

b To correct for multiple testing, significance is indicated if the p value is ≤0.01.

c Included intermittent explosive disorder, pathological gambling, pyromania, trichotillomania, compulsive shopping, and kleptomania.
Third, few details were available about changes in severity of illness over time (e.g., periods of partial or full recovery between anorexia nervosa and bulimia nervosa onset). These data would have provided a more comprehensive picture of the course of eating disorders and may have revealed additional correlates of diagnostic instability. Fourth, data about childhood psychiatric disorders or sexual abuse, which emerged as predictors of crossovers in previous studies (1), were not available. Fifth, onset of diagnoses and determination of crossover were based on retrospective reports and may be subject to recall biases. However, the rate of crossover from anorexia nervosa to bulimia nervosa observed in our study was similar to that reported in longitudinal studies (2–4), suggesting that recall biases may have been minimal.

Finally, the study included individuals who had at least one family member with a disorder, which may have influenced the results. For example, we found a higher rate of crossover from bulimia nervosa to anorexia nervosa, compared with previous studies. Although this discrepancy might be related to the particular group of participants selected (e.g., all anorexia nervosa probands had a family member with bulimia nervosa), it could also be a more accurate reflection of true crossover rates, given that the number of participants was substantially larger than in previously reported studies. In addition, none of the cited studies, except for one (1), had the specific goal of exploring the crossover phenomenon, and most of them were designed as longitudinal outcome studies. Thus, with respect to crossover, the earlier studies present limitations, such as lack of study groups strictly defined with respect to crossover (i.e., at recruitment, the bulimia nervosa group in these earlier studies may have included individuals with a history of anorexia nervosa). In addition, earlier studies reported rates of crossover that refer only to individuals who still had an eating disorder diagnosis at follow-up, and thus these studies did not explore possible crossovers before recovery. Moreover, we focused on crossovers between full syndromal anorexia nervosa and bulimia nervosa, whereas the majority of previous studies evaluated the development of bulimic behaviors, rather than the full bulimia nervosa syndrome, in individuals with restricting-type anorexia nervosa. Our study, although exploratory in nature, has the advantage of diagnostic rigor and specific attention to the crossover phenomenon. However, the extent to which our findings were biased by our recruitment procedures, rather than depicting an accurate association between bulimia nervosa and anorexia nervosa crossover, is unclear. Resolution of this issue awaits future research employing less select groups of participants that parallel the group in our study in size and diagnostic characterization.

In summary, this study furthers our understanding of diagnostic crossover in eating disorders by suggesting that there are features, such as self-directedness, that affect general diagnostic instability, while there are other factors specifically related to the initial eating disorder diagnosis that influence crossover in one direction but not the other. If findings from this exploratory study are confirmed, they may have important treatment implications. Low self-directedness has been associated with negative outcome (9, 60, 61). Indeed, high self-directedness predicts rapid and sustained response to cognitive behavior therapy in

<table>
<thead>
<tr>
<th>Participants With Initial Diagnosis of Bulimia Nervosa</th>
<th>Participants With Crossover From Bulimia Nervosa to Anorexia Nervosa</th>
<th>Participants With Bulimia Nervosa Only</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>61</td>
<td>24</td>
</tr>
<tr>
<td>26</td>
<td>28</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>29</td>
<td>31</td>
<td>121</td>
<td>47</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>68</td>
<td>73</td>
<td>179</td>
<td>70</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>29</td>
<td>31</td>
<td>60</td>
<td>23</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>57</td>
<td>61</td>
<td>115</td>
<td>45</td>
</tr>
<tr>
<td>22</td>
<td>24</td>
<td>75</td>
<td>29</td>
</tr>
</tbody>
</table>
### TABLE 3. Characteristics of Women Study Participants With Diagnostic Crossover From Bulimia Nervosa to Anorexia Nervosa and With Bulimia Nervosa Only

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants With Crossover From Bulimia Nervosa to Anorexia Nervosa (N=93)a</th>
<th>Participants With Bulimia Nervosa Only (N=257)b</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>$\chi^2$ (df=1)</th>
<th>p&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At interview</td>
<td>29.3 ± 10.0</td>
<td>27.8 ± 9.4</td>
<td>1.21</td>
<td>0.94–1.56</td>
<td>2.20</td>
<td>0.14</td>
</tr>
<tr>
<td>At bulimia nervosa onset</td>
<td>18.3 ± 4.3</td>
<td>17.6 ± 3.6</td>
<td>1.04</td>
<td>0.99–1.11</td>
<td>1.60</td>
<td>0.21</td>
</tr>
<tr>
<td>At first eating disorder symptom</td>
<td>15.9 ± 4.1</td>
<td>15.8 ± 3.4</td>
<td>1.00</td>
<td>0.93–1.08</td>
<td>0.01</td>
<td>0.91</td>
</tr>
<tr>
<td>Menarche</td>
<td>12.8 ± 1.6</td>
<td>12.9 ± 1.7</td>
<td>0.93</td>
<td>0.74–1.16</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At interview</td>
<td>19.3 ± 2.1</td>
<td>22.7 ± 3.0</td>
<td>0.10</td>
<td>0.05–5.50</td>
<td>64.44</td>
<td>0.0001</td>
</tr>
<tr>
<td>Maximum</td>
<td>23.1 ± 2.8</td>
<td>25.5 ± 3.2</td>
<td>0.39</td>
<td>0.27–5.56</td>
<td>34.70</td>
<td>0.0001</td>
</tr>
<tr>
<td>Minimum</td>
<td>15.7 ± 1.8</td>
<td>19.3 ± 1.7</td>
<td>0.005</td>
<td>0.0001–0.024</td>
<td>90.96</td>
<td>0.0001</td>
</tr>
<tr>
<td>Multidimensional Perfectionism Scale scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concern over mistakes</td>
<td>31.5 ± 9.2</td>
<td>29.0 ± 9.6</td>
<td>1.26</td>
<td>0.98–1.63</td>
<td>3.30</td>
<td>0.07</td>
</tr>
<tr>
<td>Doubts about actions</td>
<td>13.1 ± 3.8</td>
<td>11.9 ± 4.0</td>
<td>1.32</td>
<td>1.03–1.69</td>
<td>4.61</td>
<td>0.03</td>
</tr>
<tr>
<td>Personal standards</td>
<td>25.4 ± 6.3</td>
<td>24.9 ± 6.4</td>
<td>1.07</td>
<td>0.84–1.36</td>
<td>0.32</td>
<td>0.57</td>
</tr>
<tr>
<td>Parental criticism</td>
<td>11.6 ± 4.8</td>
<td>10.8 ± 4.5</td>
<td>1.21</td>
<td>0.93–1.57</td>
<td>1.94</td>
<td>0.16</td>
</tr>
<tr>
<td>Parental expectations</td>
<td>14.5 ± 5.7</td>
<td>13.7 ± 5.7</td>
<td>1.16</td>
<td>0.90–1.51</td>
<td>1.33</td>
<td>0.25</td>
</tr>
<tr>
<td>Neuroticism, Extraversion, and Openness to Experience Inventory Neuroticism score</td>
<td>117.4 ± 26.8</td>
<td>110.2 ± 23.6</td>
<td>1.25</td>
<td>0.93–1.67</td>
<td>2.40</td>
<td>0.12</td>
</tr>
<tr>
<td>State-Trait Anxiety Inventory—Form Y trait anxiety score</td>
<td>51.6 ± 14.2</td>
<td>46.6 ± 12.7</td>
<td>1.36</td>
<td>1.06–1.75</td>
<td>5.68</td>
<td>0.02</td>
</tr>
<tr>
<td>Temperament and Character Inventory scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novelty seeking</td>
<td>20.0 ± 6.3</td>
<td>22.2 ± 6.4</td>
<td>0.69</td>
<td>0.54–0.87</td>
<td>9.24</td>
<td>0.002</td>
</tr>
<tr>
<td>Reward dependence</td>
<td>16.1 ± 4.3</td>
<td>16.9 ± 4.1</td>
<td>0.82</td>
<td>0.65–1.03</td>
<td>2.65</td>
<td>0.10</td>
</tr>
<tr>
<td>Self-directedness</td>
<td>22.7 ± 9.7</td>
<td>27.0 ± 8.8</td>
<td>0.66</td>
<td>0.50–0.86</td>
<td>9.06</td>
<td>0.003</td>
</tr>
<tr>
<td>Harm avoidance</td>
<td>20.9 ± 8.0</td>
<td>17.6 ± 7.7</td>
<td>1.44</td>
<td>1.12–1.86</td>
<td>7.74</td>
<td>0.005</td>
</tr>
<tr>
<td>Self-transcendence</td>
<td>14.4 ± 6.0</td>
<td>16.2 ± 6.5</td>
<td>0.73</td>
<td>0.55–0.95</td>
<td>5.60</td>
<td>0.02</td>
</tr>
<tr>
<td>Persistence</td>
<td>5.0 ± 2.1</td>
<td>5.0 ± 2.1</td>
<td>1.02</td>
<td>0.80–1.30</td>
<td>0.02</td>
<td>0.90</td>
</tr>
<tr>
<td>Cooperativeness</td>
<td>32.4 ± 6.9</td>
<td>34.0 ± 5.8</td>
<td>0.76</td>
<td>0.60–0.95</td>
<td>4.67</td>
<td>0.03</td>
</tr>
</tbody>
</table>

a Data are missing for some participants for some characteristics.

b Chi-square values test for significant differences between groups from logistic regressions, with generalized estimating equation correction. To correct for multiple testing, significance is indicated if the p value is ≤0.01.

bulimia nervosa patients (62), and there is evidence that cognitive behavior therapy leads to increased self-directedness (63). Thus, self-directedness may influence not only the diagnostic stability of eating disorders but also their course and response to treatment. The development of additional therapeutic approaches aimed at increasing levels of self-directedness may enhance the efficacy of existing treatments. Additional techniques aimed at enhancing familial communication in cases of anorexia nervosa and at addressing impulsivity in bulimia nervosa are also likely to promote diagnostic stability and possibly shorter recovery times.

---

**References**


63. Anderson CB, Joyce PR, Carter FA, McIntosh VV, Bulik CM: The effect of cognitive-behavioral therapy for bulimia nervosa on temperament and character as measured by the temperament and character inventory. Compr Psychiatry 2002; 43:182–188
Three Psychotherapies for Anorexia Nervosa: A Randomized, Controlled Trial

Objective: Few randomized, controlled trials have examined the efficacy of treatments for anorexia nervosa. Cognitive behavior therapy and interpersonal psychotherapy are effective in a related disorder, bulimia nervosa. There are theoretical and treatment indications for these therapies in anorexia nervosa.

Method: Fifty-six women with anorexia nervosa diagnosed by using strict and lenient weight criteria were randomly assigned to three treatments. Two were specialized psychotherapies (cognitive behavior therapy and interpersonal psychotherapy), and one was a control treatment combining clinical management and supportive psychotherapy (nonspecific supportive clinical management). Therapy consisted of 20 sessions over a minimum of 20 weeks.

Results: For the total study group (intent-to-treat group), there were significant differences among therapies in the primary global outcome measure. Nonspecific supportive clinical management was superior to interpersonal psychotherapy, while cognitive behavior therapy was intermediate, neither worse than nonspecific supportive clinical management nor better than interpersonal psychotherapy. For the women completing therapy, nonspecific supportive clinical management was superior to the two specialized therapies.

Conclusions: The finding that nonspecific supportive clinical management was superior to more specialized psychotherapies was opposite to the primary hypothesis and challenges assumptions about the effective ingredients of successful treatments for anorexia nervosa.

Many randomized, controlled trials have examined psychological therapies for adults with eating disorders, most commonly bulimia nervosa. Cognitive behavior therapy has been established as the treatment of choice for bulimia nervosa (1). Interpersonal psychotherapy may also be an effective treatment in this group (2), although improvement is less rapid than with cognitive behavior therapy. There is evidence that these specialized psychotherapies are superior to other forms of psychotherapy (3).

Anorexia nervosa is a serious illness with substantial morbidity (4) and a mortality rate of approximately 5% per decade (5). Outpatient therapies are widely used in its treatment. However, few randomized, controlled trials have examined treatments for anorexia nervosa. Favorable results have been achieved by using family-based interventions with adolescents (6, 7). While some studies with adults have shown that specialized psychotherapies are superior to no treatment (8) or to a control condition of “routine” treatment (9), others have found no difference between two or more treatments (10, 11). The specific therapies studied include cognitive behavior therapy (11, 12), focal psychoanalytic psychotherapy (9), cognitive analytic therapy (9, 10), dietary counseling (12, 13), individual supportive therapy (14), and family therapy (9, 14). Taken together, these studies have not demonstrated clear superiority of any specific therapy for anorexia nervosa. More trials are needed, given the dearth of studies and the known difficulties with recruitment and retention of patients with this disorder (15).

Cognitive behavior therapy is a theoretically relevant treatment for anorexia nervosa (16, 17) with some evidence of effectiveness (17, 18). Although interpersonal psychotherapy has not been the focus of clinical trials in anorexia nervosa, the condition is readily conceptualized within the framework of interpersonal psychotherapy (19). In order to examine the effectiveness of these two psychotherapies in treating anorexia nervosa, we compared them with a control treatment called “nonspecific supportive clinical management.” This treatment provided education and nutritional advice and used supportive psychotherapy principles in responding to the patient. It was hypothesized that the two specialized psychotherapies would be more effective than nonspecific supportive clinical management.

Method

Participants

The inclusion criteria for this study were female gender, an age of 17–40 years, and the presence of current primary anorexia nervosa; the participants included individuals diagnosed according to the DSM-IV weight criterion (body mass index, <17.5), which was considered to be a strict definition of anorexia nervosa, and
those diagnosed according to a lenient weight criterion (body mass index, 17.5–19.0) (20). Individuals with a body mass index below 14.5 were considered unsuitable for outpatient psychotherapy and were referred for assessment at an inpatient unit. In light of debate as to the necessity of amenorrhea in diagnosing anorexia nervosa (21), amenorrhea was not an inclusion requirement.

The exclusion criteria were current severe major depression, psychoactive substance dependence, major medical or neurological illness, developmental learning disorder, cognitive impairment, bipolar I disorder, schizophrenia, or a chronic, refractory medical illness, developmental learning disorder, cognitive impairment. The participants completed self-report questionnaires, including the Eating Disorder Inventory—2 (25). Physical measurements were made and included height, weight, and percentage body fat. The participants were seen by the team’s dietician for nutritional counseling and food restriction and avoidance, and learned the cognitive behavior therapy skills of challenging dysfunctional thoughts and thought restructuring. Psychoeducational material was provided as take-home handouts. Phase three prepared the patient for termination, providing information on relapse and recovery and teaching strategies to reduce the risk of relapse.

**Procedure**

The structured assessment consisted of the Structured Clinical Interview for DSM-IV (22) to determine the presence of axis I disorders, the Hamilton Depression Rating Scale (23), the Eating Disorder Examination (24), the Global Assessment of Functioning (GAF) (DSM-IV, p. 32), and additional questions relating to eating and weight history, previous treatment, and demographic variables. For the primary outcome measure we developed a global anorexia nervosa measure using a four-point ordinal scale: 4=meets full criteria for the anorexia nervosa spectrum, 3=not full anorexia nervosa but having a number of features of eating disorders, 2=few features of eating disorders, and 1=no significant features of eating disorders.

The participants completed self-report questionnaires, including the Eating Disorder Inventory—2 (25). Physical measurements were made and included height, weight, and percentage body fat. The participants were seen by the team’s dietician for one or two sessions before commencing therapy.

**Therapies**

Therapy in each modality consisted of 20 hour-long manual-based sessions conducted over a minimum of 20 weeks. The therapists (including V.V.W.M., J.J., and S.E.L.) were experienced in treating individuals with eating disorders. All therapists delivered all three psychotherapies.

**Cognitive behavior therapy.** Cognitive behavior therapy for anorexia nervosa is based on the premise that the core features of anorexia nervosa, food restriction and avoidance, become entrenched habit patterns, independent of the circumstances that initiated them (26). Phase one of therapy introduced cognitive behavior therapy and its rationale and the core techniques of self-monitoring and homework. The patient’s motivation to engage in treatment was evaluated, and her ambivalence about giving up anorexia nervosa was addressed. Normal eating was prescribed, and a weight range goal was negotiated. In phase two the patient learned the cognitive behavior therapy skills of challenging dysfunctional thoughts and thought restructuring. Psychoeducational material was provided as take-home handouts. Phase three prepared the patient for termination, providing information on relapse and recovery and teaching strategies to reduce the risk of relapse.

**Interpersonal psychotherapy.** Interpersonal psychotherapy for anorexia nervosa is based on interpersonal psychotherapy for both depression (27) and bulimia nervosa (28). First, a history was taken of the patient’s life events, interpersonal relationships, and eating problems, and links among these were highlighted. Major problems were identified within the four interpersonal psychotherapy problem areas of grief, interpersonal disputes, role transitions, and interpersonal deficits. The second phase focused on the identified problem(s). In interpersonal psychotherapy for depression, each session includes an assessment of symptoms in order to make links between interpersonal issues and depressive symptoms. Fairburn’s interpersonal psychotherapy for bulimia nervosa (28) avoided discussion of bulimia nervosa symptoms. In the present study the patient’s presentation of eating disorder symptoms was used to facilitate work on the agreed interpersonal problem, but a systematic symptom review was not conducted. The final termination phase prepared the patient for independent coping.

**Nonspecific supportive clinical management.** Nonspecific supportive clinical management was developed for the present study, and its aim was to mimic outpatient treatment that could be offered to individuals with anorexia nervosa in usual clinical practice. It combined features of clinical management (29) and supportive psychotherapy (30). Clinical management includes education, care, and support (31) and fostering a therapeutic relationship that promotes adherence to treatment. Supportive psychotherapy aims to assist the patient through use of praise, re-
assurance, and advice. The abnormal nutritional status and dietary patterns typical of anorexia nervosa were central to nonspecific supportive clinical management, which emphasized the resumption of normal eating and the restoration of weight and provided information on weight maintenance strategies, energy requirements, and relearning to eat normally. Information was provided verbally and as written handouts. Other therapy content was dictated by the patient, with the therapist constrained to avoid specific strategies or foci of interpersonal psychotherapy or cognitive behavior therapy.

**Treatment integrity.** The psychotherapy sessions were audiotaped, and three randomly selected audiotapes for each patient were reviewed for adherence to the therapy protocol, which was measured by using an adaptation of the Collaborative Study Psychotherapy Rating Scale. The mean scores for each therapeutic modality were appropriately elevated, relative to the other modalities, on its respective subscale of the Collaborative Study Psychotherapy Rating Scale, indicating satisfactory adherence to therapy.

**Assessments.** Each participant completed assessments after her 10th therapy session and after her final session. The assessments consisted of physical measurements, body image assessment, self-report questionnaires, and a structured clinical interview with a clinician unaware of the participant's treatment condition. The interview included the global anorexia nervosa measure, Eating Disorder Examination, Hamilton depression scale, and GAF. The patient also completed the Eating Disorder Inventory—2.

**Data Analyses**

The outcome measures were determined a priori. The primary outcome variable was the global anorexia nervosa rating. Secondary outcome variables addressed aspects of anorexia nervosa as assessed by the Eating Disorder Examination and Eating Disorder Inventory, weight, body mass index, and other physical measures. Tertiary outcome variables addressed overall functioning and symptoms of disorders other than anorexia nervosa, such as depression.

All analyses were performed by using SPSS. Intent-to-treat analyses were conducted by using the last observation carried forward when data were missing for noncompleters. Values on the global anorexia nervosa measure were compared among treatment groups by using the nonparametric Kruskal-Wallis test. Pairwise comparisons among groups were made by using the Mann-Whitney U test. For secondary and tertiary outcome measures, repeated-measures analysis of variance was used to compare the treatment groups on the change between baseline and the end of treatment. When there was a significant difference among groups, pairwise least significant difference tests were used to compare pairs of means. The independence of the treatment effect from the effects of baseline differences was examined by using logistic regression. The criterion used for statistical significance was $p<0.05$ in two-tailed tests.

**Results**

Publicity resulted in over 400 inquiries. Of the 135 individuals who were interviewed, 78 were eligible. Fifty-six of these consented to participate in the study. At randomization the three treatment groups were similar on most measures, except that there were significant differences among the groups in the rates of lifetime comorbid diagnoses of bulimia nervosa and panic disorder. In the cognitive behavior therapy group 63% (N=12) had a lifetime history of bulimia nervosa, compared with 31% (N=5) and 19% (N=4) in the nonspecific supportive clinical management and interpersonal psychotherapy groups. For panic disorder, the lifetime rates were 26%, 19%, and 0% for the groups receiving cognitive behavior therapy (N=5), nonspecific supportive clinical management (N=3), and interpersonal psychotherapy (N=0), respectively.

Of the 56 women randomly assigned to treatment, 35 completed therapy, defined as attending at least 15 of the 20 therapy sessions. Among the 21 women who did not complete treatment, dropouts occurred over the course of therapy; the median number of sessions completed was 7, with a range of 1–14. Four of these participants were hospitalized for weight loss or complications of anorexia nervosa (after completing two, seven, nine, and 12 sessions; three were assigned to interpersonal psychotherapy, and one was assigned to nonspecific supportive clinical management). One hospitalized patient later died, probably from medical complications of anorexia nervosa. Comparing the participants who did and did not complete treatment revealed a significant difference in mean baseline weight (46.9 kg and 45.6 kg, respectively). On all other variables at baseline the two groups were not significantly different. However, the four women who were hospitalized had a lower weight (mean=41.7 kg), percentage body fat, and GAF score than all other participants.

**Relationship of Global Outcome to Other Measures**

The primary outcome measure was the global anorexia nervosa rating. Participants with a global rating of 4 were minimally improved from baseline in terms of weight, body mass index, and Eating Disorder Examination measures, but they reported lower symptom levels on the Eating Disorder Inventory. Those with a global rating of 3 had gained weight, had minimal changes on the Eating Disorder Examination, and reported considerable symptoms on the Eating Disorder Inventory. Those with a global rat-
TABLE 2. Secondary and Tertiary Outcomes for 56 Women With Anorexia Nervosa in a 20-Week Comparison of Cognitive Behavior Therapy, Interpersonal Psychotherapy, and Nonspecific Supportive Clinical Management

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline (N=56)</th>
<th>Cognitive Behavior Therapy (N=19)</th>
<th>Interpersonal Psychotherapy (N=21)</th>
<th>Nonspecific Supportive Clinical Management (N=16)</th>
<th>Analysis of Variance in Change From Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Weight [kg]</td>
<td>46.4</td>
<td>3.9</td>
<td>48.6</td>
<td>5.5</td>
<td>49.0</td>
</tr>
<tr>
<td>Body mass index [kg/m²]</td>
<td>17.3</td>
<td>1.1</td>
<td>18.1</td>
<td>1.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>18.9</td>
<td>3.4</td>
<td>22.0</td>
<td>5.3</td>
<td>20.7</td>
</tr>
<tr>
<td>Eating Disorder Examination scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrainta</td>
<td>3.9</td>
<td>1.3</td>
<td>2.8</td>
<td>1.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Eating concerns</td>
<td>2.8</td>
<td>1.3</td>
<td>1.7</td>
<td>1.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Weight concerns</td>
<td>3.1</td>
<td>1.7</td>
<td>2.4</td>
<td>1.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Shape concerns</td>
<td>3.8</td>
<td>1.3</td>
<td>2.7</td>
<td>1.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Eating Disorder Inventory—2 scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drive for thinness</td>
<td>11.7</td>
<td>5.4</td>
<td>7.9</td>
<td>6.5</td>
<td>9.5</td>
</tr>
<tr>
<td>Bulimia</td>
<td>3.1</td>
<td>4.0</td>
<td>1.5</td>
<td>4.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Body dissatisfaction</td>
<td>7.7</td>
<td>7.0</td>
<td>5.8</td>
<td>6.9</td>
<td>7.3</td>
</tr>
<tr>
<td>Ineffectiveness</td>
<td>8.3</td>
<td>6.5</td>
<td>5.4</td>
<td>5.4</td>
<td>6.7</td>
</tr>
<tr>
<td>Perfectionism</td>
<td>7.2</td>
<td>3.0</td>
<td>5.4</td>
<td>3.5</td>
<td>7.3</td>
</tr>
<tr>
<td>Disturb</td>
<td>2.6</td>
<td>3.4</td>
<td>2.6</td>
<td>3.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Interpersonal awareness</td>
<td>10.4</td>
<td>6.0</td>
<td>6.4</td>
<td>4.0</td>
<td>6.7</td>
</tr>
<tr>
<td>Maturity fears</td>
<td>3.6</td>
<td>3.8</td>
<td>2.1</td>
<td>2.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Asceticism</td>
<td>7.0</td>
<td>4.8</td>
<td>5.4</td>
<td>3.5</td>
<td>4.2</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>5.9</td>
<td>4.4</td>
<td>3.3</td>
<td>5.8</td>
<td>4.4</td>
</tr>
<tr>
<td>Insecurity</td>
<td>3.2</td>
<td>4.2</td>
<td>3.6</td>
<td>5.1</td>
<td>3.2</td>
</tr>
<tr>
<td>General psychopathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Assessment of Functioning scoreb</td>
<td>48.8</td>
<td>5.6</td>
<td>53.2</td>
<td>9.5</td>
<td>51.1</td>
</tr>
<tr>
<td>Hamilton Depression Rating Scale score</td>
<td>12.6</td>
<td>6.9</td>
<td>6.9</td>
<td>7.8</td>
<td>9.9</td>
</tr>
</tbody>
</table>

*Significant overall difference (p<0.05) among groups (analysis of variance)*; Post hoc tests indicated that cognitive behavior therapy and nonspecific supportive clinical management were each superior to interpersonal psychotherapy (p<0.05, pairwise least significant difference tests).

*Significant overall difference (p<0.05) among groups (analysis of variance)*; Post hoc tests indicated that nonspecific supportive clinical management was superior to each of the other two therapies (p<0.05, pairwise least significant difference tests).

Primary Outcome

Primary outcome data are shown in Table 1. The percentages of individuals completing therapy and the time to discontinuation for the three therapeutic modalities were not significantly different. The main reasons for discontinuing treatment also did not differ across the three therapy groups: one subject from each group moved from the area, four improved (cognitive behavior therapy, N=1; interpersonal psychotherapy, N=2; nonspecific supportive clinical management, N=1), seven disliked treatment (cognitive behavior therapy, N=3; interpersonal psychotherapy, N=2; nonspecific supportive clinical management, N=2), and three stopped coming to treatment sessions and assessments (cognitive behavior therapy, N=2; interpersonal psychotherapy, N=1; nonspecific supportive clinical management, N=0).

Intent-to-treat analyses were conducted by using the last observation carried forward for noncompleters. Patients who dropped out before the midtreatment assessment were given a rating of 4, and the rating from the (blind) assessment at midtreatment or end of treatment was used for those who had completed one of these assessments. Significant differences in the primary outcome were found among all participants across therapies. Mann-Whitney U tests showed that nonspecific supportive clinical management was superior to interpersonal psychotherapy. Cognitive behavior therapy and interpersonal psychotherapy did not differ significantly from one another, nor did cognitive behavior therapy and nonspecific supportive clinical management. For the total study group, a global rating of 1 or 2 was achieved in two (10%) of the 21 patients receiving interpersonal psychotherapy, six (32%) of the 19 receiving cognitive behavior therapy, and nine (56%) of the 16 assigned to nonspecific supportive clinical management.

There were significant differences in primary outcome among completers across therapies. Mann-Whitney U tests showed that nonspecific supportive clinical management was superior to cognitive behavior therapy and to interpersonal psychotherapy, while cognitive behavior therapy and interpersonal psychotherapy did not differ significantly from one another. Among those completing therapy, a global rating of 1 or 2 was achieved in two (17%) of the 12 interpersonal psychotherapy patients, five (42%) of the 12 patients completing cognitive behavior therapy, and nine (82%) of the 11 receiving nonspecific supportive clinical management.
Secondary and Tertiary Outcomes

Table 2 shows mean values for the secondary and tertiary outcome measures at the end of treatment for the total study group. There were no significant differences by therapy on any of the physical measures, including weight, body mass index, and percentage body fat. On the Eating Disorder Examination restraint subscale, cognitive behavior therapy and nonspecific supportive clinical management were both superior to interpersonal psychotherapy and did not differ from one another. On the GAF, nonspecific supportive clinical management was superior to both cognitive behavior therapy and interpersonal psychotherapy, which did not differ from one another.

For those completing therapy, there was no difference by therapy group on weight outcome. There were significant differences on all four Eating Disorder Examination subscales and on the drive for thinness subscale of the Eating Disorder Inventory. For restraint and shape concerns (Eating Disorder Examination) and drive for thinness (Eating Disorder Inventory), post hoc tests showed that nonspecific supportive clinical management was superior to interpersonal psychotherapy, while for eating concerns and weight concerns (Eating Disorder Examination), both nonspecific supportive clinical management and cognitive behavior therapy were superior to interpersonal psychotherapy. On the GAF, nonspecific supportive clinical management was superior to both other treatments, which did not differ from one another.

Confounding Factors

Logistic regression was used with a dichotomized form of the global anorexia nervosa rating as the dependent measure to examine whether differences in outcome among therapy groups could be explained by baseline differences (p=0.10) among treatment groups. The analyses were conducted with treatment as a binary comparison (cognitive behavior therapy or interpersonal psychotherapy versus nonspecific supportive clinical management) and as a three-way comparison. Differences in outcome among the treatment groups were not explained by any difference among treatment groups at baseline (p=0.050 for the binary comparison, p=0.033 for the three-way comparison).

Discussion

To our knowledge, this study is the first randomized, controlled trial to compare cognitive behavior therapy and interpersonal psychotherapy with nonspecific supportive clinical management for adults with anorexia nervosa. At the end of treatment, 9% of the subjects had a very good outcome and a further 21% had improved considerably; 70% either did not complete treatment or made small or no gains. Contrary to our hypotheses, the patients who received nonspecific supportive clinical management had an outcome as good as or better than the outcomes of those who received specialized psychotherapies, with 56% of the participants who received nonspecific supportive clinical management given a global rating of 1 or 2, compared with 32% and 10% of those receiving cognitive behavior therapy and interpersonal psychotherapy, respectively. On a number of secondary outcome measures, the outcomes for cognitive behavior therapy were intermediate between the results for nonspecific supportive clinical management and interpersonal psychotherapy.

Inclusion of patients in the lenient weight range resulted in a study group that differed from the subjects in many other studies. However, the subgroups based on strict and lenient weight criteria were indistinguishable on almost all measures of severity at baseline (20), and inclusion of strict/lenient as a covariate in the logistic regression did not explain the differences in outcome across therapies.

Other methodological issues include the relatively short duration of therapy and the relatively small number of subjects. Although the study group was smaller than the groups in clinical trials for bulimia nervosa or major depression, in the eight existing randomized, controlled trials for adults with anorexia nervosa of which we are aware (8–12, 14, 18, 36), the numbers of subjects were modest (range=24–90, mean=50.7, median=34.0), reflecting difficulties with recruitment and retention, both of which are features of anorexia nervosa trials (15). These methodological limitations, however, would not account for the differential treatment outcome. Further, the differences among the treatment groups in rates of comorbidity were not found to contribute to the differential treatment outcome.

Interpersonal psychotherapy was the least effective of the three therapies. Possible explanations include the relative lack of symptom focus, the relatively long time taken to decide on the problem area, thus reducing the length of the middle phase of interpersonal psychotherapy, and the lack of reactivity of the symptoms of anorexia nervosa to interpersonal triggers, which may hinder exploration of links between interpersonal issues and eating disorder symptoms. Interpersonal events may be less rewarding for or more avoided by anorexia nervosa patients, and thus a personal rationale for interpersonal psychotherapy may be missing for this patient group at this stage of illness. However, we found no difference across the three therapies in patients’ ratings of treatment credibility or satisfaction, indicating that differences in how patients perceived the three therapies did not account for the differences among therapy groups. Therapy had a relatively early focus on affect and interpersonal functioning, which may be difficult for anorexia nervosa patients.

Possible reasons for poor outcomes with cognitive behavior therapy include the large amount of psychoeducational material and extensive skills acquisition, inability to generate alternatives to fixed cognitions stemming from the cognitive rigidity of anorexia nervosa patients, and the ego-syntonic nature of anorexia nervosa, causing difficulty in actively working toward change in a direct therapy.
such as cognitive behavior therapy. Cognitive behavior therapy may be less effective for patients with anorexia nervosa, who have high rates of obsessional personality traits (37), as has been found in depressed patients (38). It may be more effective after an initial weight gain.

It may be, however, that nonspecific supportive clinical management is a particularly suitable treatment for acute anorexia nervosa. Both clinical management and supportive psychotherapy have been found to be highly effective treatments with other disorders (39, 40). In this study it was delivered by clinicians experienced with eating disorders, was practiced according to a detailed treatment manual, and included the potentially powerful components of psychoeducation and a focus on normalizing eating (shared with cognitive behavior therapy). Thus, the therapy included detailed discussion of ways to increase food choices and quantities in order to gain weight. As the remaining session content was based on issues the patient chose to present, nonspecific supportive clinical management may allow an increased sense of patient autonomy and control. A key feature of nonspecific supportive clinical management may be the important nonspecific factors of psychotherapy: the therapeutic alliance, empathy, positive regard, and support for a patient group greatly in need of these.

To our knowledge, this is the first comparison of cognitive behavior therapy and interpersonal psychotherapy with a control therapy for adults with anorexia nervosa. Some previous randomized, controlled trials for anorexia nervosa have included aspects of nonspecific supportive clinical management (9, 11, 13) but with important differences: the therapists were not eating disorders specialists (9, 11), therapy was disrupted when therapists left the service (9), and the sessions were shorter and fewer than in the other therapies (9, 11). In the present study, the proportion of patients with a very good or good outcome (30%) was similar to that in other studies (9, 10). Hall and Crisp’s comparison of dietary advice and psychotherapy (13) showed improvement in both groups, with the group receiving dietary advice having greater weight gain.

A stepped approach has been advocated in the treatment of eating disorders (26), and nonspecific supportive clinical management may be a good first phase of treatment. Further research is needed to identify which patients will respond to a particular first-line treatment. It is not known whether some therapy models are better for some individuals or whether individuals who fail to respond to one therapy will also fail to respond to another therapy model.

In conclusion, the results of the present study caution against assuming that more specialized psychotherapies are necessarily more effective in the clinical management of anorexia nervosa. As this appears to be the first time that nonspecific supportive clinical management has been the subject of a treatment trial for anorexia nervosa, replication of the study is necessary. However, the outcome in the present study indicates that nonspecific supportive clinical management may be a valuable treatment for adults with anorexia nervosa.

References

31. Joyce PR: The clinical management of depression, in Affective Disorders. Edited by Joyce PR, Romans SE, Ellis PM, Silverstone TS. Christchurch, New Zealand, Christchurch School of Medicine, Department of Psychological Medicine, 1995, pp 35–45
33. Evans MD, Piasecki JM, Kriss MR, Hollon SD: Raters’ Manual for the Collaborative Study Psychotherapy Rating Scale—Form 6 (CSPRS-6). Minneapolis, University of Minnesota and St Paul-Ramsey Medical Center, 1984
35. SPSS: Statistical Package for the Social Sciences (Version 10.0.5). Chicago, SPSS, 1999
40. Treatment Protocol Project: Management of Mental Disorders. Darlinghurst, New South Wales, Australia, World Health Organisation Collaborating Centre for Mental Health and Substance Abuse, 1997
The Incidence of Anorexia Nervosa on Curaçao

Hans W. Hoek, M.D., Ph.D.
Peter N. van Harten, M.D., Ph.D.
Karin M.E. Hermans, M.D.
Melanie A. Katzman, Ph.D.
Glenn E. Matroos, M.D.
Ezra S. Susser, M.D., Dr.P.H.

Objective: Although anorexia nervosa was once thought to occur only in affluent societies, cases have now been documented across the globe. To examine whether anorexia nervosa emerges in societies undergoing socioeconomic transition, the authors studied the incidence of anorexia nervosa on the Caribbean island of Curaçao.

Method: The authors contacted the full range of community health and service providers on Curaçao, including dietitians, school counselors, and all 82 general practitioners. They also studied inpatient records for 84,420 admissions to Curaçao General Hospital and two private hospitals in 1995–1998. Probable-incident subjects were interviewed.

Results: The incidence rates in 1995–1998 per 100,000 person-years for anorexia nervosa on Curaçao were 1.82 (95% confidence interval [CI]=0.74–2.89) for the total population and 17.48 (95% CI=4.13–30.43) for the high-risk group of 15–24-year-old females. No cases were found among the majority black population. For the Curaçao mixed and white population, the incidence rate per 100,000 person-years for anorexia nervosa was 9.08 (95% CI=3.71–14.45).

Conclusions: The overall incidence of anorexia nervosa on Curaçao is much lower than in the affluent societies of the United States and Western Europe. Within Curaçao, sociocultural factors appear to be associated with differential incidence rates of anorexia nervosa. The incidence of anorexia nervosa among the majority black population is nil, while the incidence among the minority mixed and white population on Curaçao is similar to that of the United States and the Netherlands.

Anorexia nervosa, once thought to occur only in affluent societies, has now been documented across the globe (1). Although neurobiological factors have been implicated in the emergence of anorexia nervosa, there is evidence to suggest that social factors play an important role as well (2–7). The disorder appears to be related to socioeconomic development, changing roles for women, a sociocultural emphasis on thinness, and a shift in eating patterns (8). Studies have found evidence of changing attitudes about ideal body size and eating patterns accompanying the introduction of television and other methods of merchandizing to increasingly affluent societies (9).

Societies undergoing socioeconomic transition also experience an epidemiological transition as life expectancy increases and noncommunicable diseases become predominant. It is not yet known whether this epidemiological transition will include the emergence of anorexia nervosa, as its incidence has not previously been studied outside the most highly developed societies (1). We report here on the incidence of anorexia nervosa in a society undergoing socioeconomic transition: the Caribbean island of Curaçao.

We hypothesized that like other so-called diseases of civilization, anorexia nervosa in newly affluent societies should be approaching the incidence levels of high-income countries, such as the Netherlands and the United States. Reports have documented that at least some cases of anorexia nervosa are now occurring in societies across the globe, including Curaçao (1, 10).

Curaçao presented a unique opportunity to examine this question. Curaçao is the main island in a group of five islands collectively known as the Netherlands Antilles and is still part of the kingdom of the Netherlands. Like many Caribbean islands, the society has origins in plantation slavery. Curaçao is now one of only a few of these islands to be classified by the World Bank as a high-income country, albeit the least affluent within that grouping. According to the 2001 Curaçao census (11), 23% of those 15 years and older had no income, compared to 29% in 1992 (12). Dutch and U.S. television is widely available, and other cultural influences of North America and Europe are increasingly evident. Moreover, the island also provided the opportunity to compare incidence across sociocultural groups. Like many transitional societies, Curaçao includes a well-to-do minority subgroup, which adopts many of the cultural norms of affluent societies, as well as a relatively poor majority.

Curaçao is particularly well suited for epidemiological studies. The vast majority of health professionals are trained in the Netherlands. The comprehensive health care system and population databases (modeled on the Dutch systems) made it possible to identify incident cases...
as well as to compute the denominator for the incidence rate. To our knowledge, the present study is more comprehensive than any previous incidence study on anorexia nervosa.

**Method**

This study on the incidence of anorexia nervosa was conducted with the approval of the Netherlands Antilles Foundation for Clinical Higher Education, which, in association with Groningen University Hospital in the Netherlands, was responsible for overseeing medical research conducted on Curaçao. All patients interviewed signed informed consent forms.

At the beginning of the study in 1995, Curaçao had a population of 1,49,903 (13). Census data indicate that in 2001, 82% of the population was born on Curaçao and another 6% on one of the other islands of the Netherlands Antilles (11). Papiamentu, the local language, was spoken at home in the vast majority (81%) of the households. The official language was Dutch and was used in schools, but it ranked second as a spoken language. The census indicated that in 2001, 82% of the population spoke Dutch, but it ranked second as a spoken language. According to the Curaçao Health Study (N=2,248) in 1993/1994, 79% of the population of Curaçao was identified as black, 13% as mixed, 7% as white, and 1% as Asian (14). This study also revealed rates of obesity (body mass index >30 kg/m²) among the Curaçao population in 1993/1994 (15); 18.7% of the Curaçao men were obese, compared to 20.6% of the men in the United States in 1988–1994 and 27.7% in 1999–2000 (16). The obesity rate of 36.2% for Curaçao women was higher than the obesity rate of 25.9% for U.S. women in 1988–1994 and 34% in 1999–2000. The mean body mass index for women on Curaçao is similar to that of black women and much higher than that of white women in the United States (28.3, 28.2, and 26.0, respectively) (10, 14–16). Of the obese Curaçao women with low socioeconomic status, only half considered themselves to be too fat (15).

**Sources**

We used both record reviews and personal contacts with service providers to identify “possible” cases of eating disorders. In order to encompass all possible cases, we systematically identified possible cases in both outpatient and inpatient settings. To identify possible cases in outpatient settings, beginning in 1998, we contacted the full range of community health and mental health clinics that might see patients with eating disorders (Table 1). In addition, we contacted relevant individual health care professionals: all 82 general practitioners, seven of eight psychologists on the island, two psychiatrists with a private practice, seven of eight pediatricians, three gynecologists, and three doctors of internal medicine. All 10 practicing dietitians were willing to provide information. We also informed the three school counselors at the higher grade schools we could find on the island about the study. One of them was a graduate-level psychologist with an official counseling function at the University College of Netherlands Antilles on Curaçao, and the other two were teachers with additional tasks as counselors.

We provided all sources with systematic information about eating disorders and asked them to inform us about any possible cases first seen during 1995–1998. We defined “possible” cases in accordance with a previous general practice study in the Netherlands (17) and communicated this definition in writing and in person when feasible.

With respect to inpatients in the three existing Curaçao general hospitals, we reviewed the records of all admissions to the Curaçao Saint Elisabeth Hospital and the two smaller private general hospitals for the period 1995–1998 (Table 2). We screened these records for an ICD-9 diagnosis of an eating disorder, as well as for 26 diagnostic terms that might indicate missed cases. These diagnostic terms were the same as those used in our previous study (10) and by Lucas and colleagues (18) in the United States and relate to weight loss, vomiting, or other gastrointestinal problems or menstrual problems. If any of these diagnostic terms had been used on a patient’s chart, an eating disorder expert from the research team (K.M.E.H. or H.W.H.) reviewed the chart, and if the record suggested any symptom or sign of an eating disorder, then it was a “possible” case.

**Patients**

Using this method, we identified 52 outpatients and 12 inpatients (Table 1 and Table 2) as “possible” cases, representing 47 individuals (some patients were identified by more than one source). The other physicians involved were contacted and interviewed to determine which of these 47 were “probable” cases. A probable case was one for which there was suggestive evidence that the case might meet the diagnostic criteria of DSM-IV for an eating disorder. Such cases were verified as having been first detected in 1995–1998. Twenty-four cases met these criteria.

With regard to the 24 probable cases, health care professionals located the subjects and obtained their written informed consent for our team to conduct a diagnostic interview. One subject did not give informed consent and was not included in the study. Another woman could not be traced, but the physician was able to provide detailed information. Of the remaining 22 probable cases, 18 subjects could be interviewed in person; the other four subjects were living abroad, but it proved possible to interview a first-degree relative for each of the four on Curaçao.

After written informed consent had been obtained, the subjects were extensively interviewed and assessed by trained interviewers, according to the eating disorders section of the Structured Clinical Interview for DSM-IV (SCID) (19). In addition to the SCID items, we used further assessments developed by the authors as an aid to eliciting the information required for diagnosis. Current weight and height were measured in a standardized way with the official hospital scales. The body mass index at the time of detection has been based on weight and height data from med-

---

**TABLE 1. Possible Cases of Anorexia Nervosa in Curaçao Outpatients in 1995–1998**

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Possible Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capriles Psychiatric Hospital</td>
<td>10</td>
</tr>
<tr>
<td>Community Mental Health Center</td>
<td>4</td>
</tr>
<tr>
<td>Public Health Service for Adolescents</td>
<td>3</td>
</tr>
<tr>
<td>Private practitioners (N=9)</td>
<td>6</td>
</tr>
<tr>
<td>Medical specialists (N=13)</td>
<td>3</td>
</tr>
<tr>
<td>General practitioners (N=62)</td>
<td>7</td>
</tr>
<tr>
<td>School counselors (N=3)</td>
<td>6</td>
</tr>
<tr>
<td>Dietitians (N=10)</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
</tr>
</tbody>
</table>


**TABLE 2. Possible Cases of Anorexia Nervosa in Curaçao Inpatients in 1995–1998**

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Number of Admissions</th>
<th>Screened</th>
<th>Possible Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saint Elisabeth Hospital</td>
<td>540</td>
<td>68,475</td>
<td>220</td>
</tr>
<tr>
<td>Taams Hospital</td>
<td>36</td>
<td>11,137</td>
<td>85</td>
</tr>
<tr>
<td>Advent Hospital</td>
<td>42</td>
<td>4,808b</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>618</td>
<td>84,420</td>
<td>396</td>
</tr>
</tbody>
</table>

*The records of Capriles Psychiatric Hospital were also studied but produced no possible cases in the inpatient units because this hospital mainly provides care for psychotic patients; see Table 1 for its outpatient cases.

*Estimated from number of admissions in 1995 (N=1,202).
ical records. Eating disorder experts from the research team (H.W.H., K.M.E.H., and M.A.K.) conducted the diagnostic interviews. The research interviewer made the diagnosis, but in all cases in which there were uncertainties, the case was discussed or reinterviewed by another researcher, and the final diagnosis was a consensus diagnosis arrived at by two members of the research team. For the four probable cases that could not be directly assessed, we reviewed all available information and made a consensus diagnosis.

**Data Analysis**

The denominators for the incidence rates for anorexia nervosa and 95% confidence intervals (CIs) were computed for the total population (605,708 person-years in 1995–1998) as well as for females (320,651 woman-years) and the high-risk group of females ages 15–24 years (40,041 woman-years). These denominators were provided by the Central Bureau of Statistics, which computed their data based on the census in 1992 (12) and 2001 (11) and the migration rates (13). Since the census data are not reported by race, to obtain the denominator for mixed/white race incidence, we multiplied the denominator for total incidence by the proportion that was mixed/white according to the Curação Health Study (15).

**Results**

Based on the interviews and the case records, we identified 11 incident cases fulfilling the DSM-IV criteria for anorexia nervosa and first detected in the study period 1995–1998 (Table 3). Six other possible cases were not included in the study because the diagnosis in three cases had been considered before 1995 and in three others for the first time in 1999.

Four of the 11 incident cases of the study period were the restrictor type of anorexia nervosa and seven the purging type. All subjects had serious weight loss; their mean body mass index at the time of detection was 15.2 kg/m² (range=12.0–18.2). By definition, all subjects had amenorrhea (one of the criteria in DSM-IV).

All 11 were female. Two of the patients were white and were born in the Netherlands. The other nine were of mixed race; there were no cases among blacks. The cases were 15, 18, 20, 22, 23, 24, 24, 26, 27, 27, and 39 years old at the time of detection. At the time of onset, nine of the subjects were 15 to 22; for one, the age at onset was unknown, and one was reported to be 36 at the time of onset.

Table 4 shows that 55% of the women with anorexia nervosa had been to college; according to the Central Bureau of Statistics, less than 10% of all Curação women in their age group had been to college (12). At the time of the interview, seven of 11 subjects (64%) earned more than 3,000 Netherlands Antilles florins ($1,695 U.S.) per month. According to the 2001 census, only 0.6% of all 15–24-year-old and 10.9% of all 25–34-year-old females on Curação earned more than 3,000 Netherlands Antilles florins. In addition, all subjects had been abroad for a year or more; three had been to the United States and eight to the Netherlands. For seven of the subjects with anorexia nervosa, the stay abroad was before the onset of anorexia nervosa; for one, the stay abroad was after the onset; and for three, this could not be determined.

The associated incidence rates per 100,000 person-years for anorexia nervosa on Curação were 1.82 (95% CI=0.74–2.89) for the total population, 3.43 (95% CI=1.40–5.46) for females, and 17.48 (95% CI=4.13–30.43) for the high-risk group of 15–24-year-old females. The associated incidence rates per year per 100,000 person-years for anorexia nervosa for the Curação mixed and white populations only were 9.08 (95% CI=3.71–14.45) for the total mixed/white population and 17.15 (95% CI=7.02–27.29) for mixed/white females.

**Discussion**

This study examined the incidence of anorexia nervosa in a society undergoing a socioeconomic transition. There were two main findings. First, the overall incidence rate for anorexia nervosa of 1.82 per 100,000 person-years on Curação is several times lower than in the United States (8.3 per 100,000) (18) and the Netherlands (8.1 per 100,000) (17). Also, the overall incidence rate of 17.5 per 100,000 person-years for the high-risk group of 15–24-year-old females is much lower than in the United States (48.4 per 100,000) (18) and the Netherlands (55.5 per 100,000) (17).

Second, we did find that sociocultural differences within the island were related to anorexia nervosa. Cases occurred only among the minority white and mixed-race populations. There were no cases among the majority black population, despite our comprehensive case-finding methods. The incidence of anorexia nervosa among the white and mixed-race Curação population (9.1 per 100,000 person-years) was similar to the incidence in the United States and the Netherlands.
We did not find evidence that socioeconomic transition had caused an emergence of anorexia nervosa. The finding of a low overall incidence rate of anorexia nervosa on Curaçao was contrary to what we had hypothesized. We had theorized that as a result of the socioeconomic transition taking place on the island, the incidence of anorexia nervosa would be approaching that of high-income countries such as the Netherlands. One possible explanation is that the social impact of higher incomes is yet to emerge because profound cultural changes may take two or more generations. A second possible explanation is that the socioeconomic transition has been uneven and incomplete. The World Bank has classified the Netherlands Antilles as a high-income country, but it has a high unemployment rate and other features that make it quite different from the United States or the Netherlands. The third and most likely possibility is that the low overall incidence rate could be explained by sociocultural factors. Such an explanation would be in line with the most intriguing finding: the absence of any cases among the majority black population. In the African-Caribbean population of Curaçao, being overweight is socially more accepted than in the white and mixed populations (10, 14, 15). The local norms about body size may have enhanced resiliency to sociocultural emphasis on thinness in the media and other influences from the United States and the Netherlands, and this might have served to protect young women in the majority black population from developing anorexia nervosa.

Our findings may be compared with those in a recent prevalence study in the United States (20) by using a two-stage case-finding method. This U.S. prevalence study found no anorexia nervosa in black women and 15 cases among young white women (lifetime prevalence of 1.5%). In this U.S. prevalence study, on average, the black households had lower family incomes and lower educational levels than the white households. This comparison should, however, be interpreted with much caution, given the differing sociocultural context of Curaçao and the United States.

The high incidence among the white and mixed-race populations—as high as in the Netherlands (17) and in the United States (18)—was also remarkable. One factor that may partly explain this result is that the anorexia nervosa cases belong to a high-income subculture that is similar in many ways to that of affluent societies. High-income subcultures within developing countries do sometimes exhibit a health profile that resembles that of developed countries. This phenomenon has been documented for a long time, for example, among whites in South Africa (21).

Another important contributing factor may be the experience of going abroad. Most of the cases on Curaçao were women who had developed anorexia nervosa after visiting the Netherlands or the United States (usually as students). This parallels reports of eating disturbances in students moving from a developing to a developed country (8, 22, 23).

The study has some limitations. First, despite extensive screening, we may have missed cases of anorexia nervosa. In addition, case-detection methods by the health system may have differed for the more affluent white and mixed-race population compared with the majority black population. Because ethnicity is not registered in the population and health registers, we cannot examine possible differences in help-seeking behavior or clinician bias toward the different ethnic subgroups. However, it is unlikely that our results can be explained by the failure to detect cases among the majority black population. All socioeconomic classes have access to health care, regardless of ability to pay. Cases of bulimia nervosa and binge eating disorder were detected among black women in the course of the anorexia study. It seems implausible that cases of anorexia nervosa were missed, while other eating disorders—that are usually harder to detect (24)—were identified. Although clinician bias cannot be ruled out, clinicians received extensive written and oral information on the study. Because four of the authors currently or previously worked on the island, they were familiar with most physicians on the island. Furthermore, health care workers were alerted to the topic after a British Broadcasting Corporation documentary reported on an earlier investigation of anorexia nervosa on Curaçao (10).

A second limitation, as in all other incidence studies on anorexia nervosa so far, is that we relied mainly on health care providers. The research team trained and educated health care professionals to improve identification and services on the island. The inclusion of health care providers was more comprehensive than in any previous study, and we made an attempt to include cases from three school counselors. We were able to trace some possible cases at higher grade schools. However, we had no sources in primary or secondary schools. To study possibly undetected cases in all Curaçao schools or among the general population would require an additional comprehensive two-stage study with a screening phase and clinical interviews of a large sample.

A third limitation is that in the absence of census information on ethnicity data, we could not examine differences in income and education between the anorexia nervosa cases and the population within the mixed-race group. We compared the income and education of the cases with that of all persons of the same sex and age group in the census data.

We conclude that the socioeconomic transition on Curaçao was not sufficient for the emergence of anorexia nervosa among the majority black population. The incidence among the minority mixed and white population on Curaçao is similar to the incidence in affluent societies as the United States and Western Europe. In contrast to earlier work (10) done on Curaçao that did not include interviews, the current results underscore the powerful role of the sociocultural context in the development of anorexia nervosa (25). It is nonetheless compatible with a neu-
obiological basis for the development of the disorder in individuals (2, 3, 26, 27). One possibility is that the development of anorexia nervosa in most cases requires a combination of a sociocultural environment, which imposes stress on eating behaviors and an individual vulnerability that patterns the response to stress.

Received Oct. 14, 2003; revisions received March 22 and June 18, 2004; accepted Aug. 11, 2004. From Parnassia–The Hague Psychiatric Institute; Dr. D.R. Capriles Hospital, Curacao; the Department of Epidemiology, Mailman School of Public Health, Columbia University, New York; the Department of Psychiatry, University of Groningen, Groningen, the Netherlands; Weill-Cornell Medical Center, New York; New York State Psychiatric Institute, New York; and Symfora Group, Amersfoort, the Netherlands. Address correspondence and reprint requests to Dr. Hoek, Parnassia–The Hague Psychiatric Institute, Mangostraat 15, 2552 KS the Hague, the Netherlands; w.hoek@parnassia.nl (e-mail).

Funded by a grant from the Netherlands Antilles Foundation for Clinical Higher Education.

The authors thank the health care workers of Curacao, especially Tineke Alberts, Ph.D., and Izzy Gerstenbluth, M.D., of the Public Health Service, and Orlando Gerards, M.D., of the Community Mental Health Service, and Daphne van Hoeken, Ph.D., at Parnassia–The Hague Psychiatric Institute for help with data analysis.

References


Fiona Wardell, M.A. (Hons.)
Juliet P. Vyvyan, B.Sc., M.Sc.
Simon A. Naji, B.Sc.
Gordon J. Prescott, B.Sc., M.Sc.
John M. Eagles, M.B., Ch.B., M.Phil., F.R.C.Psych.

Objective: Most previous studies of mortality in anorexia nervosa patients have shown an increased risk of premature death but have been limited by methodological constraints. This study aimed to overcome some of these constraints by having a large original sample size, diagnosis confirmed by case note review, a long duration of follow-up, and a clear base population.

Method: The authors identified 524 anorexia nervosa cases seen in specialist services in Northeast Scotland; anorexia nervosa diagnosis was confirmed by scrutinizing case notes. Those who had died were identified from the National Health Service register or register of deaths. The death rates and causes of death were analyzed.

Results: Twenty-three patients died, giving a crude death rate of 4.4% and a standardized mortality rate of 3.3 (95% CI=2.2–4.9). In only one-third of the cases was anorexia nervosa on the death certificate, but an eating disorder or other psychiatric pathology probably contributed to several of the other deaths. Older age at the time the patient was seen at the specialist service was the only identifiable risk factor in the group of patients who died. The median length of time between diagnosis and death was 11 years.

Conclusions: Anorexia nervosa is associated with increased risk of premature death. It is possible that death rates could be reduced by early diagnosis and by long-term specialist care.

In outcome studies of anorexia nervosa, crude mortality figures are often quoted (1) and show a very wide variation, from 0% to 18% (2, 3). This variation is not surprising considering the very different methodologies used. Initial patient selection can at one extreme be based on national case register information of all diagnosed cases (4), and at the other be based on carefully diagnosed but highly selected patients in a tertiary referral inpatient setting. For the former there may be doubts about diagnostic accuracy; for the latter there may be doubts about how representative the cohort will be of all patients with anorexia nervosa in a given population. Indeed, in some tertiary referral centers, the base population is not known. Follow-up periods have varied from 5 to 33 years (3, 5), and crude mortality figures quoted will naturally rise with longer follow-up periods (i.e., as a cohort ages, more cases will die). Many studies have not attempted to calculate standardized mortality ratios, but increasingly standardized mortality ratios and other standardized methods of reporting mortality are being published (1, 6). Most of these studies have shown a standardized mortality ratio between 4 and 13 for anorexia nervosa, with a few notable outliers reporting very low (2, 7) or high (8) standardized mortality ratios. Another limitation in much of the literature is sample size. For example, the study by Norring and Sohlberg (8) had a sample size of just 25. Most clinically based studies have had fewer than a hundred patients at inception, with four studies having between 100 and 210 patients (7, 9–11); a fifth study included 332 patients (12).

The current study attempts to overcome some of these methodological limitations. It is a long-term follow-up study (up to 35 years) of nearly all patients from a catchment area in Northeast Scotland (Grampian Region). In this locality, all patients referred to National Health Service psychiatric, psychological, and other mental health services—whether inpatient, day patient or outpatient—are identifiable, and there is very limited access to and uptake of private services. All the case notes of suspected patients were checked for diagnostic accuracy, and this produced a large intake cohort of 524 cases.

The aims of the study were to establish the standardized mortality ratio for patients with anorexia nervosa seen in specialist services and to analyze the causes of death.

Method

Patients with a diagnosis of any eating disorder category between 1965 and 1999 were identified from the Grampian psychiatric case register, patient administration systems, and the Eating Disorder Service database. All case notes were checked to confirm or refute a diagnosis of anorexia nervosa. Over these years, diagnostic systems changed from ICD-7 to ICD-10, and case notes did not always contain sufficient information to make a diagnosis of anorexia nervosa according to current ICD-10 or DSM criteria. In older ICD versions there were no codes for eating disorders. However, the Aberdeen Psychiatric Case Register had a category for.
ANOREXIA NERVOSA MORTALITY


<table>
<thead>
<tr>
<th>Case</th>
<th>Year of Diagnosis</th>
<th>Age at Diagnosis (years)</th>
<th>Age at Death (years)</th>
<th>Years From Diagnosis to Death</th>
<th>Cause of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1982</td>
<td>17</td>
<td>23</td>
<td>6</td>
<td>None recorded</td>
</tr>
<tr>
<td>B</td>
<td>1967</td>
<td>26</td>
<td>28</td>
<td>2</td>
<td>None recorded</td>
</tr>
<tr>
<td>C</td>
<td>1969</td>
<td>21</td>
<td>48</td>
<td>27</td>
<td>Unascertainable pending toxicology</td>
</tr>
<tr>
<td>D</td>
<td>1991</td>
<td>19</td>
<td>20</td>
<td>1</td>
<td>Acute myocardial infarction</td>
</tr>
<tr>
<td>E</td>
<td>1989</td>
<td>47</td>
<td>49</td>
<td>2</td>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td>F</td>
<td>1971</td>
<td>16</td>
<td>42</td>
<td>26</td>
<td>Anorexia nervosa</td>
</tr>
<tr>
<td>G</td>
<td>1990</td>
<td>24</td>
<td>28</td>
<td>4</td>
<td>Anorexia nervosa</td>
</tr>
<tr>
<td>H</td>
<td>1984</td>
<td>19</td>
<td>35</td>
<td>16</td>
<td>Bronchopneumonia</td>
</tr>
<tr>
<td>I</td>
<td>1993</td>
<td>34</td>
<td>40</td>
<td>6</td>
<td>Cardiac arrhythmia</td>
</tr>
<tr>
<td>J</td>
<td>1996</td>
<td>28</td>
<td>33</td>
<td>5</td>
<td>Diabetic ketoacidosis</td>
</tr>
<tr>
<td>K</td>
<td>1968</td>
<td>18</td>
<td>50</td>
<td>32</td>
<td>Gastrointestinal hemorrhage</td>
</tr>
<tr>
<td>L</td>
<td>1982</td>
<td>18</td>
<td>33</td>
<td>15</td>
<td>Ligature strangulation</td>
</tr>
<tr>
<td>M</td>
<td>1965</td>
<td>19</td>
<td>52</td>
<td>33</td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>N</td>
<td>1979</td>
<td>19</td>
<td>29</td>
<td>10</td>
<td>Natural causes</td>
</tr>
<tr>
<td>O</td>
<td>1979</td>
<td>20</td>
<td>27</td>
<td>7</td>
<td>Natural causes probably hypoglycemia</td>
</tr>
<tr>
<td>P</td>
<td>1998</td>
<td>21</td>
<td>23</td>
<td>2</td>
<td>Opiate intoxication</td>
</tr>
<tr>
<td>Q</td>
<td>1975</td>
<td>34</td>
<td>38</td>
<td>4</td>
<td>Psychological malfunction arising from mental factors (psychogenic dysmenorrhea)</td>
</tr>
<tr>
<td>R</td>
<td>1982</td>
<td>45</td>
<td>61</td>
<td>16</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>S</td>
<td>1980</td>
<td>45</td>
<td>59</td>
<td>14</td>
<td>Respiratory arrest</td>
</tr>
<tr>
<td>T</td>
<td>1988</td>
<td>25</td>
<td>37</td>
<td>12</td>
<td>Septicemia shock</td>
</tr>
<tr>
<td>U</td>
<td>1990</td>
<td>23</td>
<td>32</td>
<td>9</td>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>V</td>
<td>1992</td>
<td>21</td>
<td>23</td>
<td>2</td>
<td>Carbon monoxide poisoning (car exhaust)</td>
</tr>
<tr>
<td>W</td>
<td>1992</td>
<td>15</td>
<td>15</td>
<td>0.5</td>
<td>Degos syndrome</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anorexia nervosa (as primary or contributory cause of death)</td>
<td>8</td>
<td>34.8</td>
</tr>
<tr>
<td>Cancer</td>
<td>3</td>
<td>13.0</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2</td>
<td>8.7</td>
</tr>
<tr>
<td>Probable suicide</td>
<td>2</td>
<td>8.7</td>
</tr>
<tr>
<td>Opiate intoxication</td>
<td>1</td>
<td>4.3</td>
</tr>
<tr>
<td>Unknown/unclear</td>
<td>7</td>
<td>30.4</td>
</tr>
</tbody>
</table>

anorexia nervosa beginning in 1965. ICD-9 was introduced in 1980 with a category for anorexia nervosa (307.1), and all those patients were included for case note review. All notes with a coding for any eating disorder (F50) from ICD-10 were scrutinized in case of mistakes in coding. As in a previously published study (13), a diagnosis of anorexia nervosa was confirmed if the case notes contained the following: 1) a recorded diagnosis of anorexia nervosa, 2) significant weight loss, 3) amenorrhea or evidence of characteristic psychopathology, and 4) no other diagnosis of severe mental illness such as depression or schizophrenia during the index episode. The intention in choosing these criteria was to include all patients with anorexia nervosa and not to exclude those for whom there was insufficient recorded data. In some cases, particularly the earlier ones, the characteristic psychopathology was not described but could be inferred from the degree of weight loss, resistance to normal eating, and the presence of amenorrhea. It seemed important to include those patients even though the full modern diagnostic criteria were not documented. When there was genuine doubt about the diagnosis, patients were excluded. Of the patients with a case note diagnosis of anorexia nervosa, 32% were excluded because they did not meet the study criteria.

For the majority of patients, their index first contact was with general psychiatric services. In those services, no standardized assessments were performed from which severity of illness could be determined.

No patients were involved in any controlled therapeutic trial. Most would have had an eclectic variable mix of outpatient and inpatient treatment with a variety of individual, group, and family therapeutic approaches. It was therefore impossible to assess any associations between therapy and outcome, including mortality.
TABLE 3. Demographic and Clinical Characteristics of Deceased and Surviving Anorexia Nervosa Patients From All Cases (N=524) Seen in Specialist Services in Northeast Scotland, 1965–1999

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Deceased Cases (N=23)</th>
<th>Surviving Cases (N=501)</th>
<th>Mean Difference</th>
<th>95% CI</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>χ²</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>87.0</td>
<td>467</td>
<td>92.3</td>
<td>1.31</td>
</tr>
<tr>
<td>Age at presentation (years)</td>
<td>Mean: 25.0</td>
<td>SD: 9.8</td>
<td>Mean: 16.6</td>
<td>SD: 6.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Body mass index at presentationa</td>
<td>16.8</td>
<td>2.6</td>
<td>16.6</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Age at death (years)</td>
<td>35.8</td>
<td>12.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of illness (years from diagnosis to death)</td>
<td>10.9</td>
<td>10.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 4. Age-Standardized Mortality Ratios for Anorexia Nervosa Cases in Northeast Scotland, Correcting for Index Contact With Specialist Service

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Observed Deaths</th>
<th>Expected Deaths</th>
<th>Standardized Mortality Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3</td>
<td>0.81</td>
<td>3.7</td>
<td>1.2–11.4</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>6.21</td>
<td>3.2</td>
<td>2.1–5.0</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>7.03</td>
<td>3.3</td>
<td>2.2–4.9</td>
</tr>
</tbody>
</table>

The crude mortality rate was calculated simply by dividing the number of deaths by the total number of the cohort who were traced. To calculate the standardized mortality ratio, the time of interest for each patient was taken as the year the patient was seen at the clinic to either the year of death or mid-2002. For each individual this time was broken into episodes according to 5-year age bands, and into grouped time bands from tables of national statistics (before 1964–1979, 1980–1989, and from 1990 onward). Appropriate death rates for men and women in the Grampian population in 1974, 1985, and 1995 were obtained from the Demography and Dissemination Branch of the General Register Office for Scotland. These were used to calculate standardized mortality ratios and 95% confidence intervals. The methods were those of Clayton and Hill (14), implemented by using the Stata statistical package (15).

Discussion

This large follow-up study from a known population base, with diagnosis confirmed by case note review for each patient, shows that there is a more than threefold increased risk of premature death in patients diagnosed with anorexia nervosa. This finding is in keeping with most of the literature, although the standardized mortality ratio in published studies has ranged from 0.71 to 17.8 (6, 7, 11, 16, 17). It is very difficult to know what might have happened to the 17 untraced patients. It seems unlikely that they died in a developed country, since there are international systems in place to backtrack deaths to the U.K. Register of Births, Marriages and Deaths. It seems more likely that they moved abroad and were not picked up by the Community Health Index system.

In a smaller, earlier study in Northeast Scotland, a slightly higher standardized mortality ratio of 4.7 was calculated and compared with a lower standardized mortality ratio of 1.4 at St. George’s Hospital in London (9). It was suggested in this previous study that the presence of a well-developed specialist service for eating disorders at St. George’s Hospital may have been responsible for lower mortality there. Although there is now an outpatient specialist treatment service in the region based in Aberdeen, it is too early to see any effect on mortality rates that this
might have produced, bearing in mind the median time from presentation to death is 11 years.

Anorexia nervosa was only featured on about one-third of death certificates but may well have played a part in the death in other cases. In the two patients who died from complications of diabetes, the eating disorder may have been a contributory factor. Neilsen et al. (18) found a standardized mortality ratio for diabetes of 4.1 and for anorexia nervosa a standardized mortality ratio of 8.9. However, when these conditions were combined, the standardized mortality ratio was 14.5, suggesting it is a very dangerous combination. Patients are known to lower the dose of insulin as a means of purging calories through glycosuria (18).

Given the recorded cause of death, it is likely that two patients committed suicide and one other who died of opiate intoxication was a possible suicide, which suggests a crude suicide mortality rate of 0.57%. In a meta-analysis by Harris and Barraclough (19), the standardized mortality ratio for suicide in anorexia nervosa was 23, and some have suggested that while improved treatment and services may prevent death from anorexia nervosa and the consequences of starvation, there may be an increase in suicide (12). There were insufficient numbers in our study to explore this theory.

A surprising gap in the data was the lack of any understandable cause of death in six cases. We attempted to find more clinical information from general hospital case notes, but these were either unavailable or did not add to what was on the death certificate. It is possible that some of these deaths may have been related to anorexia nervosa, suicide, or other psychological disturbance, which seemed to be the case in the one which was attributed, rather oddly, to “psychological malfunction arising from mental factors (psychogenic dysmenorrhea).”

Most of the literature has suggested an increased mortality for anorexia nervosa, and this has been well reviewed by others (1), but there are some studies that have shown no increase in mortality (2, 7). In the study by Strober et al. (2), it is of interest that subjects in their cohort were younger at the time they first were seen in a specialist service (between 12 years and 17 years, 11 months) relative to our cohort, which had a mean age at index contact of 19.9 years for the surviving cases and 25.0 years for the cases who died. Naturally, it would be expected that older patients would be at greater risk of death, but it may also reflect delays in seeking treatment. Also, Strober et al.’s cohort had the benefit of early and intensive treatment, which continued after hospital discharge for most patients. The treatment available to our cohort was, prior to 1994, mostly in general psychiatric and psychological treatment services rather than in an intensive long-term program specializing in the treatment of eating disorders. For many patients in the study of Strober et al., it still took many years to reach full recovery from the illness, but their exposure to treatment may have been a significant factor in lowering mortality. Another study that showed an unexpectedly low mortality was undertaken as part of the Rochester epidemiology project (7). The possible reasons for the finding of a low standardized mortality ratio are well discussed by Sullivan (20). In particular, the definition and ascertainment methods used probably identified many milder cases seen in primary care. In contrast most other studies have identified patients seen in specialist services, sometimes highly specialized, tertiary care inpatient settings.

Because the numbers in any mortality study will inevitably be small, it is very difficult indeed to identify predictive factors. The only clear association with mortality in this study was older age at the time the patient was seen in the specialist service. Other possible predictive factors of mortality and poor outcome will be reported in a separate study comparing patients who died or otherwise had poor outcome with those who had a good outcome.

Northeast Scotland provides a unique setting within which it has been possible to study the epidemiology of a large cohort of patients with anorexia nervosa. First, there are case register and other patient administration records dating back to 1965. Second, there is very limited psychiatric practice outside of the National Health Service, and there is a readily accessible National Health Service psychiatric service. In community studies (21), patients with a definite diagnosis of anorexia nervosa who had not been referred to specialist services are seldom found. While some patients may initially be managed in primary care or referred to general hospital services because they are seen as less stigmatizing than mental health services, the great majority will subsequently be referred to psychiatric services in Northeast Scotland. Third, this study includes all patients seen by specialist services, not just those being admitted for inpatient treatment or only those referred to tertiary referral centers or academic centers, as is the case in many studies (2, 8, 10–12, 17). Fourth, the catchment population is well defined. Therefore, this cohort represents a reasonably complete estimate of all patients with anorexia nervosa and consequently a reasonable estimate of the risk of death from anorexia nervosa in a defined catchment population of 580,000.

The findings in this and many earlier studies confirm the medical seriousness of anorexia nervosa. It often runs a chronic course, and our study confirms that death due to anorexia nervosa can occur after many years of illness. Early access to specialist services should minimize delays in diagnosis, and even in those who have been ill for many years, continuing specialist intervention is necessary to try to reduce the risk of death in the long term.
This study was supported by a grant from the Chief Scientist's Office of the Health Department of the Scottish Executive (grant number K/OPR/2/2/D394).

The authors thank Graeme MacLennan for his help with the statistical analyses.

References

Neuropsychological Impairment and Its Neurological Correlates in Adult Offspring With Heightened Risk for Schizophrenia and Affective Psychosis

Erland W. Schubert, M.D.
Thomas F. McNeil, Ph.D.

Objective: Schizophrenia is generally considered to be a neurodevelopmental disorder reflected in findings of neuropsychological impairments and neurological abnormality in patients and their relatives. The authors investigated whether neuropsychological impairments are related to neurological abnormality and whether such deficits also characterize risk for affective psychosis.

Method: In a longitudinal study with a 93% rate of effective follow-up, the authors investigated neuropsychological impairment and its relation to neurological abnormality at a mean age of 22.3 years in 74 offspring of mothers with a history of psychotic disorders (38 offspring with heightened risk for schizophrenia and 36 with risk for affective psychosis) and 88 normal-risk offspring born to mothers with no history of psychosis.

Results: Offspring with genetically heightened risk for schizophrenia showed significantly impaired verbal memory, selective attention, and grammatical reasoning, compared with normal-risk offspring. Having impaired verbal memory, attention, and grammatical reasoning functions identified a significantly larger subgroup (16%) among offspring with heightened risk for schizophrenia than among offspring with heightened risk for affective psychosis (0%) and among normal-risk offspring (3%). Multiple neuropsychological functions were significantly related to neurological abnormality in offspring with heightened risk for schizophrenia and in normal-risk offspring but not among offspring with heightened risk for affective psychosis. The extension of schizophrenia and affective psychosis risk groups to include additional offspring of mothers with psychosis-spectrum disorders yielded results similar to those for the core risk groups.

Conclusions: The neurocognitive dysfunction attending heightened risk for schizophrenia is likely based on genetically mediated neurodevelopmental factors, with schizophrenia and affective psychosis belonging to different biological spheres.

S

schizophrenia is generally considered to be a disorder with neurodevelopmental roots reflected in findings of brain abnormalities, neurological abnormality, and neuropsychological impairments in both patients and their relatives (1–15). Patients with schizophrenia have a broad range of neuropsychological impairments in abstraction, executive functions, working and long-term memory, sustained and selective attention, and motor and perceptual-motor functions (e.g., references 6–8). Children and other relatives of patients with schizophrenia show an almost identical but less severe pattern of neuropsychological impairments (e.g., references 9–14). Especially, deviations in verbal memory and attention are associated with genetic risk for later development of schizophrenia-spectrum disorders (14–16). It is unclear whether these neuropsychological impairments are concentrated in a subgroup similar to the neurologically “high-scoring” subgroup (25%–50%) found among high-risk offspring of parents with schizophrenia (12, 17–19).

Some family, twin, and linkage studies suggest that schizophrenia and affective psychosis are at least to some degree genetically linked (20–22), and the relatives of patients with schizophrenia and affective psychosis have increased risk for both affective disorder and schizophrenia (23, 24). Neurological abnormalities and neuropsychological impairments are qualitatively similar in the two patient groups but quantitatively more severe in schizophrenia patients, and the offspring of schizophrenia patients have more neurological abnormalities than the offspring of mothers with affective psychosis (7, 8, 17, 25, 26). The neurodevelopmental hypothesis is not usually applied to the development of affective psychosis, and further study should be done to determine whether neuropsychological impairments are less strongly associated with risk for affective psychosis than with risk for schizophrenia, show a different pattern, characterize an especially poor-scoring subgroup, and are related to neurological abnormalities in a specific group of offspring at genetic risk for psychosis.
In this study we investigated 1) whether neuropsychological performance differs across groups of young adult offspring with heightened risk for schizophrenia or affective psychosis, relative to comparison offspring at normal risk, 2) whether verbal memory and attention are particularly impaired in offspring with heightened risk for schizophrenia, 3) whether impairment in verbal memory and attention is especially prominent in a subgroup of offspring with heightened risk for schizophrenia, 4) whether the pattern of neuropsychological impairments is the same for “core” versus extended “spectrum” groups of offspring with heightened risk for schizophrenia or affective psychosis, and 5) whether neuropsychological impairments correlate positively with neurological abnormalities in the different risk and comparison groups.

**Method**

**Subjects**

The subjects came from the adult follow-up of the prospective Swedish High-Risk Project; follow-up for this study occurred when the subjects were a mean age of 22.3 years (27). Heightened offspring risk for psychopathology was defined on the basis of a history of psychosis in the (index) mother, diagnosed by senior project diagnosticians (L. Kaij, M.D.; A. Malmquist-Larsson, M.D.) per all known psychiatric records for the woman and, where relevant, her biological relatives at entrance to the project. Normal offspring risk was defined on the basis of an absence of a psychiatric history in both the (comparison) mother and biological father, all defined by Research Diagnostic Criteria (28). In total, 166 (93.3%) of the 178 offspring were followed up during a 4-year period, in order to standardize age at the follow-up examination. Of these, 162 subjects (i.e., 74 high-risk offspring and 88 normal-risk offspring) had data on both neuropsychological function and neurological abnormalities. Four other subjects included in the general follow-up could not be assessed for neuropsychological and neurological abnormalities because of poor current mental health (one high-risk subject and one normal-risk subject), a problematic examination situation (one normal-risk subject), or recent death (one normal-risk subject).

The total high-risk group with neuropsychological and neurological abnormality data consisted of 28 offspring of mothers with schizophrenia (schizophrenia core group), 22 offspring of mothers with affective psychosis (affective core group, with 16 offspring of mothers with bipolar disorder and six offspring of mothers with unipolar depression), 15 offspring of mothers with schizoaffective psychosis and 36 offspring of mothers with affective-spectrum psychosis and 36 offspring of mothers with affective-spectrum psychosis. The extension of the core schizophrenia and affective risk groups to include additional subjects with maternal psychosis spectrum disorders had previously shown little effect on the rate and pattern of mental disturbance or neurological abnormality in the respective risk groups in our study (17, 28), indicating that the factors influencing mental health and neurological abnormality are equally present in the additional psychosis-

**Table 1. Characteristics of Normal-Risk and High-Risk Offspring in a Study of Neuropsychological Impairment and Neurological Abnormality in Adult Offspring of Mothers With a History of Psychotic Disorder**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Normal-Risk Offspring (N=88)a</th>
<th>Offspring of Mothers With Schizophrenia (N=38)b</th>
<th>Offspring of Mothers With Affective Psychoses (N=36)c</th>
<th>Offspring of Mothers With Schizophrenia (N=28)</th>
<th>Offspring of Mothers With Affective Psychoses (N=22)d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Age (years)</td>
<td>22.4</td>
<td>1.0</td>
<td>22.4</td>
<td>1.2</td>
<td>22.3</td>
</tr>
<tr>
<td>Gender, male</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Lower parental socioeconomic status</td>
<td>42</td>
<td>48</td>
<td>25</td>
<td>66</td>
<td>19</td>
</tr>
<tr>
<td>Education &gt;9 years</td>
<td>86</td>
<td>98</td>
<td>35</td>
<td>92</td>
<td>32</td>
</tr>
<tr>
<td>Left-handedness</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Good motivation a</td>
<td>88</td>
<td>100</td>
<td>37</td>
<td>97</td>
<td>36</td>
</tr>
<tr>
<td>Current axis I disorder</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td>16</td>
<td>9</td>
</tr>
</tbody>
</table>

a Offspring of mothers with no history of psychosis.
b Includes 28 offspring of mothers with schizophrenia and 10 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.
c Includes 16 offspring of mothers with bipolar psychotic disorder and six offspring of mothers with unipolar psychotic depression.
d Includes 22 offspring of mothers with affective psychosis and 14 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.

**In this study we investigated 1) whether neuropsychological performance differs across groups of young adult offspring with heightened risk for schizophrenia or affective psychosis, relative to comparison offspring at normal risk, 2) whether verbal memory and attention are particularly impaired in offspring with heightened risk for schizophrenia, 3) whether impairment in verbal memory and attention is especially prominent in a subgroup of offspring with heightened risk for schizophrenia, 4) whether the pattern of neuropsychological impairments is the same for “core” versus extended “spectrum” groups of offspring with heightened risk for schizophrenia or affective psychosis, and 5) whether neuropsychological impairments correlate positively with neurological abnormalities in the different risk and comparison groups.**
TABLE 2. Neuropsychological Test Scores of Normal-Risk Adult Offspring and High-Risk Adult Offspring of Mothers With a History of Psychotic Disorder

<table>
<thead>
<tr>
<th>Neuropsychological Test</th>
<th>Normal-Risk Offspring (N=88)a</th>
<th>Offspring of Mothers With Schizophrenia-Spectrum Psychoses (N=38)b</th>
<th>Offspring of Mothers With Affective-Spectrum Psychoses (N=36)c</th>
<th>Offspring of Mothers With Schizophrenia (N=28)</th>
<th>Offspring of Mothers With Affective Psychoses (N=22)d</th>
<th>Across-Groups Comparisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word-pair test</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>χ² (df=2) p</td>
</tr>
<tr>
<td>Immediate</td>
<td>26.15 ± 3.42</td>
<td>24.63 ± 3.57</td>
<td>25.94 ± 3.18</td>
<td>24.04 ± 3.79</td>
<td>25.95 ± 3.02</td>
<td>6.04 &lt; 0.05</td>
</tr>
<tr>
<td>1-hour delay</td>
<td>24.34 ± 4.38</td>
<td>21.89 ± 4.43</td>
<td>23.91 ± 4.06</td>
<td>20.89 ± 4.41</td>
<td>24.19 ± 3.98</td>
<td>9.28 0.01</td>
</tr>
<tr>
<td>Persuasive responses (%)</td>
<td>14.64 ± 8.73</td>
<td>14.30 ± 7.80</td>
<td>13.42 ± 7.81</td>
<td>13.64 ± 7.39</td>
<td>14.50 ± 8.59</td>
<td>0.05 n.s.</td>
</tr>
<tr>
<td>Grammatical reasoning test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of logical difficulty</td>
<td>5.52 ± 0.86</td>
<td>5.29 ± 0.84</td>
<td>5.12 ± 1.01</td>
<td>5.29 ± 0.94</td>
<td>5.00 ± 1.22</td>
<td>8.86 &lt; 0.02</td>
</tr>
<tr>
<td>Correct hits</td>
<td>34.18 ± 6.06</td>
<td>32.50 ± 5.86</td>
<td>31.38 ± 5.42</td>
<td>32.11 ± 5.96</td>
<td>31.29 ± 5.82</td>
<td>8.00 &lt; 0.02</td>
</tr>
<tr>
<td>Correct rejections</td>
<td>47.31 ± 3.95</td>
<td>43.68 ± 11.14</td>
<td>42.00 ± 11.76</td>
<td>43.79 ± 11.25</td>
<td>42.76 ± 13.33</td>
<td>9.91 0.007</td>
</tr>
<tr>
<td>Errors</td>
<td>5.40 ± 3.95</td>
<td>6.50 ± 5.25</td>
<td>6.88 ± 4.52</td>
<td>6.39 ± 4.52</td>
<td>6.14 ± 4.00</td>
<td>3.35 n.s.</td>
</tr>
<tr>
<td>Verbal fluency test</td>
<td>16.53 ± 3.66</td>
<td>17.61 ± 4.45</td>
<td>16.96 ± 3.64</td>
<td>17.46 ± 4.54</td>
<td>17.08 ± 3.47</td>
<td>1.66 n.s.</td>
</tr>
<tr>
<td>Digit span test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td>15.74 ± 6.30</td>
<td>17.66 ± 6.69</td>
<td>18.08 ± 6.90</td>
<td>17.82 ± 6.34</td>
<td>19.45 ± 6.92</td>
<td>4.50 n.s.</td>
</tr>
<tr>
<td>Backward</td>
<td>13.37 ± 5.97</td>
<td>14.03 ± 6.38</td>
<td>14.47 ± 7.08</td>
<td>14.21 ± 6.76</td>
<td>17.05 ± 7.36</td>
<td>0.58 n.s.</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>25.45 ± 9.22</td>
<td>28.29 ± 11.28</td>
<td>25.50 ± 11.40</td>
<td>26.50 ± 8.96</td>
<td>27.77 ± 13.56</td>
<td>2.76 n.s.</td>
</tr>
<tr>
<td>B</td>
<td>64.06 ± 23.97</td>
<td>72.21 ± 23.74</td>
<td>67.64 ± 26.31</td>
<td>70.00 ± 23.19</td>
<td>67.77 ± 21.84</td>
<td>4.04 n.s.</td>
</tr>
<tr>
<td>Selective attention test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compound hits</td>
<td>105.0 ± 10.1</td>
<td>99.6 ± 11.6</td>
<td>102.5 ± 10.4</td>
<td>100.0 ± 12.3</td>
<td>102.0 ± 11.5</td>
<td>6.06 &lt; 0.05</td>
</tr>
<tr>
<td>Correct hits</td>
<td>47.8 ± 5.6</td>
<td>45.11 ± 6.6</td>
<td>46.30 ± 5.3</td>
<td>45.3 ± 7.3</td>
<td>46.6 ± 5.0</td>
<td>5.29 0.07</td>
</tr>
<tr>
<td>Errors</td>
<td>3.56 ± 2.84</td>
<td>3.66 ± 3.89</td>
<td>2.91 ± 2.02</td>
<td>3.57 ± 3.06</td>
<td>2.33 ± 1.39</td>
<td>0.76 n.s.</td>
</tr>
<tr>
<td>Block design test</td>
<td>31.64 ± 6.19</td>
<td>30.29 ± 6.04</td>
<td>32.33 ± 5.57</td>
<td>30.43 ± 6.23</td>
<td>32.91 ± 6.02</td>
<td>2.54 n.s.</td>
</tr>
<tr>
<td>Finger-tapping test, alternate right/left</td>
<td>3.70 ± 0.60</td>
<td>3.59 ± 0.70</td>
<td>3.89 ± 0.54</td>
<td>3.67 ± 0.57</td>
<td>3.98 ± 0.51</td>
<td>3.77 n.s.</td>
</tr>
<tr>
<td>Reaction time</td>
<td>234.2 ± 27.0</td>
<td>248.2 ± 51.6</td>
<td>239.4 ± 28.2</td>
<td>238.5 ± 27.0</td>
<td>235.7 ± 27.9</td>
<td>1.76 n.s.</td>
</tr>
</tbody>
</table>

a Offspring of mothers with no history of psychosis.
b Includes 28 offspring of mothers with schizophrenia and 10 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.
c Includes 22 offspring of mothers with affective psychosis and 14 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.
d Includes 16 offspring of mothers with bipolar psychotic disorder and six offspring of mothers with unipolar psychotic depression.
e Comparison of normal-risk offspring, offspring of mothers with schizophrenia-spectrum psychoses, offspring of mothers with affective-spectrum psychoses (Kruskal-Wallis test, two-tailed, approximated by chi-square test).
f Wilcoxon-Mann-Whitney test, two-tailed.

spectrum subjects as in the “core” risk groups. The schizophrenia and affective spectra were selected as the primary diagnostic groupings for the present data analyses (with the core schizophrenia and affective diagnostic groups studied secondarily). The reason for this choice was that the spectrum approach increased group size and statistical power and tended to generally promote similarity in offspring background characteristics across the risk and comparison groups (Table 1).

After complete description of the study to the subjects, written informed consent was obtained.

**Neuropsychological Investigation**

The subjects were followed up during a full day of assessment at their local general practitioners' offices, which provided a standardized and neutral environment in each subject's own geographical area. The standardized procedure during the morning session included the neuropsychological testing and the neurological examination.

The selection of the neuropsychological tests and their particular scores for these analyses was based on 1) the goal of exploring a wide range of neuropsychological and cognitive functions (rather than representing a complete neuropsychological investigation), 2) previous evidence of impaired performance among schizophrenia patients and their relatives (6–14), and 3) avoidance of redundant scores (i.e., different scores showing a Spearman correlation (r_s)>0.50 in the total current study group) for the same neuropsychological test (such redundancy concerned different scores for the Wisconsin Card Sorting Test, for the selective attention test, and for the grammatical reasoning test).

The following tests/scores were selected for use in the current analyses: 1) the word-pair test (number of word pairs that the subject could recall immediately and after a 1-hour delay) (29); 2) Trail Making Test A and B (number of seconds to completion) (30); 3) digit span test (number of correctly reproduced digits) both forward and backward (31); 4) verbal fluency test (number of words reported) (32); 5) Wisconsin Card Sorting Test (percent errors and percent perseverative responses) (33); 6) block design test (points for correct design and speed to completion) (31); 7) finger-tapping test, alternating right/left hand, implemented on a microcomputer (number of taps per second) (34); 8) reaction time test (simple visual reaction time, mean reaction time in milliseconds) (34); 9) selective attention test (correct hit or rejection of the letter K among other letters (34), scored as compound hits, correct hits, and errors); and 10) grammatical reasoning test (34)
Neurological Examination

The neurological examination was based on a comprehensive, standardized assessment scale (5, 17), previously used by us to study adult schizophrenia patients and their siblings (5). The neurological abnormality scale comprises investigation of motor coordination, muscle power, muscle tone, sensory functions, reflexes, and cognitive functions. The total neurological abnormality score was the sum of the scores on 44 items (potentially ranging from 0 to 124).

The interrater reliability for this neurological assessment was determined by testing agreement on the total scores with an experienced physician (S. Ismail, M.D.) for 20 subjects (10 patients with psychosis and 10 hospital personnel). The interrater coefficient (intraclass correlation) was 0.97 (F = 65.44, df = 9, 10, p < 0.001).

Statistical Methods

Because of the score distributions, nonparametric tests were used for primary comparison of the groups and investigation of correlation between neuropsychological impairments and neurological abnormalities. The Kruskal-Wallis test was used to compare schizophrenia-spectrum psychosis, affective-spectrum psychosis, and normal-risk offspring groups on quantitative scores for the neuropsychological tests (Table 2). When Kruskal-Wallis analysis showed a significant or nearly significant across-group difference, secondary comparisons of pairs among these three groups were done by using Mann-Whitney tests.

Based on our previous research (17, 27), a score beyond the 90th percentile in the “poor-scoring” direction on the neuropsychological tests for the normal-risk subjects was operationally defined as “poor scoring” for all subjects and was used for classification of performance with respect to confounders and subgroups.

Five neuropsychological test scores (Table 3) were selected for investigating the effect of possible confounders on the significant across-group differences. The five neuropsychological tests that were chosen showed not only a significant across-group difference on the Kruskal-Wallis test but also the possibility of estimating an appropriate 90th percentile cutoff level. (The latter criterion was not fulfilled for the grammatical reasoning test’s level of logical difficulty score because of the score distribution.)

Three subject characteristics (gender, parental socioeconomic status, and current axis I disorder) were identified as potential confounders on the basis of the across-group differences in offspring characteristics shown in Table 1. The method suggested by Greenland (35) was used to define confounders for multivariate logistic regression analysis. Starting with the univariate model, variables were entered into bivariate and multivariate models if they changed the effect estimate by 10% or more (which was found for gender) and were excluded from the model if the effect estimates changed less than 5% (which was found for parental socioeconomic status and current axis I disorder). Logistic regression analysis adjusted for gender, with odds ratios and 95% confidence intervals (CIs), was thereafter used for analysis of the rate with which high-risk subjects scored above the 90th percentile cutoff level for the five selected neuropsychological test scores.

Based on previous findings and the current results, the same tests of verbal memory, attention, and grammatical reasoning functions (Table 3) were used further to investigate the relative size of a “poor-scoring” subgroup among high-risk (compared with normal-risk) subjects. The relative frequency of subjects showing “poor-scoring” on one, two, or all three of these functions was compared across the study groups by using Fisher’s exact test, with odds ratios and 95% CIs.

Spearman’s rank-order correlations were used to examine the relation between neuropsychological and neurological abnormalities within the schizophrenia-spectrum psychosis, affective-spectrum psychosis, and normal-risk offspring groups, separately.

Statistical significance was defined as p < 0.05, two-tailed, with 0.10 > p > 0.05 denoting results that approached significance.

The neuropsychological performance for the core schizophrenia-risk and affective-risk groups was evaluated visually rather than submitted to formal statistical analysis.

Results

Neuropsychological Performance

Significant differences or results approaching significance were found across schizophrenia-spectrum psycho-

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>z</td>
<td>p</td>
<td>z</td>
</tr>
<tr>
<td></td>
<td>2.44</td>
<td>&lt;0.02</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>0.04</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>1.82</td>
<td>0.07</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>2.24</td>
<td>&lt;0.03</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td>&lt;0.02</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>&lt;0.03</td>
<td>0.98</td>
</tr>
</tbody>
</table>
A neuropsychological poor-scoring subgroup was operationally defined by the proportion of individuals within each study group who scored above the 90th percentile on the verbal memory (word-pair test), attention (selective attention test), and grammatical reasoning scores shown in Table 3. Summarizing across these neuropsychological functions, significantly more schizophrenia-spectrum offspring (than normal-risk offspring) scored above this cutoff on one or more of these neuropsychological functions (Step A), on two or more of the functions (Step B), and on all three functions (Step C) (Table 4).

In contrast, significantly more affective-spectrum offspring (than normal-risk offspring) scored above the 90th percentile cutoff only on one or more of these functions (Step A) but not on two or more or on all three functions (Steps B and C).

Directly comparing the two risk groups, schizophrenia-spectrum offspring more frequently scored above the 90th percentile than affective-spectrum offspring on all three functions (Step C) \((p<0.03, \text{Fisher's exact test; odds ratio}=14.6, 95\% \text{CI}=0.79–269.5)\) but not on two or more functions (Step B) \((p=0.25, \text{Fisher's exact test; odds ratio}=2.21, 95\% \text{CI}=0.67–7.27)\) or one or more functions (Step A) \((p=1.00, \text{Fisher's exact test; odds ratio}=1.00, 95\% \text{CI}=0.40–2.49)\).

Furthermore, the schizophrenia-spectrum offspring showed increasing differences (odds ratios) from normal-risk offspring when progressing from Step A to Step C. In contrast, affective-spectrum offspring did not differ from normal-risk offspring at Step B and Step C, and the odds
showed significantly impaired verbal memory, selective
spring with genetically heightened risk for schizophrenia
Discussion

Logical abnormality score at a mean age of 22.3 years was
neurological abnormality score was significantly corre-

e Fisher
1.71
–
–
–
22.53 0.022 0.35 0.02
2.20 0.63–7.75 0.29
14.66 0.0064 2.20 0.63–7.75 0.29
5.30 1.25–22.52 0.022 0.35
0.02–6.88 0.56

A poor score was operationally defined as a score beyond the 90th percentile (in the “poor-scoring” direction) in the normal-risk group. The selected neuropsychological tests were the word-pair, grammatical reasoning, and selective attention tests (see Table 3).

Includes 22 offspring of mothers with affective psychosis (16 with bipolar psychotic disorder and six with unipolar psychotic depression) and 14 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.

Includes 28 offspring of mothers with schizophrenia and 10 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.

d Includes 14 offspring of mothers with schizoaffective psychosis or unspecified functional psychosis.

e Fisher’s exact test, two-tailed.

ratios for those comparisons decreased moving from Step A to Step C.

Neuropsychological Function and Neurological Abnormality

In the schizophrenia-spectrum group, the total neurological abnormality score at a mean age of 22.3 years was significantly correlated with scores for word-pair test—immediate recall and 1-hour delayed recall, grammatical reasoning test—correct rejections, verbal fluency, digit span—forward and backward, block design test, and finger-tapping test (Table 5). In the normal-risk group, the total neurological abnormality score was significantly correlated with scores for the word-pair test—immediate recall and 1-hour delayed recall; Wisconsin Card Sorting Test—percent errors and percent perseverative responses; grammatical reasoning test—level of logical difficulty, correct hits, and correct rejections; verbal fluency; digit span—forward and backward; selective attention test—compound hits; block design test; finger-tapping test; and reaction time. All significant correlations represented a positive relationship between neuropsychological impairments and neurological abnormalities.

In contrast, no significant correlation was found between the total neurological abnormality score and any neuropsychological test results in the affective-spectrum group. However, correlations approaching significance were found between the total neurological abnormality score and scores for 1) the word-pair test—immediate recall and 1-hour delayed recall and 2) Trail Making Test A.

Discussion

In this prospective, longitudinal study, young adult offspring with genetically heightened risk for schizophrenia showed significantly impaired verbal memory, selective attention, and grammatical reasoning, compared with normal-risk offspring. Having impairment on all three functions identified a significantly larger subgroup among offspring of mothers with schizophrenia-spectrum psychosis (16%) than among offspring of mothers with affective-spectrum psychosis (0%) and normal-risk offspring of mothers with no history of psychosis (3%). Several neuropsychological test scores were significantly related to neurological abnormalities at the same age in schizophrenia-spectrum offspring and in normal-risk offspring, but not among affective-spectrum offspring. In total, these findings suggest that the neurodevelopmental correlates of risk for developing schizophrenia differ from those for developing affective psychosis.

We confirmed previous findings of verbal memory impairments in patients with schizophrenia, their children, and other relatives (6–14, 16), and also showed that verbal memory tends to discriminate offspring of mothers with schizophrenia-spectrum psychosis from offspring of mothers with affective-spectrum psychosis. In addition, verbal memory was significantly correlated with neurological abnormalities among schizophrenia-spectrum offspring. These findings are in line with the previous suggestion that verbal memory dysfunction may be a specific trait marker for liability for schizophrenia (14–16).

The grammatical reasoning test was the only neuropsychological test on which both high-risk groups scored significantly more poorly than normal-risk offspring. The grammatical reasoning test is an especially demanding and complex cognitive test that assesses sensorimotor function, visual scanning and perception, and semantic and logical processing. Both high-risk groups had fewer correct hits and achieved a lower level of difficulty, but did not have more errors, than normal-risk offspring. These findings suggest that high-risk offspring process the infor-
mation in a logically correct manner but at a slower rate than normal-risk offspring. As this test involves integration of multiple neuropsychological functions as well as brain regions, the slower processing may indicate difficulty in integration of these functions or regions (2). Similarly, the schizophrenia-spectrum offspring did not make more errors but achieved fewer correct hits on the selective attention test, compared with the normal-risk offspring. This result may again indicate slower information processing rather than a specific attentional dysfunction.

The identification of a substantial subgroup (16%–50%, depending on the criterion used) (Table 4) of individuals with impairments in verbal memory, attention, and/or grammatical reasoning among the schizophrenia-spectrum offspring corresponds to the occurrence of a neurologically “high-scoring” subgroup (25%–50%) of schizophrenia offspring observed in our own and other studies of high-risk offspring in infancy, childhood, adolescence, and adulthood (12, 17–19). These repeated findings support the hypothesis of a genetically mediated disturbance in the neurodevelopmental maturation in a subgroup of individuals at presumed risk for schizophrenia.

Multiple neuropsychological test scores were significantly related to neurological abnormalities at the same age among schizophrenia-spectrum offspring and among normal-risk offspring, while no significant relation was seen among affective-spectrum offspring. The absence of such a relation among these latter offspring cannot be explained by lack of neurological abnormalities, as normal-risk offspring and affective-spectrum offspring had a similar mean level of neurological abnormalities (17). In total, these findings suggest that the neurocognitive dysfunction attending increased risk for schizophrenia is based on neurological/biological/cerebral factors (rather than, for example, deviations in motivation or personality), that schizophrenia is likely a genetically mediated neurodevelopmental disorder, and that schizophrenia and affective psychosis belong to different biological spheres.

The extension of core schizophrenia and affective psychosis risk groups to include offspring of mothers with additional psychosis-spectrum disorders had little effect on the rate and pattern of neuropsychological function in the respective risk groups. This result is in line with previous findings (17, 27) and seems to indicate that the factors influencing neuropsychological function are equally present.
in the psychosis-spectrum offspring and in the core group offspring.

The strengths of the study were the prospective design, the high rate of follow-up of adult offspring, the existence of different high-risk groups, the narrow age range at examination, the control of possible confounders, the use of an extensive standardized examination routine conducted by one investigator uninformed with regard to the subject's study group, and access to other project data.

The limitations of this study include the small numbers of subjects in specific high-risk groups (yielding low statistical power and limiting the possibility for evaluating negative findings), the lack of screening of comparison subjects' parents for any indicators of cognitive performance, and the lack of independent assessment of the neurological and neuropsychological judgments by different examiners.

In summary, 1) impairments in verbal memory, selective attention, and complex executive function; 2) a substantial poor-scoring subgroup; and 3) significant correlations between neuropsychological functions and neurological abnormalities were found among offspring with increased genetic risk for developing schizophrenia but not among offspring at risk for affective psychosis. This finding supports the position that schizophrenia is a genetically mediated neurodevelopmental disorder whose etiology at least partly differs from that of affective psychosis.

References


27. Schubert EW, McNeil TF: Prospective study of adult mental disturbance in offspring of women with psychosis. Arch Gen Psychiatry 2003; 60:473–480
30. Reitan RM: Trail Making Test: Manual for Administration, Scoring, and Interpretation. Bloomington, Indiana University Medical Center, Department of Neurology, Section of Neuropsychology, 1956
34. Levander S: An Automated Psychological Test Battery, IBM-PC version (APT-PC): Research Reports From the Department of Psychiatry and Behavioural Medicine, University of Trondheim, Vol 11, No 65. Trondheim, Norway, University of Trondheim, 1988
Maternal Exposure to Toxoplasmosis and Risk of Schizophrenia in Adult Offspring

Alan S. Brown, M.D.
Catherine A. Schaefer, Ph.D.
Charles P. Quesenberry, Jr., Ph.D.
Liyan Liu, M.D., M.Sc.
Vicki P. Babulas, M.P.H.
Ezra S. Susser, M.D., Dr.P.H.

Objective: The authors examined the relationship between maternal antibody to toxoplasmosis and the risk of schizophrenia and other schizophrenia spectrum disorders in offspring. Toxoplasmosis is known to adversely affect fetal brain development.

Method: In a nested case-control design of a large birth cohort born between 1959 and 1967, the authors conducted serological assays for Toxoplasma antibody on maternal serum specimens from pregnancies giving rise to 63 cases of schizophrenia and other schizophrenia spectrum disorders and 123 matched comparison subjects. Toxoplasma immunoglobulin (IgG) antibody was quantified by using the Sabin-Feldman dye test. The IgG titers were classified into three groups: negative (<1:16), moderate (1:16–1:64), and high (≥1:128).

Results: The adjusted odds ratio of schizophrenia/schizophrenia spectrum disorders for subjects with high maternal Toxoplasma IgG antibody titers was 2.61 (95% confidence interval = 1.00–6.82). There was no association between moderate Toxoplasma IgG antibody titers and the risk of schizophrenia/spectrum disorders.

Conclusions: These findings suggest that maternal exposure to toxoplasmosis may be a risk factor for schizophrenia. The findings may be explained by reactivated infection or an effect of the antibody on the developing fetus. Given that toxoplasmosis is a preventable infection, the findings, if replicated, may have implications for reducing the incidence of schizophrenia.

Several maternal infections have been associated with an elevated risk of schizophrenia in offspring (1–3). Evidence suggests that infections known to cause congenital CNS anomalies in humans, including rubella (4), herpes simplex (5), polio, and varicella-zoster virus (1), might be related to the risk of schizophrenia. Toxoplasmosis is also associated with maldevelopment of the CNS, and ecological data have led to the suggestion that this organism might be involved in the etiology of schizophrenia (6).

Therefore, we used serological methods to investigate whether maternal exposure to toxoplasmosis is associated with an increased risk of schizophrenia in adult offspring.

Toxoplasma gondii, the cause of toxoplasmosis, is a ubiquitous intracellular parasite (7, 8). When primary infection occurs during pregnancy, the offspring have a markedly increased risk of CNS congenital abnormalities, including microcephaly, hydrocephalus, mental retardation, convulsions, cerebral calcifications, and chorioretinitis (7–9). Delayed neurologic sequelae, including lower IQ, retarded psychomotor development, and sensorineural deafness, have also been demonstrated in subjects who were exposed in utero, even among those with subclinical infection during the neonatal period (7, 8).

Serological methods for detection of Toxoplasma include direct detection of the parasite, immunoassays for serum immunoglobulin (IgM antibody, and elevation of maternal IgG antibody to Toxoplasma. Elevation of Toxoplasma IgG may reflect primary active or reactivated infection; however, unlike IgM antibody, which is a specific indicator of recent infection, increased IgG may persist for years in subjects with dormant infection (7). Increased IgG titers to Toxoplasma have been associated with both severe and subtle neuropsychiatric abnormalities (10).

In the present study, we conducted assays of archived prenatal serum specimens drawn prospectively in subjects with schizophrenia and matched comparison subjects from the Prenatal Determinants of Schizophrenia Study, a large birth cohort investigation (11), to examine the relationship between elevated Toxoplasma antibody and risk of adult schizophrenia.

Method

Description of the Cohort

The Prenatal Determinants of Schizophrenia Study has been described previously (11) and will be only briefly reviewed here. The cohort members were enrolled in the Child Health and Development Study (12), which took place from 1959 to 1967. The Child Health and Development Study recruited nearly every pregnant woman under obstetric care from the Kaiser Foundation Health Plan in Alameda County, Calif. The 19,044 live-born offspring of these women were automatically enrolled in the Kaiser Foundation Health Plan. The Child Health and Development Study collected data from maternal medical records, maternal interviews, and other sources.

The Prenatal Determinants of Schizophrenia Study cohort consisted of the 12,094 live births who belonged to the Kaiser Perma-
TABLE 1. Distribution of Immunoglobulin G Toxoplasmosis Antibody Titers With the Sabin-Feldman Dye Test in Subjects With Schizophrenia Spectrum Disorders and Comparison Subjects

<table>
<thead>
<tr>
<th>Antibody Titera</th>
<th>Subjects With Schizophrenia Spectrum Disorders</th>
<th>Comparison Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Reference (&lt;1:16)</td>
<td>45</td>
<td>71.4</td>
</tr>
<tr>
<td>Moderate (1:16–1:64)</td>
<td>5</td>
<td>7.9</td>
</tr>
<tr>
<td>High (1:128–1:1024)</td>
<td>13</td>
<td>20.6</td>
</tr>
</tbody>
</table>

aAntibody titers are presented only for subjects who were seropositive; titers <1:16 are considered seronegative by definition.

Maternal serum samples were obtained during pregnancy for virtually all subjects, were frozen immediately, and were archived at −20°C in a single repository. All specimens were uniformly handled and stored in accordance with a strict protocol.

Collection of Maternal Sera

The outcome was schizophrenia and other schizophrenia spectrum disorders, defined from previous studies (13) as any of the following: schizophrenia, schizoaffective disorder, delusional disorder, psychotic disorder not otherwise specified, and schizotypal personality disorder. Case ascertainment involved three steps: 1) ascertainment of potential cases from computerized records, 2) chart review of potential cases to confirm eligibility for assessment, and 3) diagnostic interview (or chart review) and consensus diagnosis. Case ascertainment was conducted by a computerized record linkage between the Child Health and Development Study and Kaiser Permanente Medical Care Plan identifiers by using inpatient, outpatient, and pharmacy registries. Subjects from the hospital registry were screened for potential schizophrenia spectrum disorders based on diagnoses of ICD-9 295–299 and psychiatrist review of all psychiatric and medical records. Patients from the outpatient registry screened positive if they were assigned ICD-9 diagnoses of 295, 297, 298, or 299. Subjects from the pharmacy registry screened positive based on a history of antipsychotic treatment.

There were 13 deceased subjects among those who screened positive for potential schizophrenia spectrum disorders (N=183). Among the 170 remaining potential subjects with schizophrenia spectrum disorders, 146 (86%) were contacted to schedule a diagnostic interview.

Clinicians with at least a master’s degree in a mental health field who were trained to reliability administered the Diagnostic Interview for Genetic Studies to potential subjects with schizophrenia spectrum disorders (14). Consensus of three experienced research psychiatrists was used to obtain DSM-IV diagnoses based on a review of the Diagnostic Interview for Genetic Studies narrative and medical records and discussions with the interviewer. The Diagnostic Interview for Genetic Studies was completed by 107 (73%) of the 146 contacted potential subjects with schizophrenia or schizophrenia spectrum disorders. For the 76 potential subjects who were not interviewed, chart reviews by experienced clinicians were conducted; all diagnoses were confirmed by a research psychiatrist. These procedures yielded a total of 71 subjects with schizophrenia spectrum disorders, 44 of whom received the Diagnostic Interview for Genetic Studies and 27 of whom were diagnosed by chart review. Among these 71 subjects with schizophrenia spectrum disorders, 64 had available prenatal sera. The diagnoses of these subjects were schizophrenia (N=38), schizoaffective disorder (N=15), delusional disorder (N=1), schizotypal personality disorder (N=5), and other schizophrenia spectrum psychosis (N=5).

All subjects in the Prenatal Determinants of Schizophrenia Study provided written informed consent for human investigation. The study protocol was approved by the institutional review boards of the New York State Psychiatric Institute and the Kaiser Foundation Research Institute.

Laboratory Assay

All of the assays were performed with researchers blind to case/comparison status in the Toxoplasma Serology Laboratory at the Palo Alto Medical Foundation Research Institute, the Toxoplasma reference laboratory for the United States (15), under the direction of Dr. Jack Remington.

Three assays were used in the present study (15). The first two concern the assessment of Toxoplasma IgG antibody titer. In accordance with established practice, samples were first screened for the IgG antibody titer by using the screen agglutination test. Then, to definitively establish the presence of Toxoplasma IgG antibody, the Sabin-Feldman dye test (16), the reference standard for the serological detection of Toxoplasma antibody (7), was performed in the samples that screened positive on the agglutination test.

To examine whether recent infection with toxoplasmosis occurred in our samples, we assayed for Toxoplasma IgM antibody using the double-sandwich enzyme-linked immunosorbent assay (IgM ELISA) (17). Given the extremely low likelihood that positive IgM antibody would be found in a sample with a negative antibody result on the IgG screening agglutination test, the IgM ELISA was performed only on subjects who screened positive in the agglutination test.

Categorization of Exposures

Following established practice, the screen agglutination test results for Toxoplasma IgG were categorized as a dichotomy (positive/negative).

The Sabin-Feldman dye test (16) yields IgG antibody titer results in serial twofold dilutions. All dye test IgG titers <1:16 are considered negative. Given the lack of clear precedents in the literature for the classification of Toxoplasma IgG antibody titers, we categorized the subjects with positive IgG titers based on the distribution of the titers in our sample. “High” titer was defined as a Toxoplasma IgG antibody titer of ≥1:128; this category represented approximately the highest 10th percentile (10.5%) of IgG titers for comparison subjects in our study group. The moderate titer group was defined as an IgG antibody titer of 1:16–1:64 and consisted of the remaining subjects with positive IgG antibody titers. This classification strategy provided sufficient subject numbers in each of the exposure groups to permit a meaningful analysis of the data, while also allowing us to examine the effect of different magnitudes of antibody titer on the risk of schizophrenia spectrum disorders. The reference category consisted of subjects with negative IgG antibody titers.

In accordance with the methods used by the Toxoplasma Serology Laboratory at the Palo Alto Medical Foundation Research Institute (17), IgM ELISA antibody titer values >1.8 were considered positive, values from 1.7 to 1.9 were equivocal, and values ≤1.6 were negative.

http://ajp.psychiatryonline.org

Am J Psychiatry 162:4, April 2005
Analytic Strategy

The analysis was based on a nested case-control design (18) in which the comparison subjects for each case are selected to represent the population at risk when the case was ascertained. In the Prenatal Determinants of Schizophrenia Study, cases were ascertained on the first date of medical attention for schizophrenia spectrum disorders.

Eligible Subjects With Schizophrenia Spectrum Disorders

Among the 71 subjects with schizophrenia spectrum disorders, 64 had at least one available prenatal serum sample: 58 (90.6%) of these 64 subjects had either schizophrenia or schizoaffective disorder. The last serum sample available for each pregnancy (late third trimester or perinatal) was used. This provided the greatest opportunity to detect Toxoplasma infection if it occurred at all during pregnancy. Even if exposure occurred in early pregnancy, it was unlikely to have been missed by our assay method because Toxoplasma IgG antibodies generally remain elevated for many months or years after infection (7).

Eligible Comparison Subjects

Eligible comparison subjects (N=10,768) were selected from offspring without schizophrenia spectrum disorders in the Prenatal Determinants of Schizophrenia Study cohort after excluding siblings of subjects with schizophrenia spectrum disorders, subjects with major affective disorders, and subjects without prenatal sera.

Matching Procedure

Matching by a nested case-control design ensured that each case and its corresponding comparison subject were followed for equal lengths of time from birth until first treatment of the case. Comparison subjects were matched to subjects with schizophrenia spectrum disorders on membership in the Kaiser Permanente Medical Care Plan at the time the case was ascertained, date of birth (±28 days), gender, and number and timing (±28 days) of the first maternal blood sample taken during the index pregnancy (2, 11).

To conserve the sera, two comparison subjects were selected at random from the pool of potential matched comparison subjects for each case and were further matched on gestational age by requiring that they were drawn within 42 days of the serum sample of the case. This selection process resulted in 124 matched comparison samples (60 sets with 1:2 matching and four sets with 1:1 matching). A comparison in a 1:1 matched set was eliminated from the analysis because of an insufficient quantity of serum, resulting in 63 subjects with schizophrenia spectrum disorders and 123 comparison subjects (60 sets matched at 1:2 and three sets matched at 1:1). Gestational ages (in days) of sera from subjects with schizophrenia spectrum disorders (mean=271.7, SD=24.8) and comparison subjects (mean=275.3, SD=26.3) did not differ significantly from one another (t=0.67, df=184, p=0.51).

Appropriate to the nested case-control study design, point and interval estimates of odds ratios were obtained by fitting conditional logistic regression models for matched sets (19). We first tested the relationship between high maternal Toxoplasma IgG and risk of schizophrenia spectrum disorders. We then examined whether moderate Toxoplasma IgG was associated with schizophrenia spectrum disorders. Statistical significance was judged at α=0.05.

After first obtaining unadjusted estimates of the association between Toxoplasma IgG and schizophrenia spectrum disorders, we then assessed the following covariates as potential confounders: maternal age (<35 [reference], ≥35), maternal ethnicity (Caucasian [reference], African American, other), maternal socioeconomic status, defined as maternal education (<high school, high school only [reference], some college/college graduate), and gestational age of the serum sample (in days after last menstrual period).

Results

Demographic Characteristics

The mean ages of the subjects with schizophrenia spectrum disorders and the comparison subjects, respectively, were 24.2 years (SD=4.8) and 25.2 years (SD=4.9) (t=0.09, df=184, p=0.93). The proportion of male subjects was 50% (21 of 42) of the subjects with schizophrenia spectrum disorders and 66.7% (82 of 123) of the comparison subjects. Maternal age did not differ significantly between the subjects with schizophrenia spectrum disorders (mean=30.0, SD=6.2) and the comparison subjects (mean=28.6, SD=6.1) (t=−1.06, df=184, p=0.29).

Screen Agglutination Test

Of the 186 samples tested, 55 were positive on the screen agglutination test. For the subjects with schizophrenia spectrum disorders, 25 of 63 (39.7%) were positive; for the comparison subjects, 30 of 123 (24.4%) were positive.

IgG With Dye Test

The seroprevalence of IgG antibody by the Sabin-Feldman dye test among the 55 subjects who screened positive on the screen agglutination test was 18 of 63 (28.5%) of the subjects with schizophrenia spectrum disorders and 22 of 123 (17.9%) of the 123 comparison subjects. The prevalence of high Toxoplasma IgG antibody titers was greater for subjects with schizophrenia spectrum disorders than for comparison subjects (Table 1).

In the conditional logistic regression analysis, the odds ratio of schizophrenia spectrum disorders for the subjects with a high Toxoplasma IgG antibody titer was 2.42 (95% confidence interval [CI]=0.94–6.25, p<0.07) (Table 2). There was no increase in the risk of schizophrenia spectrum disorders for the group with moderate IgG antibody titers (odds ratio=1.37, 95% CI=0.40–4.73, p=0.62). The global test of association (df=2) yielded a p value of 0.19.

Testing of Covariates in Relation to Toxoplasma IgG

Of the four covariates examined in relation to the Toxoplasma IgG antibody among comparison subjects, an appreciable association was found only for maternal age: older mothers had higher IgG antibody titers than younger mothers (χ²=4.07, df=1, p=0.05). No associations were observed between Toxoplasma IgG and maternal ethnicity, education, and gestational age of the serum samples.

Given the association between toxoplasmosis and maternal age, we adjusted for this covariate (Table 2). The adjusted odds ratio of schizophrenia spectrum disorders for subjects with high IgG antibody titers was 2.61 (95% CI=1.00–6.82, p=0.051). There was no association between moderate IgG antibody titer and risk of schizophrenia spectrum disorders.
TABLE 2. Conditional Logistic Regression of Antibody to Toxoplasmosis and Risk of Schizophrenia and Other Schizophrenia Spectrum Disorders in Subjects With Schizophrenia Spectrum Disorders and Comparison Subjects

<table>
<thead>
<tr>
<th>Immunoglobulin G Titer</th>
<th>Unadjusted</th>
<th>Adjusted for Maternal Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Reference (&lt;1:16)</td>
<td>—</td>
<td>1.00</td>
</tr>
<tr>
<td>Moderate (1:16–1:64)</td>
<td>0.25</td>
<td>1.37</td>
</tr>
<tr>
<td>High (1:128–1:1024)</td>
<td>0.88</td>
<td>2.42</td>
</tr>
</tbody>
</table>

Prenatal exposure to toxoplasmosis is a plausible risk factor for schizophrenia. The parasite has a predilection for the developing fetal brain (7, 8) and results in a similar array of congenital abnormalities as rubella and other pathogens that have been implicated in schizophrenia (1, 4, 5). Postmortem and neuroimaging studies in newborns with congenital toxoplasmosis have revealed several CNS abnormalities, including enlargement of the third and lateral ventricles secondary to hydrocephalus, and intracranial calcifications, frequently of the basal ganglia, choroid plexus, and meninges. Although these findings differ in severity, and to some degree in type, from those in schizophrenia, radiological studies have not been conducted on children who were exposed in utero to toxoplasmosis but who did not have the stigmata of congenital toxoplasmosis at the time of birth. It has been shown, however, that 40%–70% of these asymptomatic newborns later developed neurocognitive and neuromotor abnormalities that resemble those found in children who were later diagnosed with schizophrenia (8). These sequelae include retarded mental and psychomotor development, mildly decreased IQ, coordination problems, neurologic soft signs, and stereotypes (7, 10).

Because antibody titers to Toxoplasma IgG may remain elevated for significant periods of time, an increase in IgG antibody may reflect an active primary infection, reactivation of infection, or a persistent immune response to a dormant infection. First, we discuss active primary infection. In the present study, none of the serum samples in our cohort tested positive for IgM-specific Toxoplasma antibody, the most robust indicator of recently acquired infection (7). In a previous birth cohort study conducted during the same years as the Child Health and Development Study, the prevalence of primary Toxoplasma infection during pregnancy was low (10), consistent with subsequent studies (7). These results indicate that primary infection is unlikely to account for the observed finding.

Second, we consider reactivation of a previous infection. Following acute Toxoplasma infection, the parasites are not completely eliminated by the immune system; rather, they become sequestered as bradyzoites in dormant cysts in affected organs, most commonly the brain, eye, heart, and skeletal muscle (7, 8). However, the cysts have been known to spontaneously rupture as a result of host immunosuppression or other factors. When this occurs, reactivation of infection is produced by conversion of bradyzoites into tachyzoites, which are released from the cyst, proliferate, and invade cells (7, 22). The resulting anamnestic re-
sponse produces elevations of IgG antibody titers. An effect of activated dormant *Toxoplasma* microcysts on brain function is suggested by studies demonstrating associations between increased *Toxoplasma* IgG antibody and first-episode schizophrenia (23), personality changes (24), and cryptogenic epilepsy (25) in adult patients. The prevalence of reactivated infection is not known.

*Toxoplasma* is transmitted to the fetus through the placenta (26). Infection is most often transmitted to the offspring when *Toxoplasma* is acquired later in pregnancy, with the highest risk in the third trimester (26). The proliferating tachyzoites in the fetal CNS and other organs destroy parasitized cells and result in an inflammatory response, leading to anoxia, cell death, and tissue necrosis (8). *Toxoplasma* also increases levels of homovanillic acid and dopamine, which are implicated in the pathogenesis of schizophrenia (27).

It has also been suggested that *Toxoplasma gondii* may result in congenital CNS abnormalities without direct transmission of the parasite to the fetus. This mechanism is proposed to involve toxofactor, a toxin released by *Toxoplasma* (28). When administered during pregnancy, toxofactor causes congenital abnormalities, particularly CNS defects, in exposed animals.

It is also possible that the risk of schizophrenia may be increased in the offspring of mothers with dormant *Toxoplasma* infection and elevated IgG antibody. Individuals with dormant *Toxoplasma* may have elevations of IgG antibody for months or years following the infection (7). Under this scenario, *Toxoplasma* IgG antibody, rather than the organism or a toxic product, may cross the placenta and cause damage to the developing fetal brain. IgG antibodies from women with spontaneous abortion (29) and from subjects with systemic lupus erythematosus (30) are known to cause teratogenic effects.

Our findings may help to distinguish between these possible explanations. The level of IgG antibody titer is generally correlated with both the severity and recency of infection. Thus, antibody titers in the “high” category are more likely to be associated with a current or recent reactivated infection than titers in the “moderate” category, which have a greater probability of reflecting dormant infection.

Given associations between prenatal exposure to other infectious agents and the risk of schizophrenia (1, 2), it also possible that an alteration of maternal immune status may have accounted for the findings. We aim to test this hypothesis in future work.

In a previous study from the Collaborative Perinatal Project, no association was found between antibody to *Toxoplasma* and the risk of adult psychosis (5). The group size of this study was small, however, and there was greater heterogeneity of psychotic disorders than the present study. Furthermore, the previous study quantified IgG antibody by solid-phase enzyme immunoassay rather than by the Sabin-Feldman dye test, the reference standard because of its high sensitivity and specificity (7).

**Limitations**

The sera in the present study had been frozen for over 30 years, raising the possibility that storage for this period of time may have altered the *Toxoplasma* antibody levels. However, this factor appears unlikely to have had an appreciable impact on our results for several reasons. First, *Toxoplasma* antibody levels are generally stable in frozen stored sera. Second, the seroprevalence of toxoplasmosis in comparison subjects was 17.9%, similar to the 17.5% seroprevalence found in a large previous study of toxoplasmosis in reproductive-age women in the United States (31). Third, we matched the comparison subjects to subjects with schizophrenia spectrum disorders on the date of birth and gestational timing, and the samples were uniformly handled and stored, indicating that storage time should not have biased the associations.

Second, we should consider the potential impact on our findings of associations between determinants of toxoplasmosis and factors related to maternal lifestyle or health. Toxoplasmosis is generally acquired by eating raw or undercooked meat containing *Toxoplasma gondii* tissue cysts, by ingesting oocysts from soil through activities such as gardening or eating unwashed vegetables or fruits, or possibly by exposure secondary to changing cat litter boxes (31). It is conceivable that mothers of future patients with schizophrenia, in relation to mothers of comparison subjects, were more likely to engage in these activities. It is also possible that factors related to determinants of toxoplasmosis and to maternal lifestyle and health in mothers of schizophrenia patients may have confounded the observed association. Maternal or family history of schizophrenia might be considered such a factor. We have not yet acquired sufficient data on family history of schizophrenia spectrum disorders to examine potential confounding by family history. There is no clear reason, however, to postulate that maternal toxoplasmosis should be related to maternal or other family history of schizophrenia after adjustment for potential demographic risk factors. Even if such a relation exists, the effect of these factors would need to be very substantial to account for the observed associations, and we did adjust for the potential confounders available in our data set (age, social class, or ethnicity); nonetheless, we cannot entirely rule out this possibility.

Third, the finding was marginally significant, and the group size was modest. Thus, independent replication of this result is essential.

**Conclusions**

In a birth cohort with archived prenatal sera, we demonstrated that elevated maternal IgG antibody to *Toxoplasma* is associated with an increased risk of schizophrenia and...
other schizophrenia spectrum disorders. Toxoplasmosis is a potentially preventable infection. Indeed, standard obstetric practice has included recommendations to pregnant women aimed at minimizing exposure to Toxoplasma, including the avoidance of cat feces and the ingestion of undercooked meat, which may contain Toxoplasma gondii oocysts (32). Toxoplasmosis has also been effectively treated in infants with antibiotics, such as pyrimethamine and sulfadiazine, which appear to reduce neurologic and other sequelae of the infection (7).

These findings add to a growing literature suggesting a relationship between in utero exposure to infectious agents that are known to disrupt fetal brain development and the risk of adult schizophrenia (1). Although replication will be necessary, these results may have implications for the prevention of schizophrenia.

References

32. CDC recommendations regarding selected conditions affecting women’s health. MMWR Recomm Rep 2000; 49(RR-2):57–75
In many different countries, severe mental disorders have been reported to occur 5–10 times more frequently among people in prison than in the general population (1, 2). However, it is unclear whether the high rates of psychotic mental disorders in prison are real or artifactual (3, 4). Alternative explanations should be considered. Differences in prevalence rates could be explained by 1) differences in sampling, which would affect for example sociodemographic characteristics of the populations studied (1); 2) differences in ascertainment methods (1); 3) differences in clinical syndromal profiles of cases (5); or 4) the toxic or withdrawal effects of psychoactive substances on mental functioning. People with psychosis live mainly in private households, but appreciable numbers are in temporary or homeless accommodations and in long-term institutions, including a range of different types of prison. In Great Britain, both primary healthcare and prison healthcare are funded from central taxation. A program of surveys assessing psychiatric morbidity and receipt of treatment was commenced during the 1990s in Great Britain (6) in order to determine progress toward health policy objectives. A program of surveys assessing psychiatric morbidity and receipt of treatment was commenced during the 1990s in Great Britain (6) in order to determine progress toward health policy objectives. Nationally representative randomly selected samples of the private household and adult prison populations were selected from appropriate sampling frames. The aim of this study was to compare rates of psychotic mental disorder in these two national samples using identical methods of ascertainment and to seek explanation for the differences found.

**Method**

Psychiatric morbidity surveys were carried out in the general population throughout Great Britain (6). Interviewers with a minimum of 3 years experience with the Office for Population Censuses and Surveys (now the Office for National Statistics) carried out initial interviews. Delivery points (N=18,000) were drawn from the continuously updated Small Area Postcode Address File, stratifying for socioeconomic grouping within the English regions, Wales, and Scotland (7). Adults (N=12,730) were selected from 15,765 private households (Figure 1). Subjects selected for interview were those who 1) endorsed any of the five symptom groups (mania, thought disorder, paranoia, delusions, or auditory hallucinations) covered by questions from the self-report Psychosis Screening Questionnaire (8); 2) reported that they were taking antipsychotic medication; or 3) reported that they had been given a diagnosis of psychotic illness by a physician. Interviews were performed as soon afterward as possible by clinicians trained in the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (9). Survey preparatory fieldwork had previously shown that psychosis cases identified by SCAN were not missed using these selection rules.

The SCAN is a semistructured interview covering axis I nonpsychotic and psychotic disorders, substance-related disorders, and organic brain disorders. Specific numbered sections of the SCAN cover particular syndromal groups or types of symptoms. For ex-
and psychotic symptoms, there are numbered sections covering clearcut hallucinations (section 17), subjectively described thought disorder and experiences of will replacement (section 18), and delusions (section 19) (9). The SCAN was developed by a joint task force of the World Health Organization and the U.S. Alcohol, Drug, and Mental Health Administration (9). Independently assessed agreement (using kappa) has been shown to exceed 0.8 for schizophrenia and affective psychosis (10, 11) and substance-related disorders (10). The SCAN consists of an interview guide providing the wording of questions about symptoms and a detailed glossary of symptoms. Every symptom and its threshold is defined in the glossary, and clinically experienced interviewers are trained to use question probes until the presence of each symptom can be confirmed or ruled out (12). In contrast to self-report questionnaires, in the SCAN the interviewer judges whether symptoms are present. Each type of psychotic symptom was rated individually for the year prior to interview. As part of the SCAN, clinicians also judged whether a symptom was attributable to any of a list of established toxic or withdrawal effects of alcohol or drug use listed in the SCAN glossary. We used published ICD-10 diagnostic algorithms (13) applied to SCAN symptom ratings to establish the presence of nonorganic psychosis in the year before interview, using ICD-10 codes F20–F31 (schizophrenia, delusional disorder, schizoaffective disorder, manic or bipolar affective disorder) and the ICD-10 F32–33 codes (depressive disorder) that require psychotic symptoms (14). Interviewers also coded any clearly established medical and organic causes such as epilepsy or brain damage according to ICD-10 criteria. In the absence of a SCAN interview in the household sample, a project diagnosis of probable functional psychosis was still made for anyone who reported taking antipsychotic medication and who also reported that they had a psychotic illness or that their doctor had told them that they had a psychotic illness, since this combination of responses was found to be most closely related to the presence of psychosis in those also clinically assessed (15).

All 131 prisons in England and Wales—was used to select prisoners according to the following sampling fractions: one in 34 male sentenced prisoners (1:50 in the final 4 weeks of the survey); one in eight remanded men (pretrial custody prisoners); and one in three of all women prisoners (16). Past research had shown different rates of disorder in these groups. These sampling fractions were therefore chosen in order to achieve sufficient numbers of interviews in these groups to provide percentages with standard errors of typically 1% and rarely more than 2%. A 1-in-5 random subsample of those who consented to participate were also interviewed with the SCAN (Figure 1). Symptoms in the past year were rated for diagnostic purposes as in the household survey. However, ratings for the past month were also made in order to register more transient symptoms of psychosis. In those not selected, a project diagnosis of probable functional psychosis (16) was made for anyone fulfilling any two of the following: a positive response to Psychosis Screening Questionnaire question “hearing voices”; a self-reported diagnosis of psychotic disorder; current antipsychotic medication; or a history of mental illness (admission to a mental hospital or mental illness in medical records), since this combination of responses was found to be most closely related to the presence of psychosis in those also assessed with the SCAN (16). The criteria for a project diagnosis of probable functional psychosis differed between the two surveys because, compared with those selected for assessment by a psychiatrist, the probability of being psychotic in those who were not selected for the SCAN interview was the same in the prison sample, whereas it was intended to be zero in the household sample.

In both surveys, written informed consent was obtained after the survey procedures had been fully explained. Standardized questions were asked covering employment and education, general health, access to health services, and treatment for general health and mental health problems. Respondents were asked the question: “In the past 12 months, have you spoken to a general practitioner or family doctor on your own behalf, either in person or by telephone, about being anxious, or depressed or a mental, nervous, or emotional problem?” They were also asked “Are you taking any pills or tablets or any other medicine by mouth which have been prescribed for you?” Medicines were checked...
and coded during subsequent data entry according to the British National Formulary (17). In the prison survey, clinical interviewers also had access to prison health care records covering time in custody.

We compared the two samples in terms of sociodemographic profiles and the prevalence of probable functional psychosis in the past year. Such comparisons can be biased for two reasons: first, because of differences between the characteristics of a sample and those of the population from which it is drawn and second, because statistical adjustments made to correct for these differences can appear to be overprecise unless the method chosen for calculating standard errors takes into account the adjustment method. Prevalence estimates were adjusted by the use of weights calculated to reflect differences between the original characteristics of the populations studied and the interviewed samples. Replication (resampling) methods were used for analysis to produce valid estimates of variances for data with complex sampling structures including clustering and nonresponse (18). A finite population correction was made for the female prisoners, since more than 20% had been sampled. In order to explore factors possibly involved in prevalence differences between the two samples, we reexamined the prison sample after adjustment according to the household population characteristics. Differences between the two samples in clinical profile were evaluated by comparing the proportion of specific types of psychotic symptoms in prison and household respondents in whom the SCAN assessment had been completed. Use of health services and of treatments was also compared.

### Results

Response rates in both surveys were good (7, 16) (Figure 1). Of 12,730 eligible households, 10,108 adults (79%) cooperated. Of 3,563 eligible sampled prisoners, 3,142 (88%) cooperated. Of 749 household residents who had positive psychosis screen results, 473 (63%) had a SCAN interview by a clinician. Of 661 prisoners who completed the initial survey interview and who were randomly selected for a follow-up clinical interview, 505 (76%) completed it.

In mid-1997, the overall prison population of England and Wales (N=61,944) included 2,770 women prisoners and 12,302 male remanded prisoners, the remainder being sentenced males (16). Sociodemographic characteristics of the samples are presented in Table 1. Most prisoners were under 35 years of age, and one-quarter of the male remanded prisoners were 16 to 19 years of age, although a smaller proportion of sentenced males were in this age group (16%). Of the male remanded prisoners, 80% classified themselves as white, which was almost the same as among male sentenced prisoners (84%), female remanded prisoners (77%), and female sentenced prisoners (75%) but less than among household survey respondents (94%).

### Prevalence of Psychosis

The prevalence of probable functional psychosis in the past year, adjusted for nonresponse and design clustering, was 4.5 per thousand (95% CI=3.1–5.8) in the household survey. In the prison survey, the adjusted rate was over 10 times greater: 52 per thousand (95% CI=45–60). Among male and female prisoners, rates were 50 (95% CI=42–57) and 110 (95% CI=100–120) per thousand, respectively.

Possible reasons for these prevalence differences were then considered. The prison population consisted substantially of younger adult male subjects. In order to take into account age and gender differences between the household and prison populations, we estimated the prison rate of probable functional psychosis, adjusted for the characteristics of the household population, as 70 per

---

**TABLE 1. Sociodemographic Characteristics of Male and Female Subjects From Surveys of British Households and Prisons to Assess Prevalence of Psychosis**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Male Remanded Prisoners (N=1,250)</th>
<th>Female Remanded Prisoners (N=187)</th>
<th>Male Sentenced Prisoners (N=1,121)</th>
<th>Female Sentenced Prisoners (N=584)</th>
<th>Male Household Residents (N=4,513)</th>
<th>Female Household Residents (N=5,308)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16–19</td>
<td>25</td>
<td>14</td>
<td>16</td>
<td>14</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>20–24</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>14</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>25–29</td>
<td>21</td>
<td>28</td>
<td>22</td>
<td>23</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>30–34</td>
<td>16</td>
<td>20</td>
<td>16</td>
<td>21</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>35–39</td>
<td>8</td>
<td>6</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>40–44</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>≥45</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>11</td>
<td>11</td>
<td>15</td>
<td>14</td>
<td>56</td>
<td>55</td>
</tr>
<tr>
<td>Cohabiting</td>
<td>37</td>
<td>45</td>
<td>42</td>
<td>39</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Single</td>
<td>44</td>
<td>33</td>
<td>34</td>
<td>32</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>Widowed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Divorced</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Separated</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working (full- or part-time)</td>
<td>36</td>
<td>26</td>
<td>44</td>
<td>34</td>
<td>73</td>
<td>60</td>
</tr>
<tr>
<td>Unemployed</td>
<td>34</td>
<td>24</td>
<td>28</td>
<td>23</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Economically inactive</td>
<td>30</td>
<td>50</td>
<td>27</td>
<td>43</td>
<td>15</td>
<td>34</td>
</tr>
</tbody>
</table>

*Living off crime, having a long-term illness, bringing up a family, other (retired, student, visiting the country).*
Crude rates in subgroups of prison respondents undergoing the SCAN were also compared with rates of probable psychosis in the full samples. The unadjusted rate (per thousand) of psychotic mental disorder in 394 male prisoners who underwent a full SCAN clinical interview was 100 (95% CI=61–139) in remanded prisoners and 70 (95% CI=31–101) in sentenced prisoners. Unadjusted probable psychosis rates in 2,371 male remanded and sentenced prisoners were 90 (95% CI=70–110) and 40 (95% CI=20–60), respectively. Probable psychosis rates in prisoners who described themselves as black African, African Caribbean, and “black other” were low in male remanded prisoners (20 per thousand [95% CI=0–40]) and male sentenced prisoners (30 per thousand [95% CI=10–50]).

## Frequency of Specific Psychotic Phenomena

Although the small number of subjects with active clinically significant psychotic symptoms limited comparisons, unadjusted comparisons of clinical symptom profiles were carried out by using SCAN interview ratings from the two surveys. Approximately 8% of household survey respondents were selected for the SCAN, and 473 completed SCAN interviews were available (Figure 1); the 1-in-5 random subsample of prisoners (N=661) selected for a SCAN interview yielded 505 completed interviews. In general, we found remarkably similar frequencies of psychotic phenomena in prisoners and household residents (Table 2). Subjectively described thought disorder and experience of replacement of will (SCAN section 18) was infrequently rated but was no more common among prisoners than among household residents (Table 2). Organic factors were seldom judged to influence such symptoms.

Nonverbal and olfactory hallucinations appeared somewhat more prevalent in household residents with psychosis (Table 3). Visual hallucinations were found both in household residents and particularly in prisoners categorized with organic psychosis but, even when this was taken into account, it did not alter the small differences between household residents and prisoners (Table 3). No type of delusion occurred more often in prisoners (Table 3).

As expected, psychotic mental health problems attributed by the interviewers to toxic or withdrawal effects of alcohol or drug use were more common in the prison population. Among prisoners with SCAN-rated nonaffective psychotic symptoms (SCAN sections 17 to 19), 25% were judged to have such an origin. Two prisoners with nonaffective psychotic symptoms rated on the SCAN had ratings attributed to a possible organic cause. In only one prisoner whose psychotic mental health problems were attributable to alcohol or drug use was the organic cause judged clearly to be the major cause of those symptoms. No respondent in the household population with SCAN ratings of nonaffective psychotic symptoms received a clinician attribution of their psychosis to use of alcohol or drugs. Using the SCAN data, we found hardly any difference between the annual and 1-month psychosis estimate in prison.

Prison survey respondents who did not undergo a SCAN interview were classified as having probable functional psychosis if they endorsed two criteria: current psychotic experiences between the annual and 1-month psychosis estimate in prison.

**Probable Psychosis and the Use of Services and Medication**

Of adults from the household survey with a project diagnosis of probable functional psychosis, 63% had consulted a general practitioner (primary care physician) for a mental health problem during the previous year; the rates among the prison populations were 40% for remanded male prisoners, 37% in sentenced male prisoners, and 57% in sentenced female prisoners. Across all samples, female remanded prisoners were most likely to report having been admitted to a mental hospital (22%); only 8% of male sentenced prisoners had ever been admitted to a mental hospital. These proportions are substantially higher than those reported in the general population (19). Use of prescribed medication for psychosis and related conditions was remarkably similar in the two groups with probable psychosis: 34% among household residents and 38% and 34% in male remanded and sentenced prisoners, respectively. However, in the much smaller female prison population, the rates were 53% and 62% in remanded and sentenced prisoners, respectively.

## Discussion

Reasons for the substantially higher prevalence of psychosis in the prison population need to be examined critically. Prevalence estimates based on the smaller number of prisoners interviewed clinically confirmed the high estimates made in those not selected for interview with the
SCAN. In the two-phase household survey, respondents whose psychosis screen results were negative were not clinically evaluated (Figure 1). If 200 of the over 9,000 negative screen results had been clinically assessed, the discovery of 10 missed psychosis cases would have been necessary to explain the excess prevalence of prison psychosis. This number of false negatives seems unlikely, given that the selection rule was derived from an extensive presurvey fieldwork study (15). However, a second British household survey has since been completed that included the same selection questions and a modified design in which SCAN interviews were also carried out in a random sample who would not have been selected; this design did yield some initially missed positive cases. The report of this second household survey includes a detailed discussion of the psychosis prevalence estimate (20): taking into account the subjects with negative initial screen results, the psychosis prevalence estimate for household residents was 11.1 per thousand (95% CI=5.2–17.0), which was still well below the rate found in the present sample of prisoners randomly selected to undergo a SCAN assessment.

Another explanation for higher rates of psychosis in the prison population might lie in between-group differences in age and socioeconomic status (Table 1). Out of interest, we reweighted the prison data to have the same age, gender, and socioeconomic status as the general population, but this did not explain the increased prevalence in the prisons. Drug misuse and dependency may be significant prior risk factors for the development of severe mental disorder in vulnerable persons (21). In the present prison survey we found that a history of first use of amphetamines or cocaine before the age of 16 and severe cannabis or cocaine dependence were related to an increased risk of probable functional psychosis (22). However, our clinical interviewers could only attribute the presence of psychotic symptoms to the effects of such substances in approximately 1 in 4 psychosis cases. The relationship between psychoactive substance misuse and the development of nonorganic psychotic mental disorder appears to be associated with earlier age at initiation of drug use and with severity of dependence (22), possible indicators of prolonged heavy use. To date, it is not clear how the effects of chronic cannabinoid and psychostimulant use, the main substances implicated, contribute to or aggravate psychotic disorders.

Some differences in receipt of services and treatments for psychosis were found, but again these were not sufficient to explain the overall difference in psychosis prevalence. The use of this information could be limited by the uncertain reliability of self-report data on treatment and service use. Our detailed examination of individual psychosis syndromes and types of symptoms did not point to

### TABLE 3. Specific Psychotic Symptoms in the Preceding Year Reported by British Household and Prison Survey Respondents

<table>
<thead>
<tr>
<th>Psychotic Phenomenon and Symptom</th>
<th>Item Numbera</th>
<th>Prevalence in Preceding Year (%)</th>
<th>Household Survey Respondents (N=473)</th>
<th>Prison Survey Respondents (N=505)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hallucinations (SCAN section 17)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonverbal auditory hallucinations</td>
<td>17.3</td>
<td>23</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Internal auditory hallucinations</td>
<td>17.7</td>
<td>37</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Auditory hallucinations with voices commenting on respondent</td>
<td>17.8</td>
<td>28</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Second-person auditory hallucinations</td>
<td>17.9</td>
<td>27</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Third-person auditory hallucinations (with or without second person auditory hallucinations)</td>
<td>17.9</td>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Mood-incongruent auditory hallucinations</td>
<td>17.10</td>
<td>15</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Special features of auditory hallucinations</td>
<td>17.12</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Visual hallucinations unformed (after clinically excluding drug and organic causes)</td>
<td>17.15</td>
<td>17</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Visual hallucinations fully formed</td>
<td>17.16</td>
<td>9</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Scenic visual hallucinations</td>
<td>17.17</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Olfactory hallucinations</td>
<td>17.22</td>
<td>14</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Subjectively described thought disorder and experience of replacement of will (SCAN section 18)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delusional mood and perplexity</td>
<td>18.1</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Delusions that thoughts are being read</td>
<td>18.2</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Thought insertion</td>
<td>18.3</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Thought broadcast</td>
<td>18.7</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Thought comment</td>
<td>18.8</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Thought block</td>
<td>18.9</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Delusions (SCAN section 19)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delusions of being spied upon</td>
<td>19.3</td>
<td>13</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Delusions of reference</td>
<td>19.4</td>
<td>14</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Delusional misinterpretation</td>
<td>19.5</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Quotation of ideas</td>
<td>19.6</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Delusional ideas of reference based on guilt</td>
<td>19.10</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Delusional ideas of reference based on expansive mood</td>
<td>19.11</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Delusions of persecution</td>
<td>19.12</td>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Delusions of conspiracy</td>
<td>19.13</td>
<td>12</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

From the Schedules for Clinical Assessment in Neuropsychiatry (9).
between-group differences in the proportion of case types for the two populations. It only served to confirm the much higher frequency of equivalent, recognizable psychotic phenomena in prisoners.

**Previous Survey Comparisons**

Other general population surveys that used similar assessment techniques have found similar prevalence rates to the Great Britain household study included here. For example, using the SCAN Jablensky and colleagues found a 1-year prevalence rate of psychosis between 4 and 7 per thousand in the adult population of urban Australia (2). Prison surveys have generally found higher rates but have used different assessment approaches. A systematic review of 62 surveys from 12 countries found a psychotic illness rate of 37 per thousand (95% CI=33–41) in 18,530 male prisoners, with a combined estimate of 4 times the general population rate (1). Unfortunately, the different time periods used in this combined estimate were not specified (1). Only one previous systematic direct comparison of prison and community data has been made (23): as determined with a fully structured interview, the lifetime rate of manic episodes for the prison population of Chicago was significantly higher than the community rate (Epidemiologic Catchment Area survey [24]), but the lifetime rate for schizophrenia/schizophreniform disorder was not significantly higher. However, doubts about the high rate of psychosis in prisons have persisted since the influential study of 2,070 adults in prisons throughout England in 1988 by Gunn et al. (25), who reported the prevalence (in the past month) of psychosis “warranting hospital treatment” to be 2% in 1,769 sentenced men and in 301 women. Case finding was based on ICD-9, a clinician rating of psychotic phenomena, and a consensus case conference for randomly sampled respondents and difficult cases. However, other work in prisons in England is in line with our findings: a consecutive series of 569 subjects recommended to one prison (26), evaluated with a structured clinical interview covering psychosis criteria in detail, found 70 lifetime cases per 1,000. Our psychosis cases may also include a wider range of psychotic disorders than the detailed SCAN clinical interview has identified. The prison psychosis rate may also have risen in the past decade because of the considerable expansion of prison places in England, the attendant closure of acute psychiatric in-patient beds, and the inadequacy during the intervening decade of court diversion and prison reception health screening (27).

**Implications**

Comparison of psychosis prevalence rates obtained from prison and household population surveys that used a similar ascertainment approach and involved 13,250 adults confirms that rates in prison are about 10 times higher. It appears unlikely that this excess can simply be attributed to the lower socioeconomic status and age of prisoners. Psychotic episodes in prison have the same form and clinical presentation of illness as those in the community, dispelling the notion that their clinical presentations are spurious or due to a putatively distinct “prison-psychosis” (5). The present British comparison provides firmer evidence that the substantially higher prevalence of psychotic disorder in prisons, which is also found in many other countries (1), is genuine and deserving of the full range of medical treatments and social care known to improve functioning and outcome. Policy makers and researchers now need to take these findings into account.

The public health policy implications of the high rate of severe mental illness in prisons and the lack of appropriate specialized care for this group have been commented upon (1, 28). When we compared adults with psychosis in the community with those in prison, fewer prisoners had consulted a doctor about their mental health problem in the last year. Good quality prison healthcare or mental health “in reach” services could provide effective treatment in many such cases. Elsewhere we have reported on a follow-up study of the prisoners with probable psychosis, most of whom had returned to the community at least once during the following year: less than a quarter had appointments with psychiatric professionals. For those with violent or sexual offenses, the rate only rose to 41% (29).

Health and social care reforms to address this through closer links between prison and primary health care and specialist services are said to be underway in England and Wales (30), but there seems to be little attention to the problem elsewhere.

A crucial question is what health care and risk management are provided for severely mentally disturbed adults, particularly those who have committed serious and violent offenses, following release into the community. This is a priority for future research. We also require more detailed studies of drug misuse in populations with a high prevalence of psychotic symptoms, such as prisons. A growing body of longitudinal research (31) will also help to clarify factors that might contribute to these prevalence differences, including previous history of drug use and dependence, the response of services, and the widespread introduction of effective treatments.
PSYCHOSIS PREVALENCE

Both surveys were funded by Department of Health (England) contracts to the Office for National Statistics, Social Survey Division, London, which carried out fieldwork.

References

Impact of Referral Source and Study Applicants’ Preference for Randomly Assigned Service on Research Enrollment, Service Engagement, and Evaluative Outcomes

Cathaleene Macias, Ph.D.
Paul Barreira, M.D.
William Hargreaves, Ph.D.
Leonard Bickman, Ph.D.
William Fisher, Ph.D.
Elliot Aronson, Ph.D.

Objective: The inability to blind research participants to their experimental conditions is the Achilles’ heel of mental health services research. When one experimental condition receives more disappointed participants, or more satisfied participants, research findings can be biased in spite of random assignment. The authors explored the potential for research participants’ preference for one experimental program over another to compromise the generalizability and validity of randomized controlled service evaluations as well as cross-study comparisons.

Method: Three Cox regression analyses measured the impact of applicants’ service assignment preference on research project enrollment, engagement in assigned services, and a service-related outcome, competitive employment.

Results: A stated service preference, referral by an agency with a low level of continuity in outpatient care, and willingness to switch from current services were significant positive predictors of research enrollment. Match to service assignment preference was a significant positive predictor of service engagement, and mismatch to assignment preference was a significant negative predictor of both service engagement and employment outcome.

Conclusions: Referral source type and service assignment preference should be routinely measured and statistically controlled for in all studies of mental health service effectiveness to provide a sound empirical base for evidence-based practice.

Participants in research on mental health services are rarely blind to their experimental assignments, and so attrition caused by disappointment in service assignment is a well-recognized risk. Researchers can minimize overall attrition, as well as the threat of differential attrition (1), by screening out applicants with a priori service preferences. However, some applicants may choose not to disclose their preference to ensure that they have a chance to be assigned to a favored program, or they may feel they do not have a personal preference over and above the obvious one posed by the study design. For instance, when the experimental comparison is between a new intervention and services as usual, many applicants may prefer the new program even before they learn much about it.

Unfortunately, random assignment will not always equalize preexisting study enrollee characteristics across experimental groups (2), and this will always be the case when study enrollees tend to prefer one service condition over another. Common sense tells us that service assignment preference must first be balanced within the total enrollee group for it to be distributed equitably across experimental conditions. For example, if 60% of enrollees have a preference for condition A and 40% have a preference for condition B, then, with true equivalence across conditions, service A would have 60% pleased and 40% disappointed assignees, while service B would have 40% pleased and 60% disappointed assignees. To the extent that participant satisfaction or disappointment influences service engagement and study outcomes (3, 4), research findings will be biased in spite of random assignment (5–7).

Previous experience with mental health services may influence applicants’ attitudes toward research participation and the experimental services offered by a study (8, 9). Applicants referred by agencies not designed to provide continuity in outpatient care (e.g., homeless shelters, emergency units) are typically more receptive to a study offering any type of continuous care than are applicants referred by mental health centers (10, 11). On the other hand, study applicants already engaged in outpatient services should be more likely to have a service preference. If a research study enrollee is assigned to a control condition that appears to be no better than previously received services, the enrollee may enter into the assigned program halfheartedly or refuse services altogether (1, 12). Likewise, if study enrollees have to relinquish a current service to participate in an experimental program, they may delay contacting the assigned program or withdraw from project participation (13, 14).
The present study investigates the impact of preference in experimental condition assignment on study applicants’ motivation to enroll in a mental health services research project, engage in assigned services, and pursue the project’s targeted outcomes.

**Method**

Data were taken from the Massachusetts Employment Intervention Project, which was funded from 1995 to 2000 through the Employment Intervention Demonstration Program of the Substance Abuse and Mental Health Services Administration (15).

**Experimental Programs**

Program of assertive community treatment (PACT) model. PACT (16, 17) is a mobile treatment team that provides a variety of mental health and medical services (18). The program in Worcester, Mass., was created by Leonard Stein, M.D., and Jana Frey, Ph.D., of Madison, Wis., in 1996. Fidelity was verified through annual site visits by Gary Bond, Ph.D., and Dr. Frey. Following the recommended 1:10 staff-to-consumer ratio, the ceiling on PACT enrollment was 90 participants, with an enrollment rate of five to seven participants per month.

**Clubhouse model**. A clubhouse (19) is a facility-based day program offering membership in a supportive community. A defining aspect is the 9–5 “work-ordered day” in which members and staff work side by side to perform voluntary work essential to the clubhouse (20). Genesis Club, Inc., in Worcester, which had an average annual enrollment of 400 members, was certified by the International Center for Clubhouse Development as having full compliance with the Standards for Clubhouse Programs (21).

**Study Group**

Individuals were eligible for the study if they resided in the vicinity of Worcester, were 18 years old or older, did not have severe retardation (IQ greater than 60), were currently unemployed, and were given a primary DSM-IV diagnosis of a schizophrenia spectrum disorder, major depression, or bipolar disorder. Diagnostic information was obtained through medical records. When there was no clear diagnosis, eligibility was confirmed by DSM-IV diagnostic assessments (22) conducted by psychiatric residents at the University of Massachusetts Medical School.

Recruitment was from February 1996 through May 1998. Presentations were made at local agencies and advocacy groups, and self-referrals were elicited through flyers, radio, and newspapers. Recruiters explained that study enrollees would be expected to slowly relinquish their current clinical services if assigned to PACT, or to relinquish any current day program if assigned to the clubhouse. During intake sessions, interviewers read a description of the study aloud, explained randomization, and asked each applicant if he or she would be willing to participate fully in either experimental program regardless of personal preference. Applicants who expressed understanding and agreed to participate provided written informed consent.

**Measures of Service Assignment Preference**

As Table 1 shows, one-half (51% [N=158]) of the 310 eligible applicants expressed a preference for being randomly assigned to one of the two experimental programs. Program recruiters recorded these preferences, and reasons for having a preference, verbatim. Assignment preferences were content-coded as pro-PACT, pro-clubhouse, and no expressed preference and then recorded after random assignment as match to preference, mismatch to preference, or no preference. Reluctance to relinquish existing services was dichotomized as reluctant to exchange current day program for clubhouse versus reluctant to exchange current clinical services for PACT.

**Measure of Type of Referral Source**

Applications were received from 42 local organizations, as well as through self-referrals and family referrals. Following previous research (10, 23), we dichotomized referral sources (or key provider agencies for self-referrals or family referrals) as having either a high level of continuity in outpatient care (routine treatment by clinicians, day programs, or case managers) or a low level of continuity in outpatient care (inpatient, emergency, and screening units of hospitals, shelters, residential treatment centers, nursing homes, or jails). As Table 1 shows, approximately 60% of both eligible and ineligible applicants were recruited from sources characterized as having a low level of continuity of care.

Type of referral source was unrelated to specific preference for PACT or clubhouse assignment, but enrollees from agencies with a high level of continuity of care were more likely than enrollees from agencies with a low level to have a service assignment preference (62% [N=32] versus 45% [N=56]) ($\chi^2=3.93, df=1, p<0.05$).

---

**TABLE 1. Progression of 465 Research Study Applicants Through the Massachusetts Employment Intervention Project and Participants’ Characteristics and Preferences for Randomly Assigned Treatment Conditions**

<table>
<thead>
<tr>
<th>Study Status</th>
<th>Referred by Agency With Low Level of Continuity</th>
<th>Number of Days to Status</th>
<th>Male Gender</th>
<th>Age (years)</th>
<th>Ethnic Minority</th>
<th>Diagnosis of Schizophrenia</th>
<th>Preference for PACT</th>
<th>Preference for Clubhouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied (N=465)</td>
<td>286 62</td>
<td>251 54</td>
<td>37.9 11</td>
<td>93 20</td>
<td>38.5 12</td>
<td>5 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lost (N=24)</td>
<td>15 63</td>
<td>20 83</td>
<td>38.5 12</td>
<td>5 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineligible (N=131)</td>
<td>81 62</td>
<td>77 59</td>
<td>37.0 11</td>
<td>30 23</td>
<td>38.2 11</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible (N=310)</td>
<td>190 61</td>
<td>154 50</td>
<td>38.7 11</td>
<td>58 19</td>
<td>67 22</td>
<td>91 29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refused (N=133)</td>
<td>65 49</td>
<td>58 44</td>
<td>38.6 11</td>
<td>21 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled (N=177)</td>
<td>125 71</td>
<td>42.6 60</td>
<td>38.1 10</td>
<td>37 21</td>
<td>91 51</td>
<td>45 25</td>
<td>43 24</td>
<td></td>
</tr>
<tr>
<td>PACT (N=88)</td>
<td>63 72</td>
<td>36.3 38</td>
<td>52 59</td>
<td>37.1 9</td>
<td>20 23</td>
<td>52 59</td>
<td>16 18</td>
<td>18 21</td>
</tr>
<tr>
<td>Clubhouse (N=89)</td>
<td>62 70</td>
<td>48.9 76</td>
<td>44 49</td>
<td>39.0 11</td>
<td>17 19</td>
<td>39 44</td>
<td>29 33</td>
<td>25 28</td>
</tr>
<tr>
<td>Engaged (N=150)</td>
<td>107 71</td>
<td>41.0 91</td>
<td>81 54</td>
<td>38.3 10</td>
<td>32 21</td>
<td>79 53</td>
<td>35 23</td>
<td>38 25</td>
</tr>
<tr>
<td>PACT (N=84)</td>
<td>60 71</td>
<td>15.3 19</td>
<td>49 58</td>
<td>37.1 9</td>
<td>20 24</td>
<td>49 58</td>
<td>16 19</td>
<td>16 19</td>
</tr>
<tr>
<td>Clubhouse (N=66)</td>
<td>47 71</td>
<td>73.7 129</td>
<td>32 49</td>
<td>39.7 11</td>
<td>12 18</td>
<td>30 46</td>
<td>19 29</td>
<td>22 33</td>
</tr>
<tr>
<td>First job (N=88)</td>
<td>62 71</td>
<td>221.0 196</td>
<td>46 52</td>
<td>37.0 10</td>
<td>18 21</td>
<td>44 50</td>
<td>24 27</td>
<td>15 17</td>
</tr>
<tr>
<td>PACT (N=47)</td>
<td>33 70</td>
<td>230.8 189</td>
<td>32 68</td>
<td>36.9 10</td>
<td>11 23</td>
<td>27 57</td>
<td>10 21</td>
<td>6 13</td>
</tr>
<tr>
<td>Clubhouse (N=41)</td>
<td>29 71</td>
<td>209.8 205</td>
<td>14 34</td>
<td>37.1 10</td>
<td>7 17</td>
<td>17 42</td>
<td>14 34</td>
<td>9 22</td>
</tr>
</tbody>
</table>

---

a Percents based on numbers given in first column.

b Program of assertive community treatment.
Statistical Analyses

Three survival (event history) analyses (24) were conducted to test the impact of recruitment variables on length of time to study enrollment, time to service engagement, and time to first competitive job. Survival analysis tracks the timing of specific events in relation to the participant’s date of study entry, regardless of individual length of observation period. SPSS version 11.0 (SPSS, Inc., Chicago) was used for all data analyses.

Time to study enrollment was measured as calendar days from application to date of random assignment to service. Period of observation was from application to the end of the recruitment period, which was comparable for enrolled (mean=462 days) and nonenrolled (mean=426 days) applicants. One applicant who took 18 months to enroll was dropped from the analysis as an extreme outlier. Applications received later in the recruitment process were processed more rapidly and, hence, were more likely to result in enrollment. Therefore, timing of application and timing of enrollment were correlated (r=–0.40, p<0.001), violating the assumption of stability of the dependent measure in respect to absolute time. For this reason, life tables were not constructed for this analysis of project enrollment, but timing of application was included as a control variable in the Cox regression.

Time to program engagement was measured as calendar days from application to date of random assignment to service. Period of observation was from application to the end of the recruitment period, which was comparable for enrolled (mean=462 days) and nonenrolled (mean=426 days) applicants. One applicant who took 18 months to enroll was dropped from the analysis as an extreme outlier. Applications received later in the recruitment process were processed more rapidly and, hence, were more likely to result in enrollment. Therefore, timing of application and timing of enrollment were correlated (r=–0.40, p<0.001), violating the assumption of stability of the dependent measure in respect to absolute time. For this reason, life tables were not constructed for this analysis of project enrollment, but timing of application was included as a control variable in the Cox regression.

Time to first job was measured as calendar days from enrollment to the first day of the participant’s first competitive job (any job located in a mainstream, integrated setting paying at least minimum wage and lasting more than 5 days). Time to first job for censored (never-employed) participants was coded as days from enrollment to the participant’s 24-month anniversary date, the follow-up timeframe adopted by the larger multisite-supported Employment Intervention Demonstration Program. Omitted from the analysis were five enrollees who died during the 24-month observation period, the PACT participant who refused use of his data, and another PACT participant who was a crossover to the clubhouse. As with time to program engagement, match to preference and no preference were combined as the reference category for mismatch to preference in the life table.

Results

Predictors of Research Study Enrollment

Table 1 presents an overview of participants’ progress through the research study. Study applicants who were ineligible or lost to follow-up (N=155) were more likely than eligible applicants (N=310) to be male (63% versus 50%) (χ²=6.93, df=1, p<0.01). Otherwise, status in the project was unrelated to demographic variables.

As expected, eligible applicants referred by agencies not designed to provide continuity in outpatient care (e.g., hospitals, shelters) enrolled in the project at a higher rate than applicants referred by community mental health centers and other agencies with a high level of continuity of care (66% [N=125] versus 43% [N=52]) (χ²=15.14, df=1, p<0.001). This was true for those with a service assignment preference (65% [N=56] versus 44% [N=31]) (χ²=7.25, df=1,
TABLE 2. Estimates From Cox Regression of Time to Enrollment on Gender, Referral Source, Service Assignment Preference, and Reluctance to Switch Services (N=309)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard Rate</th>
<th>Wald</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application timing</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Male gender</td>
<td>0.31</td>
<td>0.15</td>
<td>0.73</td>
</tr>
<tr>
<td>Referral source</td>
<td>0.64</td>
<td>0.17</td>
<td>0.53</td>
</tr>
<tr>
<td>Service preference</td>
<td>0.41</td>
<td>0.18</td>
<td>0.66</td>
</tr>
<tr>
<td>Reluctance to switch</td>
<td>0.70</td>
<td>0.22</td>
<td>2.01</td>
</tr>
</tbody>
</table>

*p Significant results for full model (χ²=34.04, df=5, p<0.0001).

Table 3 shows the results of a Cox regression analysis that controlled for date of application (Table 2) confirm that applicants from agencies with a low level of continuity of outpatient care enrolled more quickly than applicants from agencies with a high level of continuity of care. Applicants who preferred assignment to PACT also enrolled faster, and those reluctant to exchange current services for new ones took longer to enroll. As the hazard rate shows, applicants reluctant to switch services were twice as likely as other applicants to withdraw their applications. These same findings are obtained when ethnicity (five missing cases) and age (21 missing cases) were added as regression model covariates.

Predictors of Service Program Engagement

 Applicants from agencies with a high level of continuity of outpatient care had equivalent rates of service engagement (84%) to those from agencies with a low level of continuity of care (87%), and these two groups were comparable in demographics. On the other hand, fewer enrollees who were mismatched to their service preference (75% [N=35]) became active in their assigned program compared with matched enrollees (93% [N=38]) and enrollees with no service preference (90% [N=77]) (χ²=7.69, df=2, p<0.05).

As Table 1 shows, the clubhouse was assigned more enrollees who were matched to their service preference as well as more enrollees who had not wanted to be assigned there. That is, the clubhouse received 20% more participants than PACT who had any service preference (66% versus 40%) (χ²=9.10, df=1, p<0.01). Fortunately, preference-matched, mismatched, and no-preference enrollees had equivalent rates of schizophrenia spectrum disorder diagnoses (52%, 51%, and 52%, respectively) and were similar in other background variables, allowing statistical control of program preference in service engagement and employment analyses.
Predictors of Competitive Employment

Univariate analyses revealed no significant differences in employment rates between participants assigned to the service they wanted (46% [N=19]), those assigned to the service they did not want (43% [N=20]), and those who had no assignment preference (58% [N=49]), or between referees from agencies high (50% [N=26]) versus low (50% [N=62]) in continuity of outpatient care. Competitive employment rates were fairly comparable for PACT (53% [N=47]) and clubhouse (47% [N=41]), and participants in both programs began their first competitive job an average of 7 months after enrollment. A fourth of all employed PACT participants (26% [N=12]) and 42% of employed clubhouse participants (N=17) took their first competitive job within 90 days. Interestingly, employed PACT participants were more likely to be men (68% [N=32]), while employed clubhouse participants were more likely to be women (66% [N=27]) (χ²=10.11, df=1, p<0.01).

A Cox regression analysis (Table 4) revealed a significant main effect for mismatch to service preference that paralleled the main effect obtained in the Cox regression analysis of service engagement and contradicts the univariate employment findings. When we controlled for length of time to service engagement, program assignment, demographics, and two work-related background variables (baseline interest in work and receipt of at least 1 hour of vocational services), enrollees mismatched to their service preference took significantly longer to begin a competitive job than those who had no service preference. As the hazard rate shows, applicants mismatched to a service they did not want were twice as likely to never become employed as applicants with no service preference. When entered as a separate block, match and mismatch significantly improved fit of the regression model: the difference between the –2 log likelihoods of the full versus the reduced model was significant (χ²=17.72, df=9, p<0.001).

Survival Function Plots

Figure 1 shows the survival functions for PACT and clubhouse participants mismatched to their service preference compared with participants either matched with their preferred service or without a preference. The left panel plots these unadjusted survival functions for time to service engagement, and the right panel plots the functions for time to service engagement.
Discussion

The results of the enrollment analysis demonstrate that service preferences can influence research project enrollment and limit sample representativeness to the extent that preference correlates with previous service experience or consumer characteristics. Results of the service engagement and employment analyses demonstrate that applicant service preference continues to be salient even after project enrollment, predicting both engagement in the assigned service and motivation to pursue service-related outcomes. Applicants randomly matched to their service preference were 64% more likely than applicants with no preference to become engaged in assigned services. Applicants mismatched to their service preference were 77% less likely to become engaged in services and twice as likely to never be competitively employed. Since service preference was relatively balanced within this comparison of two high-quality interventions (27), and since the clubhouse condition received by chance more pro-PACT and more pro-clubhouse enrollees, the effects of service preference were not immediately obvious. However, had a majority of the enrollees favored one program over another, even perfect equivalence through random assignment would have given the preferred program an unfair advantage, i.e., more matched and fewer mismatched enrollees. For this reason, applicant preferences for program assignment should be routinely measured and statistically controlled for in all mental health service evaluations.

Extent of the Problem
Within Mental Health Services Research

Only a handful of published service evaluations have reported and controlled for diversity in sample recruitment sources or study applicants’ preferences in service assignment. Even fewer published studies describe experimental program requirements for relinquishing current services that would duplicate experimental ones, and it is unclear how often such expectations appear in informed consents. This widespread disregard of the potential for service assignment preference to bias research findings is particularly disheartening because the design of most evidence-based (randomized controlled) service evaluations predispose applicants to prefer one experimental condition over the other. Many service evaluations compare a new or better service with services as usual or with a control condition not designed to achieve the targeted outcomes, as when day treatment is the control condition for a study of competitive employment. Service preference is confounded with experimental program differences in such studies and, unless measured, is impossible to statistically control. For this reason, evidence-based practices that have been evaluated by randomized controlled evaluations that did not document and statistically control for enrollees’ service assignment preference may still lack evidence of effectiveness (28).

Recommendations for Future Research

Our study findings underscore the importance of routinely documenting research applicants’ current services, any reluctance to relinquish an existing service, and preference or dislike for experimental services. Our recruiters asked applicants for a verbal commitment to participate in either experimental service if randomly assigned there, and applicant responses to planned probes were recorded verbatim. Other researchers may also benefit from asking open-ended questions and tracking verbatim dichotomous responses. Checklists and rating scale measures could also be constructed. However, this added burden to researchers may be unreasonable, given the demonstrated strength of categorical predictors in the present study. The more difficult decision is whether to limit the sample to only those individuals who have no service preference or, instead, to invite all eligible applicants to enroll and statistically control for preexisting preferences. The former solution appears more rigorous if eligible nonenrollees can be tracked in the same way as enrollees (29, 30). The latter alternative is more practical if recruitment is costly or slow.

A pilot survey of the target population could provide an estimate of the prevalence of service preferences to guide research design decisions (31). The high rate of expressed preferences among our own study enrollees suggests that we would have greatly limited our study group’s size and representativeness had we chosen to include only applicants with no preference. Also, applicants sometimes indicated that they would hide their service preference if it would exclude them from the study, so no service preference as an inclusion criterion may be unfeasible. Regardless, recruiters should carefully explain randomization and expectations for participation in each study condition, especially if applicants are active in services that might duplicate or compete with experimental services. The applicant may not have a service assignment preference until he or she fully understands the research study.
Conclusions

The randomized controlled trial is the gold standard for psychiatric research, and it is well suited to most clinical intervention and pharmaceutical studies. However, inability to blind participants to their experimental assignment is the Achilles’ heel of services research. We must begin to measure and control for applicant preference in random assignment if we are to build a strong empirical base for evidence-based practice in mental health services.

Received Jan. 31, 2003; revision received July 15, 2003; accepted April 9, 2004. From McLean Hospital, Harvard Medical School; the University of California, San Francisco; Vanderbilt University, Nashville, Tenn.; the University of Massachusetts Medical School, Worcester; and Stanford University, Stanford, Calif. Address correspondence and reprint requests to Dr. Macias, Community Intervention Research, McLean Hospital, 115 Mill St., Belmont, MA 02478; cmacias@mclean.harvard.edu (e-mail).

Supported by NIMH grant MH-060828, Substance Abuse and Mental Health Services Administration collaborative agreement SM-51831, the van Ameringen Foundation, the Rhodebeck Charitable Trust, and Llewellyn & Nicholas Nicholas.

The authors thank William Shadish, Danson Jones, and Charles Rodican for their thoughtful critiques of the manuscript.

References

Phenylthiocarbamide Perception in Patients With Schizophrenia and First-Degree Family Members

Paul J. Moberg, Ph.D.
David R. Roalf, B.S.
Catherine C. Balderston, M.S.
Stephen J. Kanes, M.D., Ph.D.
Raquel E. Gur, M.D., Ph.D.
Bruce I. Turetsky, M.D.

Objective: The inability to taste phenylthiocarbamide (PTC) has been associated with medical and neurological illnesses not typically related to taste. The authors examined PTC sensitivity in schizophrenia patients and their non-ill relatives to determine whether this represented a vulnerability marker.

Method: PTC sensitivity was assessed in 42 schizophrenia patients, 23 healthy comparison subjects, and 12 first-degree relatives of the patients.

Results: More nontasters were found among patients and family members than healthy comparison subjects. Among patients, nontasters had more positive symptoms. Differences were not explained by sex, age, medication, smoking, or cognitive impairment.

Conclusions: The prevalence of PTC nontasters was greater among schizophrenia patients and non-ill first-degree family members. Phenotypic variation in PTC sensitivity is genetic in origin. This suggests a higher risk for illness among subjects with recessive alleles.

(S) tudies of sensitivity to the bitter-tasting antithyroid compound phenylthiocarbamide (PTC) have shown this to be an inherited trait determined by a dominant allele (1, 2). A G protein-coupled bitter taste receptor, TAS2R38 (also known as PTC or TAS2R), located on chromosome 7q, accounts for 85% of the variance in taste (3). A second minor locus is found at 16p (4). About 30% of the U.S. population are PTC “nontasters” (i.e., tt), while approximately 70% are “tasters” (i.e., TT or Tt) (4, 5). Although nontaster status has been linked to a variety of medical disorders, there have been few investigations of PTC taster status in schizophrenia (6–8). Abnormalities in the function and/or expression of G protein signaling pathways have been reported in patients with schizophrenia and appear to be implicated in prepulse inhibition of the startle reflex and negative symptoms (9, 10). Therefore, we anticipated that patients would show a differential pattern of PTC tasting status relative to healthy comparison subjects and that a similar pattern would be observed in their non-ill first-degree relatives.

Method

Forty-two patients with schizophrenia (30 men and 12 women), 23 healthy volunteers (11 men and 12 women), and 12 first-degree relatives (four men and eight women) were recruited from the University of Pennsylvania Schizophrenia Research Center. The relative cohort included one parent, eight siblings, and three adult offspring of six patients. The patients received the Structured Clinical Interview for DSM-IV, Patient Edition (11), a physical examination, and routine laboratory tests. The patients were rated on the Brief Psychiatric Rating Scale (BPRS) (12), the Scale for the Assessment of Positive Symptoms (13), and the Scale for the Assessment of Negative Symptoms (13), and the Scale for the Assessment of Positive Symptoms (14). Family members were assessed with the Structured Clinical Interview for Personality Disorders (15). All probands met DSM-IV criteria for schizophrenia with no other concurrent diagnoses. Family members and healthy comparison subjects were free of any current axis I diagnosis or axis II cluster A personality disorder.

Subjects were excluded for a history of neurological disorder, head trauma, loss of consciousness, substance abuse/dependence, a medical condition that might alter cerebral functioning, a recent respiratory infection, or any condition that could affect taste functioning. Written informed consent was obtained after a complete description of the study was given to the subjects.

There were no differences in age among the patients (mean=36.7, SD=12.7), family members (mean=41.2, SD=18.8), and healthy comparison subjects (mean=30.8, SD=12.8) (F=2.5, df=2, 74, p=0.09). The probands had a greater proportion of African Americans than the healthy comparison subjects (χ²=12.8, df=6, p=0.05). All patients were stable outpatients at the time of testing. Mean duration of illness was 11.8 years (SD=8.6). Twenty-five patients were receiving atypical antipsychotic medications, six were receiving typical antipsychotics, and 11 were unmedicated at the time of testing. The mean dose was 323.3 mg/day (SD=213.6) in chlorpromazine equivalents. The mean BPRS score was 30.4 (SD=8.6), indicating a low level of acute symptoms.

A PTC-impregnated strip of filter paper (Carolina Biological Supply Company, Burlington, N.C.) was placed on the tongue. The subjects were asked if they tasted anything. They were then asked to rate the intensity on a 100-mm visual analog line, ranging from 0 mm (no taste) to 100 mm (extremely strong taste). Any subject who reported an inability to taste the filter paper and an intensity of less than 6 mm was classified as a nontaster. The 6-mm value denoted the break point in the bimodal distribution of intensity ratings between the tasters and nontasters.

Results

Statistical analysis revealed significant differences in the distribution of tasters and nontasters among the three groups (χ²=8.30, df=2, p=0.02), with both patients (χ²=7.53, df=1, p=0.006) and family members (χ²=4.70, df=1, p=0.03) having a larger proportion of nontasters than...
healthy volunteers (Figure 1). Patients and family members did not differ from one another ($\chi^2=0.01$, df=1, p=0.94). Among tasters, there were no group differences in intensity ratings and no effect of diagnosis, sex, or diagnosis-by-sex interaction (all p>0.18).

Within the patient group, tasters and nontasters did not differ in sex distribution, Mini-Mental State Examination (MMSE) score, age at onset, illness duration, illness severity, deficit status, or negative symptoms. The two subgroups did differ in positive symptom ratings ($F=5.3$, df=1, 33, p=0.03), with nontasters demonstrating higher total scores on the Scale for the Assessment of Positive Symptoms. This reflected higher ratings for the hallucinations ($F=7.6$, df=1, 33, p=0.009) and delusions ($F=5.3$, df=1, 33, p=0.03) subscales. Linear regression analyses with age, sex, chlorpromazine equivalents, age at onset, illness duration, smoking history, and MMSE score as predictors did not alter the observed differences in taster status or affect PTC intensity ratings.

Discussion

These data demonstrate a higher prevalence of PTC nontasters among patients with schizophrenia and their non-ill first-degree relatives. Among healthy comparison subjects, 78% were classified as tasters, in contrast to 43% among patients and 42% among family members. The higher prevalence of nontasters in the patient and family groups is consistent with three other reports of nontaster prevalences ranging from 42% to 47% among patients with schizophrenia (6–8). A higher prevalence of nontasters in patients could not be explained by a greater proportion of African Americans because this ethnic group shows a lower prevalence of nontasters than other populations (4). The presence of a similar impairment in first-degree relatives suggests that people with at least one dominant allele may be at a lower risk than those with two recessive alleles. The high frequency of nontasters in patients and family members may be due to abnormalities in the function and/or expression of G-protein signaling pathways. Drayna and colleagues (16) found that PTC binds to both forms of the receptor (i.e., taster and nontaster) with equal affinity, but the nontaster form fails to activate G-protein. This failure in G-protein signaling may interact with other genetic and/or environmental factors to produce an increased vulnerability to illness. This hypothesis requires further investigation.

A few caveats must be noted. First, the group of family members was relatively small, and a larger cohort will be required to fully assess this difference. However, it is notable that the effect size (Cohen’s d) for the difference in taster status between family members and healthy comparison subjects is quite large ($d_{1}=0.81$, 95% confidence interval [CI]=0.08–1.53), and we would expect this effect to be robust in independent samples. Second, the comparison subject group was smaller than the group of patients.

Nevertheless, the observed proportions of tasters and nontasters in the healthy group were consistent with other population-based investigations of PTC sensitivity (4).

We conclude that PTC nontaster status may be an endophenotypic marker of an inherited neuronal abnormality that conveys risk for the development of schizophrenia.

Received Feb. 22, 2004; revision received May 4 and May 17, 2004; accepted May 26, 2004. From the Schizophrenia Research Center, Department of Psychiatry, and the Smell and Taste Center, Department of Otorhinolaryngology, Head and Neck Surgery, University of Pennsylvania School of Medicine. Address correspondence and reprint requests to Dr. Moberg, Brain-Behavior Laboratory, Department of Psychiatry, 10th Floor, Gates Bldg., University of Pennsylvania School of Medicine, 3400 Spruce St., Philadelphia, PA 19104; moberg@bbl.med.upenn.edu (e-mail).

Funded in part by NIH grants MH-63381 (to Dr. Moberg), MH-59852 (to Dr. Turetsky), MH-66121, MH-42191 (to Dr. Gur), and an Independent Investigator Award from the National Alliance for Research on Schizophrenia and Depression to Dr. Moberg.

The authors thank Ms. Janice Lieber and Mrs. Constance Lieber for their support.

References

Objective: The authors’ goal was to investigate the distribution of metabolites and voxel composition in the pons and three cerebellar subregions and compare metabolite integral values and differences in voxel composition between patients with schizophrenia and healthy subjects.

Method: Proton magnetic resonance spectroscopic imaging was used to study the cerebellum and pons of 14 patients with schizophrenia and 14 healthy comparison subjects.

Results: The voxel composition was not significantly different between the groups, but the patients with schizophrenia had significantly lower N-acetylaspartate levels in the cerebellar cortex and vermis.

Conclusions: The lower integral value of N-acetylaspartate in the cerebellar cortex and the vermis of patients with schizophrenia supports the theory of a dysfunctional corticocerebellar-thalamic-cortical circuit in schizophrenia.

The literature on schizophrenia gives increasing evidence that the cerebellum is involved in complex mental activities and plays an important role within the cerebral network, revealing a functional disconnectivity in patients with schizophrenia in the corticocerebellar thalamic cortical circuit (1, 2). Proton magnetic resonance spectroscopic imaging (1H-MRSI) allows the detection of signals from N-acetylaspartate, creatine, and choline-containing compounds. Few 1H-MRSI studies have been conducted of the cerebellum in patients with schizophrenia (3–5).

Method

Fourteen patients with schizophrenia (12 men, two women; mean age=38.9 years, SD=7.3) and 14 healthy comparison sub-
Brief Report

Further Evidence for Altered Cerebellar Neuronal Integrity in Schizophrenia

Gabriele Ende, Dr. rer. nat.
Petra Hubrich, M.D.
Sigrid Walter, Dipl. Phys.
Wolfgang Weber-Fahr, Dr. rer. nat.
Nina Kämmerer, M.D.
Dieter F. Braus, M.D.
Fritz A. Henn, M.D., Ph.D.

Objective: The authors’ goal was to investigate the distribution of metabolites and voxel composition in the pons and three cerebellar subregions and compare metabolite integral values and differences in voxel composition between patients with schizophrenia and healthy subjects.

Method: Proton magnetic resonance spectroscopic imaging was used to study the cerebellum and pons of 14 patients with schizophrenia and 14 healthy comparison subjects.

Results: The voxel composition was not significantly different between the groups, but the patients with schizophrenia had significantly lower N-acetylaspartate levels in the cerebellar cortex and vermis.

Conclusions: The lower integral value of N-acetylaspartate in the cerebellar cortex and the vermis of patients with schizophrenia supports the theory of a dysfunctional corticocerebellar-thalamic-cortical circuit in schizophrenia.

The literature on schizophrenia gives increasing evidence that the cerebellum is involved in complex mental activities and plays an important role within the cerebral network, revealing a functional disconnectivity in patients with schizophrenia in the corticocerebellar thalamic cortical circuit (1, 2). Proton magnetic resonance spectroscopic imaging (1H-MRSI) allows the detection of signals from N-acetylaspartate, creatine, and choline-containing compounds. Few 1H-MRSI studies have been conducted of the cerebellum in patients with schizophrenia (3–5).

Method

Fourteen patients with schizophrenia (12 men, two women; mean age=38.9 years, SD=7.3) and 14 healthy comparison sub-
subjects (eight men, six women; mean age=35.6 years, SD=3.7) were studied with 1H-MRSI. All subjects were right-handed (Edinburgh Inventory [6]). All patients satisfied DSM-III-R as well as ICD-10 criteria for schizophrenia and had been diagnosed for at least 6 months. Their mean Brief Psychiatric Rating Scale score was 32.3 (SD=4.8). All patients had been clinically stable for at least 3 months.

After complete description of the study to the subjects, written informed consent was obtained. The study was approved by the university ethics committee.

All 1H-MRSI studies were performed on a 1.5-T Siemens Vision (Siemens, Erlangen, Germany). A 1H-MRSI sequence with point-resolved spectroscopy volume selection was used with the volume centered on the cerebellum and pons. Measurement parameters included a field of view of 210×210 mm, 15-mm slice thickness, TE=135 msec, and TR=1.5 seconds. In addition, a three-dimensional magnetization-prepared rapid-gradient-echo data set was acquired.

Postprocessing of the 1H-MRSI data included CSF and imperfect excitation pulse correction of 1H-MRSI brain data to raise sensitivity and lower variance (7).

Voxels were selected from the pons, vermis, dentate nucleus, and cerebellar cortex. In addition to the metabolite signals, the voxel composition for each subregion was evaluated for differences in gray matter, white matter, and CSF content. Mean values per data set of spectra from each subregion are reported.

Multivariate analysis based on a general linear model was used for data analysis by the use of SPSS for Windows, release 10.1 (SPSS, Inc., Chicago). Two multivariate models were used, one for each of the subregions (a and b) and one for the three integral values of the metabolites (c). The dependent variables were 1) the integral values for each individual metabolite in the subregions, with group as the between-subject factor and age and gender as covariates, 2) the voxel composition in terms of gray matter, white matter, and CSF for the subregions, with group as the between-subject factor and age and gender as covariates, and 3) the integral values for the metabolites (N-acetylaspartate, creatine, choline), with age, gender, and voxel gray matter content as covariates and group as the between-subject factor tested for each subregion. The criterion for significance was set at p<0.05.

**Results**

The analysis of the metabolite signals with age and gender as covariates in the subregions showed significant differences for the choline signals between regions (F=4.70, df=3, 19, p<0.02). There was no significant group effect and no significant influence of gender or age. The choline signals differed most between the vermis (largest signal) and cerebellar cortex (smallest signal).

The comparison among regions of mean voxel composition in percent gray matter, white matter, and CSF showed no significant group effect and no significant influence of gender or age but showed significant differences among the evaluated subregions (F=8.37, df=3, 19, p=0.001, for gray matter; F=5.73, df=3, 19, p<0.006, for white matter; F=3.79, df=3, 19, p<0.03, for CSF).

---

**FIGURE 1. Results for the CSF-Corrected N-Acetylaspartate Signal for Three Cerebellar Subregions and the Pons in Healthy Comparison Subjects and Patients With Schizophrenia**

<table>
<thead>
<tr>
<th></th>
<th>Pons</th>
<th>Cerebellar Cortex</th>
<th>Cerebellar Vermis</th>
<th>Dentate Nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison subjects</td>
<td>N=14</td>
<td>N=14</td>
<td>N=14</td>
<td>N=14</td>
</tr>
<tr>
<td>Schizophrenia patients</td>
<td>N=13</td>
<td>N=14</td>
<td>N=14</td>
<td>N=14</td>
</tr>
</tbody>
</table>

- Significantly lower values for patients (F=10.32, df=1, 23, p=0.004).
- Significantly lower values for patients (F=14.94, df=1, 23, p=0.001).
The analysis of the metabolite signals revealed significant effects of group in the cerebellar cortex (F=3.41, df=3, 21, p<0.04) and the vermis (F=5.57, df=3, 21, p=0.006) and a significant influence of gray matter content in the cerebellar cortex (F=6.60, df=3, 21, p=0.003).

Univariate analysis revealed significantly lower values for patients of N-acetylaspartate in both regions (F=10.32, df=1, 23, p=0.004, for the cerebellar cortex; F=14.94, df=1, 23, p=0.001, for the vermis) (Figure 1) and choline in the vermis (F=4.88, df=1, 23, p<0.04). Statistics on the remaining metabolite signal values in these two regions were above the p=0.05 level (F=2.88, df=1, 23, p=0.10, for choline in the cerebellar cortex; F=3.53, df=1, 23, p=0.07, for creatine in the cerebellar cortex; F=4.03, df=1, 23, p=0.06, for creatine in the vermis). No significant age or gender influence was found. The influence of the gray matter content was significant only for the creatine signal in the cerebellar cortex (F=4.22, df=1, 23, p=0.05).

There was no significant correlation between the metabolites and duration of illness (r<sub>c</sub>=0.18>r<sub>c</sub> >-0.36, p=0.2, Spearman’s correlation).

**Discussion**

The cerebellar metabolite reductions between our patients with schizophrenia and healthy comparison subjects cannot be attributed to differences in tissue voxel heterogeneity between the two groups. The significantly lower N-acetylaspartate signals in patients with schizophrenia were restricted to the cerebellar cortex and the vermis.

Our data are in good agreement with the 1H-MRSI data of Deicken et al. (5), who also found reduced N-acetylaspartate in the vermis of patients with schizophrenia. A lower level of N-acetylaspartate is generally thought to indicate reduced neuronal viability, resulting either from structural or functional neuronal impairment.

Marcelis et al. (8) reported evidence for cerebellar, frontal, and thalamic gray matter deficits in psychotic patients. This is consistent with our own data (9) and supports the theory of a dysfunctional corticocerebellar-thalamic-cortical circuit, as proposed by Andreasen et al. (1).

Our patient group was not homogeneous regarding illness duration and medication. All patients were receiving medication, and possible medication effects cannot be excluded. The influence of the covariate gray matter on univariate N-acetylaspartate comparisons was not significant, but the influence of gray matter differences on the main effect was not totally eliminated.

Presented in part at the 9th annual meeting of the Society of Magnetic Resonance in Medicine, Glasgow, April 21–27, 2001. Received March 1, 2004; revisions received May 3, May 10, and May 26, 2004; accepted June 10, 2004. From NMR Research in Psychiatry, Central Institute of Mental Health, Mannheim, Germany. Address reprint requests and correspondence to Dr. Ende, NMR Research in Psychiatry, Central Institute of Mental Health, J5, 68159 Mannheim, Germany; ende@z-i-mannheim.de (e-mail).

Supported by grant 47/2000 from the Central Institute of Mental Health.

The authors thank Dr. Norbert Schuff for providing the spherical k-space sampling magnetic resonance spectroscopic imaging sequences, Drs. Andrew Maudsley and Brian Soher for providing the automated spectral fitting routine, and Iris Reinhard for help with statistical analysis.

**References**

The Longitudinal Course of Thought Disorder in Geriatric Patients With Chronic Schizophrenia

Christopher R. Bowie, Ph.D.
Irene Tsapelas, B.A.
Joseph Friedman, M.D.
Michael Parrella, Ph.D.
Leonard White, Ph.D.
Philip D. Harvey, Ph.D.

Objective: This study longitudinally assessed previously validated dimensions of thought disorder—verbal underproductivity and disconnection—in geriatric schizophrenia and replicated previous cross-sectional differences in communication disorders.

Communication in schizophrenia is commonly characterized by several types of abnormalities, referred to as “formal thought disorder.” Positive thought disorder (also referred to as “disconnection” or “disorganization syndrome”) involves abnormalities in the production of language, such as derailment and tangentiality, whereas negative thought disorder denotes a reduction in the amount of overall output or information in speech, which includes poverty of content of speech and poverty of speech. Empirical studies of the structure of communication disorders have indicated that poverty of content of speech is actually associated with disconnection and not verbal productivity (1).

The longitudinal course of thought disorder is not well understood. Thought disorder was found to be stable during the early stages of an acute psychotic episode (2), and the severity of both poverty of speech and disconnected speech remained stable across an acute episode (3) and a subsequent 8-month follow-up (4). Addressing the long-term course of communication disorders, Harvey et al. (5) examined age-related differences in formal thought disorder in a cross-sectional study of schizophrenia patients who ranged in age from 19 to 96. Poverty of speech was found to be more common and severe in geriatric patients. However, the disconnection component was less severe in geriatric patients. Further analyses indicated that this lower severity of disconnection was not due to differences in verbal output. Cognitive impairment was found to predict poverty of speech. These findings tentatively suggest that disconnection and poverty of speech in schizophrenia, despite being generally stable over time, may change with age and possibly with level of cognitive functioning.

The present study is a longitudinal examination of the course of these two dimensions of thought disorder in older patients with schizophrenia. Based on previous findings, we hypothesized that

1. Poverty of speech will worsen over time.
2. Disconnection will remain stable or improve over time.
3. Age and cognitive impairment will predict the baseline severity and subsequent course of poverty of speech.

Method

As part of a longitudinal study of chronic schizophrenia described elsewhere (5, 6), we selected schizophrenia patients who had follow-up assessments on the measures described.

We identified 799 schizophrenia patients at baseline and 220 with follow-up assessments. Patients were lost to follow-up because of uncooperativeness of the patient (N=87), death of the patient (N=290), or discharge to a noncooperative treatment site (N=202). The mean follow-up interval was 842 days (SD=489, range=362–2,264). Among the patients with follow-up data, 127 (57.7%) were women; the group had a mean age of 74.9 (SD=9.4, range=49–97) at baseline, a mean of 10.2 years (SD=2.7) of education, and a mean age at first hospitalization of 27.6 (SD=10.9). One hundred forty-three were residents of a psychiatric hospital for chronic illness, and 57 were residents of a nursing home who had been discharged from a psychiatric hospital for chronic illness. Exactly 77.7% of the patients were Caucasian, 16.4% were African American, and 5.9% were Hispanic. All were English speaking, without sensory deficits, and cooperative with the assessment procedures.

In this study, we report on the results of the patients’ performance on the Mini-Mental State Examination (MMSE) (7) and the interviewers’ ratings of thought disorder on the Thought, Language, and Communication Scale (8, 9). The Thought, Language, and Communication Scale ratings were based on a symptom in-
A multivariate ANOVA to examine baseline differences in cognition, poverty of speech, disconnection, age, and age at first hospitalization between those who were reevaluated and those who were lost to follow-up was not significant (Pillai’s trace $F=1.75$, df=5, 550, $p=0.12$).

The patients displayed worsening in verbal underproductivity over time, with mean poverty of speech ratings of 1.9 (SD=1.4) at baseline and 2.3 (SD=1.4) at follow-up (Pillai’s trace $F=13.3$, df=1, 219, $p<0.001$, Wilks’s lambda=0.94). Only 5% of the patients displayed improvements in poverty of speech; the majority (74%) who improved did so by 1 point. Verbally productive patients demonstrated improved disconnection from baseline (mean=4.8, SD=4.7) to follow-up (mean=3.7, SD=4.4) (Pillai’s trace $F=4.9$, df=1, 124, $p=0.03$, Wilks’s lambda=0.96). However, when entering change in poverty of speech as a covariate, this finding was no longer significant (Pillai’s trace $F=0.64$, df=1, 124, $p=0.43$, Wilks’s lambda=0.99). MMSE scores were stable for the group as a whole from baseline (mean=11.4, SD=9.8) to follow-up (mean=12.1, SD=10.2) (Pillai’s trace $F=1.3$, df=1, 188, $p=0.26$, Wilks’s lambda=0.99).

Poverty of speech was associated with MMSE scores at baseline ($r=-0.60$, $p<0.001$) and follow-up ($r=-0.58$, $p<0.001$) and severity scores on disconnection at baseline ($r=-0.45$, $p<0.001$) and follow-up ($r=-0.35$, $p<0.001$). Worsening in poverty of speech was correlated with decreases in MMSE scores ($r=-0.30$, $p<0.03$) and greater age at baseline ($r=-0.15$, $p<0.03$) but not years of education, medication status (none, typical or atypical antipsychotic), or age at first hospitalization. Disconnection was not significantly associated with MMSE scores at baseline ($r=0.03$, $p=0.71$) or follow-up ($r=0.04$, $p=0.60$). Gender was not significantly associated with disconnection at baseline ($r=0.05$, $p=0.43$, Wilks’s lambda=0.99) and severity scores on disconnection at baseline ($r=0.03$, $p=0.60$, Wilks’s lambda=0.99). The repeated-measures ANOVA examining changes in disconnection by age was not significant (Pillai’s trace $F=2.9$, df=1, 180, $p=0.09$, Wilks’s lambda=0.98). As demonstrated in Figure 1, this interaction effect suggests that the older group had a greater worsening in verbal underproductivity over time. The repeated-measures ANOVA examining changes in MMSE by age was also not significant (Pillai’s trace $F=1.4$, df=1, 180, $p=0.25$, Wilks’s lambda=0.99).

Discussion

Older schizophrenia patients with a chronic course of illness show decreases in verbal productivity over time, with this change predicted by worsening in MMSE scores. Improvement in verbal productivity over time was rare and minimal. Scores on the disconnection scale indicated an improvement over time; however, this change is an artifact of decreases in output. It may be more accurate to refer to this “improvement” as a shift in the behavioral topography of communication disorders, from disconnected to unproductive speech, as patients age. Viewed this way, thought disorder does not appear to “burn out” in chronic schizophrenia.

Age-related differences in thought disorder were noted in the cross-sectional study (5), and the present longitudinal study confirms a worsening in poverty of speech over time in the truly geriatric patients. The reduction in verbal output over time complements previous work that has
suggested a decline in cognitive and functional skills during this age range (10), particularly in frontal “executive” functions (11), as poverty of speech is associated with dorsolateral prefrontal cortex dysfunction (12). Reduced verbal output also suggests that findings of improvement in positive and disorganized symptoms with aging in schizophrenia should be carefully considered in the context of possible reductions in verbal productivity. Decreased verbal output may affect one's ability to report hallucinations and delusions, as well as reducing the tendency to produce aberrant language. Future studies may address some of the limitations of this study, such as the influence of physical illnesses on thought disorder in late life.

References


8. Andreasen NC: Thought, language, and communication disorders, I: clinical assessment, definition of terms and evaluation of their reliability. Arch Gen Psychiatry 1979; 36:1315–1321

9. Andreasen NC: Thought, language, and communication disorders, II: diagnostic significance. Arch Gen Psychiatry 1979; 36; 1325–1330


Received March 1, 2004; revision received May 31, 2004; accepted June 10, 2004. From the Department of Psychiatry, Mount Sinai School of Medicine. Address correspondence and reprint requests to Dr. Bowie, Department of Psychiatry, Mount Sinai School of Medicine, 1425 Madison Ave, 4th Floor, Box 1230, New York, NY 10029; christopher.bowie@mssm.edu (e-mail).

Funded by the Mount Sinai Mental Health Clinical Research Center on Late-Life Schizophrenia (K.L. Davis, principal investigator), the Silvio Conte Neuroscience Center (NIMH grant MH-36692; K.L. Davis, principal investigator), NIMH grant 63116 (Dr. Harvey, principal investigator), and the VA Veterans Integrated Service Network 3 Mental Illness Research, Education, and Clinical Center.
Is Combination Olanzapine and Antidepressant Medication Associated With a More Rapid Response Trajectory Than Antidepressant Alone?

Gordon Parker, M.D., Ph.D., D.Sc., F.R.A.N.Z.C.P.
Heather Brotchie, B.A., M.B., B.S.
Kay Parker, B.A. (Hons. Psych.)

Objective: The authors’ goal was to determine if prescription of antidepressant medication plus olanzapine initiates a more rapid response than prescription of antidepressant alone.

Method: Twenty patients with major depression were studied. For 2 weeks the patients were blindly assigned to receive antidepressant plus olanzapine or antidepressant plus placebo. After 2 weeks, olanzapine augmentation was initiated for patients who did not improve with placebo augmentation. Response to medication was measured primarily by Hamilton Depression Rating Scale score. Other measures were the CORE, Clinical Global Impression, Beck Depression Inventory, and Daily Rating Schedule.

Results: Hamilton depression scores improved nonsignificantly in response to olanzapine combination therapy, but that trend was not evident on any secondary measure. Four patients who did not improve while receiving antidepressant and placebo showed rapid remission following late olanzapine augmentation.

Conclusions: Failure to demonstrate any benefit from initial combination therapy may reflect an underpowered rather than a negative study. The distinct impact of late olanzapine augmentation suggests that pretreatment with an antidepressant may be required to facilitate a rapid antidepressant response to combined treatment.

The addition of an augmenting atypical antipsychotic drug to an antidepressant drug is an increasingly observed clinical practice. Rapid remission has been described in relation to olanzapine (1) and risperidone (2) augmentation. In a controlled study (3), subjects with treatment-resistant depression received olanzapine alone, fluoxetine alone, or a combination of both; the combination was associated with significantly greater and faster improvement than was either drug alone.

Such data encouraged the current study, which tests the hypothesis that prescription of olanzapine with antidepressant medication is associated with a more rapid remission of depression than prescription of an antidepressant alone. Because there is evidence suggesting that, when antidepressants “work,” clinical improvement is present in the first 10 days (4), the controlled study component was restricted to 2 weeks.

Method

The study was conducted in an outpatient practice (6.P); the inclusion criterion was that the patient be experiencing a first or new episode of DSM-IV nonpsychotic major depression warranting prescription of antidepressant medication. Patients could not have received any antidepressant medication in the preceding 2 weeks or ECT in the previous month. If a patient had previously responded well to a particular antidepressant, that antidepressant was likely to be prescribed again (or the converse). If the patient had never received medication, dual-action antidepressants were favored for melancholic depression and selective serotonin reuptake inhibitors (SSRIs) were favored for nonmelancholic depression.

The active augmentor contained 2.5 mg of olanzapine, initially one tablet at night but allowed to be raised to two tablets at night if there was no significant improvement or side effects over the first week. At 2 weeks, the blind could be broken. If the patient was then in complete remission, the augmentor was ceased. If the patient showed minimal or no improvement to identified placebo augmentation, olanzapine augmentation could then be initiated. At baseline, depression severity was rated on the 17-item Hamilton Depression Rating Scale (5) (the primary outcome measure) and psychomotor severity was rated on the CORE (6). At weekly intervals, these measures were repeated and Clinical Global Impression (CGI) estimates of improvement (% change from baseline) undertaken. Patients completed the Beck Depression Inventory (7) at baseline and weekly as well as a Daily Rating Schedule (8) assessing depression severity nightly.

Of 24 subjects approached, three declined and one dropped out after 1 week. The remaining 20 subjects form the study group; 10 were assigned blindly (by means of random numbering) to receive antidepressant plus olanzapine, and 10 were assigned to receive antidepressant plus placebo. The antidepressants prescribed were extended release venlafaxine in the morning (N=10), citalopram in the morning (N=5), mirtazapine at night (N=4), and sertraline in the morning (N=1). Patients’ antidepressant medication regimens were maintained for 4 weeks without initiation of any other psychotropic medication. When the blind was broken at 2 weeks, the four recipients of placebo augmentation judged clinically to be nonimprovers received late olanzapine augmentation (at a dose of 2.5–5.0 mg).

Analyses were conducted on an intent-to-treat basis, and the last observation carried forward strategy was employed.

Results

Seven of the patients were men and 13 were women; their mean age was 49.5 years (SD=14.5). Respective 2-week and 4-week improvement rates were 50% and 73% (Hamilton depression scores improved nonsignificantly in response to olanzapine combination therapy, but that trend was not evident on any secondary measure. Four patients who did not improve while receiving antidepressant and placebo showed rapid remission following late olanzapine augmentation.

Conclusions: Failure to demonstrate any benefit from initial combination therapy may reflect an underpowered rather than a negative study. The distinct impact of late olanzapine augmentation suggests that pretreatment with an antidepressant may be required to facilitate a rapid antidepressant response to combined treatment.
depression scale), 49% and 67% (Beck Depression Inventory), 41% and 71% (CORE), 44% and 68% (total Daily Rating Schedule), and 44% and 73% (CGI improvement).

Table 1 reports baseline, day 7, and day 14 data for the patients receiving either olanzapine or placebo augmentation over the first 2 weeks. Over that fortnight, the former improved by 59.0% on the Hamilton depression scale (compared with 42.8% for those receiving placebo augmentation), a nonsignificant (t=2.0) difference. In addition, we defined primary responders as having at least 50% improvement in Hamilton depression scale scores over that fortnight. Of the 12 primary responders, eight were in the olanzapine and four in the placebo augmentation group, a nonsignificant (t=3.3, p=0.07) difference favoring olanzapine. However, because analyses failed to identify any significant impact of olanzapine augmentation on any other outcome study variable at that 2-week review, we analyzed percentage improvement on each Hamilton depression scale item to determine if the primary measure trends might merely be an artifact of olanzapine acting on selective constructs such as sleep. No such selectivity was evident: patients receiving olanzapine augmentation tended to show greater improvement on 13 of the 17 items, without any suggested superior impact on sleep.

For the four subjects receiving initial placebo and late olanzapine augmentation following minimal response at 2 weeks, improvement (from baseline) was distinctive by day 21 (i.e., 79.7% on the Hamilton depression scale, 86.4% on the Beck Depression Inventory, 82.1% on the total Daily Rating Schedule, and 83.3% improvement on CGI) and extended by day 28. Specifically, their mean weekly Hamilton depression scale scores were 25.2, 17.7, 18.0, 4.7, and 3.0, and their respective mean Beck Depression Inventory scores were 21.2, 17.7, 14.0, 3.0, and 1.7, with the trend breaks between weeks 2 and 3 following late olanzapine augmentation.

**Discussion**

To our knowledge, this is the first study examining whether initial prescription of an atypical antipsychotic and an antidepressant together is associated with a more rapid antidepressant effect. In the light of the small number of subjects, it remains a pilot study but should assist combination and augmentation therapy study designs. Improvement of some 50% over the first 2 weeks was so high as to risk reducing the probability of any true olanzapine augmentation benefit being revealed. The nonsignificant findings that patients receiving concomitant olanzapine and antidepressant were more likely to achieve responder status (i.e., 80% versus 40%) and have lower Hamilton depression scale scores at day 14 were not mirrored in the other observer-rated measures (CORE and CGI improvement) and are therefore unlikely to reflect rater bias. The findings may reflect nuances of the Hamilton depression scale detecting a true effect in an underpowered study.

Again conceding the small number of subjects, we highlight the suggestion of a distinct and rapid olanzapine augmenting effect in the patients who initially received antidepressant alone without benefit and, compatible with earlier observations (1–3), in treatment-resistant patients. Thus, any distinct olanzapine augmentation effect may depend on the antidepressant drug being in place—a mechanism postulated to explain lithium augmentation of antidepressant drugs in unipolar depression. De Montigny et al. (9), describing a case series of eight patients who, having failed to respond to a tricyclic, then responded rapidly to lithium introduction, suggested that tricyclic pretreatment sensitized the serotonin (5-HT) receptor to create an antidepressant effect, with lithium increasing the efficacy of

<table>
<thead>
<tr>
<th>Depression Variable</th>
<th>Score</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton Depression Rating Scale</td>
<td>20.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>25.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Total Daily Rating Schedule</td>
<td>37.7</td>
<td>0.6</td>
</tr>
<tr>
<td>CORE</td>
<td>10.4</td>
<td>0.5</td>
</tr>
<tr>
<td>CGI improvement (%)</td>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td>Hamilton Depression Rating Scale</td>
<td>11.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>15.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Total Daily Rating Schedule</td>
<td>20.9</td>
<td>0.0</td>
</tr>
<tr>
<td>CORE</td>
<td>6.2</td>
<td>1.4</td>
</tr>
<tr>
<td>CGI improvement (%)</td>
<td>43.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Hamilton Depression Rating Scale</td>
<td>8.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>14.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Total Daily Rating Schedule</td>
<td>22.6</td>
<td>0.8</td>
</tr>
<tr>
<td>CORE</td>
<td>6.3</td>
<td>0.2</td>
</tr>
<tr>
<td>CGI improvement (%)</td>
<td>43.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>
the central 5-HT system. Such a model may account for the substantive antidepressant responses reported when atypical antipsychotic drugs are added to antidepressant drugs in patients not responding to the antidepressant drug alone.

Such results argue for further refined studies examining the impact of both simultaneous prescription and prescription sequencing studies involving differing antidepressant classes, because the broader action of atypical antipsychotic drugs like olanzapine (with effects on noradrenergic, dopaminergic and serotonergic neurotransmission) may produce quite different results when combined with narrow-action SSRIs than when combined with broader-action antidepressants.

References


Received April 20, 2004; revision received June 17, 2004; accepted July 7, 2004. From the School of Psychiatry, University of New South Wales. Address reprint requests and correspondence to Professor Parker, Black Dog Institute, Prince of Wales Hospital, Randwick, Australia 2031; g.parker@unsw.edu.au (e-mail).

Supported by National Health and Medical Research Council Program Grant 2223208 and by an Investigator Initiated Grant from Eli Lilly, Australia.
Suicide Risk in Placebo-Controlled Trials of Treatment for Acute Manic Episode and Prevention of Manic-Depressive Episode

Jitschak G. Storosum, M.D., Ph.D.
Tamar Wohlfarth, Ph.D.
Christine C. Gispen-de Wied, M.D., Ph.D.
Don H. Linszen, M.D., Ph.D.
Berthold P.R. Gersons, M.D., Ph.D.
Barbara J. van Zwieten, Ph.D.
Wim van den Brink, M.D., Ph.D.

Objective: The authors’ goal was to investigate whether there is a greater suicide risk in the placebo arms of placebo-controlled studies of active medication for the treatment of acute manic episode and the prevention of manic/depressive episode. If so, this would be a strong ethical argument against the conduct of such studies.

Method: All placebo-controlled, double-blind, randomized trials of medication for the treatment of acute manic episode and the prevention of manic/depressive episode that were part of a registration dossier submitted to the regulatory authority of the Netherlands, the Medicines Evaluation Board, between 1997 and 2003, were reviewed for occurrence of suicide and attempted suicide.

Results: In 11 placebo-controlled studies of the treatment of acute manic episode, including 1,506 patients (117 person-years) in the combined active compound group and 1,005 patients (71 person-years) in the combined placebo group, no suicides and no suicide attempts occurred. In four placebo-controlled studies of the prevention of manic/depressive episode, including 943 patients (406 person-years) in the combined active compound group and 418 patients (136 person-years) in the placebo group, two suicides (493/100,000 person-years of exposure) and eight suicide attempts (1,969/100,000 person-years of exposure) occurred in the combined active compound group, but no suicides and two suicide attempts (1,467/100,000 person-years of exposure) occurred in the combined placebo group.

Conclusions: Concern about greater risk of suicide or attempted suicide in the placebo group should not be an argument against the conduct of placebo-controlled trials for these indications, provided that appropriate precautions are taken.

The use of control subjects given placebo in clinical trials is associated with ethical problems, especially in cases where effective treatment is available (1, 2) and where progressive diseases involve potential deterioration that is likely to be irreversible. In Europe, however, granting a license to study treatment of bipolar disorder, manic episode, and bipolar depression and the prevention of manic/depressive episode is conditioned on demonstration of efficacy in comparison with placebo (3).

Patients with bipolar disorder are at high risk for committing suicide; the estimated rate is 400/100,000 person-years of exposure, compared with the international general population average of 17/100,000 person-years of exposure (4, 5). The rate of suicide attempts is approximately 2,000/100,000 person-years of exposure in patients with bipolar disorder (6, 7).

Recently, Goodwin et al. (8) reported that anticonvulsant treatment was associated with higher rates of suicide and suicide attempts than was lithium among patients treated for bipolar disorder for varying amounts of time. This finding is consistent with others noting that lithium, the gold standard therapy in bipolar disorder, may lower suicide risk (5, 6, 9, 10).

In addition to all other controversial issues concerning the use of placebo, the high risk of suicide in bipolar disorder and the possible suicide protection of lithium are arguments against the conduct of studies with a placebo group in this disorder (11).

The aim of this study is to investigate whether the risk of suicide and attempted suicide is indeed greater in placebo groups in studies of the treatment of acute manic episode and the prevention of manic/depressive episode in order to determine if this would be an argument against the conduct of such studies.

Method

All double-blind, placebo-controlled, randomized trials conducted for the treatment of acute manic episode and for the prevention of manic/depressive episode (3) that were part of a registration dossier submitted to the Medicines Evaluation Board of the Netherlands between the years 1997 and 2003 were reviewed for committed and attempted suicides. (In 1997, after a long period of time, the first compound for the treatment of bipolar disorder was submitted to the Medicines Evaluation Board.) The duration of the acute manic episode studies varied between 3 and 12 weeks. Prevention studies were included only if they had a minimum duration of 52 weeks.

The Medicines Evaluation Board is the regulatory authority of the Netherlands. To obtain a marketing authorization, pharmaceutical companies are required to submit a dossier to the Medicines Evaluation Board that includes all clinical trials conducted for a drug under development.

Pharmaceutical companies are required to report in their registration dossiers all efficacy and safety results, including suicides and attempted suicides (12). These dossiers contain studies that

http://ajp.psychiatryonline.org
may have been published as well as studies that will never be published. The decision of whether to grant a market authorization to a given product is based on the assessment of the complete dossier.

The original studies submitted by the companies to the Medicines Evaluation Board dossiers were selected. All suicides and suicide attempts that occurred during the placebo-controlled phase (as defined in the individual study protocols) of these studies were considered cases. The a priori definition of suicide and suicide attempt in this investigation was the definition that was used in the studies submitted to the Medicines Evaluation Board. Because the dossiers submitted are confidential and are the property of the pharmaceutical companies, the dossiers were made anonymous.

Analyses were based on the intent-to-treat population, including all patients who were randomly assigned to active medication or placebo. Suicides and attempted suicides that occurred after the placebo-controlled phase were not included in the analyses. The incidence of suicide and suicide attempts was estimated for person-years at risk, and the statistical significance of differences between groups was assessed on the basis of a Poisson model.

**Results**

During the period under investigation, 11 placebo-controlled studies for treatment of acute manic episode and four studies for the prevention of manic/depressive episode were submitted to the Medicines Evaluation Board (Table 1).

A total of 2,511 patients were included in the 11 intervention studies for acute manic episode: 1,506 in the active compound groups and 1,005 patients in the placebo groups. These patients contributed in total 188 person-years of exposure (117 person-years of exposure in the active compound groups and 71 person-years of exposure in the placebo group). The duration of the studies varied from 21 days to 84 days. Six studies had a three-arm design. Lithium was the active comparator in four studies and haloperidol in two. All studies included patients who were hospitalized at the beginning of the study. DSM criteria were used for the diagnosis in 10 studies: DSM-III-R criteria in one and DSM-IV criteria in nine. Suicidal patients were excluded at baseline in seven studies (N=1,685); suicidal patients were not explicitly excluded in the other four (N=826). Exclusion of suicidal patients was done on the basis of the clinician’s evaluation at entry to the study in five studies and based on a rating of 3 or more on item 3 of the Hamilton Depression Rating Scale in the other two.

No suicide occurred while patients were receiving the active medication, the active comparator, or the placebo during the study period.

One suicide occurred 21 days after the completion of the study in a patient from an active treatment group who was using medication at the time of suicide. One fatality occurred in the placebo group: a patient discontinued the placebo treatment 6 days after the start of the study because of insufficient response and died 20 days later as a result of a motor vehicle accident. This accident was coded as injury and not suicide.

One attempted suicide occurred in the placebo group in a patient who decided to withdraw from the study after 2 days of treatment. This patient was treated for 3 weeks with active medication and made a suicide attempt 3 days later. These cases of suicide and attempted suicide that occurred after the placebo-controlled phase of the studies were not included in the analysis.

A total of 1,361 patients were included in the four studies of the prevention of manic/depressive episode: 943 in the active compound groups, contributing 406 person-years, and 418 patients in the placebo groups, contributing 136 person-years. The duration of the studies varied from 52 to 76 weeks. Three studies had a three-arm design with lithium as active comparator; in these studies 258 patients were treated with lithium, contributing 80 person-years. Two studies included inpatients and outpatients, and the other two studies included only outpatients. DSM-III-R criteria were used to establish the diagnosis in one study, and DSM-IV criteria were used in the other three studies. Suicidal patients were excluded at baseline in all studies. In two studies the clinician determined at entry of the study whether the patient was at serious suicide risk. In the other two studies suicide risk was based on a rating of 3 or more on item 3 of the Hamilton depression scale.

Two patients committed suicide in the long-term studies of prevention of manic/depressive episode. Both suicides occurred in active compound groups. The suicide rate in the combined active compound group was 493/100,000 person-years of exposure. Another suicide occurred in the active compound group 3 weeks after the patient dropped out of the study. This patient was not included in the analysis. No suicide occurred in the placebo groups; therefore, no statistical test was possible.


<table>
<thead>
<tr>
<th>Indicationa</th>
<th>Study Duration (weeks)</th>
<th>Patients Given Placebob</th>
<th>Patients Given Active Compoundb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Rate</td>
<td>N</td>
</tr>
<tr>
<td>Treatment of acute manic episode (N=11)</td>
<td>3–12</td>
<td>1,005 0 0 0 0</td>
<td>1,506 0 0 0 0</td>
</tr>
<tr>
<td>Prevention of manic/depressive episode (N=4)</td>
<td>&gt;52</td>
<td>418 0 0 2 1,467</td>
<td>943 2 493 8 1,969</td>
</tr>
</tbody>
</table>

*a N indicates number of studies.

b N indicates number of patients. Rate is per 100,000 person-years of exposure.
There was a total of 10 attempted suicides. Two occurred in the combined placebo group and eight in the combined active compound group. The incidence rate of suicide attempts per 100,000 person-years of exposure was 1,467 in the placebo group and 1,989 in the active compound group. The differences between the two incidence rates were not statistically significant (likelihood ratio=0.15, df=1, p<0.71, likelihood ratio test).

Two patients who were in the active compound group made a suicide attempt during follow-up, i.e., after discontinuation from the study. These patients were not included in the analysis.

No suicides occurred in the lithium treatment groups. Two suicide attempts were reported in the lithium treatment groups, resulting in a suicide attempt rate of 1,801/100,000 person-years of exposure in the lithium groups.

Discussion

The results presented here indicate no greater risk of suicide among patients with acute manic episode or stabilized bipolar disorder who were treated with placebo compared with the risk of patients who were treated with an active compound under the conditions of the trials. Moreover, the inclusion in the analysis of the suicides and attempted suicides that occurred after the completion of the studies does not change this result. These findings are compatible with findings from depression (13, 14) and schizophrenia (15, 16) trials.

Several limitations of this study should be considered when interpreting its results. The first is its limited power, due to the relatively low incidence of suicide and attempted suicide even in this high-risk population (1,467 suicide attempts per 100,000 person-years in the combined placebo groups). Yet it should be kept in mind that the evidence presented here is likely to be the best available because of the extraordinary large number of patients included in the pooled analysis (N=3,872).

A bias that may limit the generalizability of the results is caused by the exclusion in these studies of patients who were at risk for committing suicide (7, 17–19). However, in the recently reported study of Goodwin et al. (8) the incidence of suicide was 66/100,000 person-years of exposure in the lithium treated patients and 155/100,000 person-years of exposure in patients given anticonvulsants (mainly divalproex), whereas the suicide rate in the active treatment arms of the placebo-controlled prevention studies of our investigation was 493/100,000 person-years of exposure. This figure indicates that patients with bipolar disorder are at high risk for suicide, even if they are considered nonsuicidal and treated with an active compound.

The studies included in our investigation are part of registration files. Because some studies are not submitted to the Medicines Evaluation Board (e.g., because a company decided to stop further development of the drug) our investigation did not cover all studies conducted in the period 1997 to 2003. However, “negative” studies and new studies that have not (yet) been published were available in the database. Therefore, this information is not likely to be affected by publication bias and is adequate for addressing the current research question.

We believe that the evidence presented gives some indication that placebo treatment does not raise the risk of suicide among patients who are eligible to participate in these kinds of trials.

Am J Psychiatry 162:4, April 2005

BRIEF REPORTS

References


Received Dec. 3, 2003; revision received Feb. 26, 2004; accepted April 5, 2004. From the Medicines Evaluation Board of the Netherlands; and the Psychiatric Department of the Academic Medical Center, Amsterdam. Address correspondence and reprint requests to Dr. Storosum, Medicines Evaluation Board of the Netherlands, Kalvermarkt 53, PO-BOX 16229, 2500 BE Den Haag, The Netherlands; J.G.Storosum@CBG-MEB.nl (e-mail).
Brief Report

Platelet Serotonin Reuptake Inhibition and Response to SSRIs in Depressed Adolescents

David A. Axelson, M.D.
James M. Perel, Ph.D.
Boris Birmaher, M.D.
George Rudolph, B.S.
Sharon Nuss, R.N.
Linda Yurasits, B.S.
Jeffrey Bridge, Ph.D.
David A. Brent, M.D.

Objective: The authors examined platelet serotonin reuptake inhibition and response to selective serotonin reuptake inhibitor (SSRI) treatment in depressed adolescents.

Method: Twenty-three depressed adolescents participating in pharmacokinetic studies of SSRIs had platelet serotonin reuptake measured before and after 14–28 days of treatment. The Clinical Global Impression (CGI) improvement rating was determined on the basis of all clinical information and was performed blind to the platelet data.

Results: Improvement in depressive symptoms as rated with the CGI improvement subscale was significantly associated with the percentage change in platelet serotonin reuptake inhibition from pre- to posttreatment. Improvement in depression was also associated with absolute decrease in platelet serotonin reuptake when adjusted for the magnitude of baseline reuptake.

Conclusions: Platelet serotonin reuptake inhibition may be an appropriate surrogate biological marker for the pharmacodynamic activity of SSRIs in depressed adolescents.

Selective serotonin reuptake inhibitors (SSRIs) have been shown to be efficacious in children and adolescents for the treatment of depression (1–3). The presumed mechanism of action for the therapeutic effect of SSRIs is blockade of the serotonin reuptake transporter on presynaptic neurons, with the subsequent increase in transmission by serotonergic neurons. Human platelets have a serotonin reuptake transporter that is identical to the one in the brain (4), thus the platelet transporter may be used as a surrogate marker for the effect of SSRIs on serotonergic neurons.

Previous positron emission tomography data have indicated that adults taking 20 mg of paroxetine or 20 mg of citalopram had a 77% mean occupancy of the serotonin transporter in the brain (5). It has also been observed that at the typical minimum therapeutic doses of SSRIs in adults (e.g., 20 mg of fluoxetine or paroxetine, 50 mg of sertraline, 40 mg of citalopram), the mean serotonin reuptake inhibition in patient platelets was 60%–80% (6).

However, the extent of platelet serotonin reuptake inhibition has not been quantified during SSRI treatment in adolescents and has not been directly related to clinical response in this population or in adults.

This study determined the change in platelet serotonin reuptake in 23 adolescents with depressive symptoms who participated in a 14–28-day pharmacokinetic study of citalopram (20 mg/day), paroxetine (20 mg/day), or sertraline (50 mg/day). The relation of serotonin reuptake inhibition to clinical response was also examined.

Method

Subjects were participants in pediatric SSRI pharmacokinetic studies that were approved by the University of Pittsburgh Institutional Review Board (7). They were referred to the study by their attending child psychiatrist for initiation of SSRI treatment. The subjects’ parents or guardians as well as those subjects aged 14 years or older provided written informed consent prior to initiation of any study procedures. Verbal assent was obtained from
Brief Report

Platelet Serotonin Reuptake Inhibition and Response to SSRIs in Depressed Adolescents

David A. Axelson, M.D.
James M. Perel, Ph.D.
Boris Birmaher, M.D.
George Rudolph, B.S.
Sharon Nuss, R.N.
Linda Yurasits, B.S.
Jeffrey Bridge, Ph.D.
David A. Brent, M.D.

Objective: The authors examined platelet serotonin reuptake inhibition and response to selective serotonin reuptake inhibitor (SSRI) treatment in depressed adolescents.

Method: Twenty-three depressed adolescents participating in pharmacokinetic studies of SSRIs had platelet serotonin reuptake measured before and after 14–28 days of treatment. The Clinical Global Impression (CGI) improvement rating was determined on the basis of all clinical information and was performed blind to the platelet data.

Results: Improvement in depressive symptoms as rated with the CGI improvement subscale was significantly associated with the percentage change in platelet serotonin reuptake inhibition from pre- to posttreatment. Improvement in depression was also associated with absolute decrease in platelet serotonin reuptake when adjusted for the magnitude of baseline reuptake.

Conclusions: Platelet serotonin reuptake inhibition may be an appropriate surrogate biological marker for the pharmacodynamic activity of SSRIs in depressed adolescents.

Selective serotonin reuptake inhibitors (SSRIs) have been shown to be efficacious in children and adolescents for the treatment of depression (1–3). The presumed mechanism of action for the therapeutic effect of SSRIs is blockade of the serotonin reuptake transporter on presynaptic neurons, with the subsequent increase in transmission by serotonergic neurons. Human platelets have a serotonin reuptake transporter that is identical to the one in the brain (4), thus the platelet transporter may be used as a surrogate marker for the effect of SSRIs on serotonergic neurons.

Previous positron emission tomography data have indicated that adults taking 20 mg of paroxetine or 20 mg of citalopram had a 77% mean occupancy of the serotonin transporter in the brain (5). It has also been observed that at the typical minimum therapeutic doses of SSRIs in adults (e.g., 20 mg of fluoxetine or paroxetine, 50 mg of sertraline, 40 mg of citalopram), the mean serotonin reuptake inhibition in patient platelets was 60%–80% (6). However, the extent of platelet serotonin reuptake inhibition has not been quantified during SSRI treatment in adolescents and has not been directly related to clinical response in this population or in adults.

This study determined the change in platelet serotonin reuptake in 23 adolescents with depressive symptoms who participated in a 14–28-day pharmacokinetic study of citalopram (20 mg/day), paroxetine (20 mg/day), or sertraline (50 mg/day). The relation of serotonin reuptake inhibition to clinical response was also examined.

Method

Subjects were participants in pediatric SSRI pharmacokinetic studies that were approved by the University of Pittsburgh Institutional Review Board (7). They were referred to the study by their attending child psychiatrist for initiation of SSRI treatment. The subjects’ parents or guardians as well as those subjects aged 14 years or older provided written informed consent prior to initiation of any study procedures. Verbal assent was obtained from
children 13 years of age or younger. Subjects had a pretreatment interview that included the depression section of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS), present episode version (8). Subjects also completed the Beck Depression Inventory (9) or Children’s Depression Inventory (10) and had a clinical interview with a board-certified child psychiatrist (D.A.A.) to confirm the appropriateness of SSRI treatment. Subjects were included in this analysis if they met DSM-IV criteria for a current depressive diagnosis. Of the 23 subjects who were depressed at intake, 13 had major depressive disorder, two had dysthyemic disorder, and eight had depressive disorder not otherwise specified. There were 10 male and 13 female subjects, with ages ranging from 13.1 to 17.7 years (mean=15.1, SD=1.2). Seventeen subjects were white, five were African American, and one was Asian American. All subjects were physically healthy and had normal physical examination results at study entry.

On the morning before starting the first dose of the SSRI, subjects had blood drawn for analysis of pretreatment platelet serotonin reuptake inhibition. They then took a daily morning dose of citalopram, 20 mg (N=13); sertraline, 50 mg (N=8); or paroxetine, 20 mg (N=2). After 14–28 days (mean=16.2) of treatment, subjects returned in the morning to have blood drawn before taking their morning dose to measure posttreatment platelet serotonin reuptake. Repeat assessments with the K-SADS depression interview and either the Beck Depression Inventory or Children's Depression Inventory were performed at that time. Subjects returned to their referring psychiatrist for follow-up treatment. Subjects did not take other psychotropic medications for at least 2 weeks before and during the study. All had negative urine drug screen results before entering the protocol.

Clinical response of depressive symptoms during the 14–28-day interval between platelet reuptake measurements was rated retrospectively by the lead author (D.A.A.) using the Clinical Global Impressions (CGI) improvement scale. The ratings were based on the K-SADS depression interviews, Beck Depression Inventory/Children’s Depression Inventory self-reports, and clinical progress notes. The CGI was performed blind to the results of the platelet data.

The method of Tuomisto and colleagues (11) was used as the procedure for the platelet serotonin reuptake assay but was modified by reducing the blood volume collected to 20 ml. Platelet concentration was determined by a Coulter count, and samples were diluted to a platelet concentration of 200,000 cell/µl. Sero\-tonin uptake was measured at nine standard concentrations of [3H]serotonin within 2.5 hours of obtaining the sample from the subject. The interday coefficient of variation of the assay ranged from 9.1% to 14.6%. For the calculation of maximum velocity (Vmax) of serotonin uptake into the platelet, the raw data (from scintillation counter measurements at each of the nine concentrations) was converted into a V versus serotonin plot that was fit to the Michaelis-Menten equation by nonlinear regression using Enzfitter version 2.0.8 ( Biosoft, Ferguson, Mo.). Vmax was measured in picomoles of serotonin/107 platelets/5 minutes. Platelet serotonin reuptake inhibition was determined by the change in Vmax from pre- to posttreatment and was expressed as the percentage change from the baseline Vmax for the primary analysis. SPSS version 11.5 was used for the nonparametric correlation analysis and for an ordinal regression analysis using a logit link function (SPSS, Inc., Chicago).

Results

Two subjects were rated as being very much improved (CGI rating of 1), six were much improved (CGI rating of 2), 10 were minimally improved (CGI rating of 3), and five were rated as not improved (CGI rating of 4); therefore, 35% (N=8) of the 23 subjects were unequivocal responders. The mean platelet reuptake inhibition was 52% (SD=27%, range=0%–87%). There were no associations between demographic variables and either clinical response or platelet serotonin reuptake inhibition. There was a significant correlation between CGI improvement ratings and platelet reuptake inhibition, with greater clinical response associated with higher inhibition (Figure 1). Ordinal regression analysis determined that improvement on the CGI improvement subscale was significantly associated with absolute change in Vmax when baseline Vmax was controlled (beta estimate=0.31; Wald statistic=6.2, df=1, p=0.01).

Conclusions

To our knowledge, this is the first study in adolescents that quantifies the pharmacodynamic effect of SSRIs on a biological marker and shows a relation of the marker to clinical response. This is a preliminary study and should be viewed with the following limitations in mind. The study sample was small and was heterogeneous in the intensity of depressive symptoms at intake. Final clinical response was determined retrospectively, although it was based on data that was obtained prospectively and conducted blind to the results of the platelet assays. The duration of SSRI treatment prior to measuring response was relatively short, so some subjects may not have had time to exhibit a complete response to the medication. Consequently, a substantial proportion of subjects were rated as having minimal improvement, and platelet results were widely distributed in this group. However, the effect size of the difference in platelet serotonin reuptake inhibition be-

---

**FIGURE 1. Platelet Serotonin Reuptake Inhibition in 23 Depressed Adolescents Treated With SSRIs, by Posttreatment CGI Improvement Rating**

<table>
<thead>
<tr>
<th>CGI Improvement Rating</th>
<th>Platelet 5-HT Reuptake Inhibition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very much improved (N=2)</td>
<td>0%</td>
</tr>
<tr>
<td>Much improved (N=6)</td>
<td>10%</td>
</tr>
<tr>
<td>Minimally improved (N=10)</td>
<td>50%</td>
</tr>
<tr>
<td>Not improved (N=5)</td>
<td>90%</td>
</tr>
</tbody>
</table>

<sup>3</sup> Significant correlation between platelet serotonin reuptake inhibition and CGI improvement rating (r=−0.44, p<0.04).
tween subjects who showed unequivocal clinical improvement (CGI rating of 1 or 2; mean inhibition=66% [SD=17]) and those who had absolutely no response (CGI rating of 4; mean inhibition=33% [SD=33]) was large (d=1.4). Despite the aforementioned limitations, the results demonstrate that platelet serotonin reuptake inhibition may be an appropriate surrogate biological marker of the pharmacodynamic activity of SSRIs.

The development of appropriate surrogate biomarkers of treatment response to SSRIs in youth can have potential use clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.

The efficacy of SSRIs for depression in the pediatric population has recently been called into question. A recently issued FDA advisory stated that despite submissions of randomized controlled trials of sertraline, paroxetine, citalopram, venlafaxine (a serotonin-norepinephrine reuptake inhibitor), and fluoxetine for pediatric depression, only fluoxetine had sufficient evidence to be accepted clinically as well as have implications for whether this class of medication has therapeutic effects in this patient population. SSRI blood levels have not been shown to correlate with clinical response in numerous studies of depression in adults (12). As it can take several weeks at the proper dose of an SSRI to show improvement in depressive symptoms, a biological surrogate marker of response that can be measured early in treatment has the potential to substantially reduce titration time to the proper dose.
Effects of Tryptophan Depletion on Serum Levels of Brain-Derived Neurotrophic Factor in Unmedicated Patients With Remitted Depression and Healthy Subjects

Alexander Neumeister, M.D.
Peixiong Yuan, M.D., Ph.D.
Theresa A. Young, M.S.
Omer Bonne, M.D.
David A. Luckenbaugh, M.A.
Dennis S. Charney, M.D.
Husseini Manji, M.D.

Objective: Data suggest the involvement of serotonergic and neurotrophic systems in major depressive disorder. To investigate their potential interaction, the authors studied changes in serum levels of brain-derived neurotrophic factor (BDNF) during tryptophan depletion and sham depletion in unmedicated patients with remitted major depressive disorder and in a group of healthy comparison subjects.

Method: Twenty-seven patients with remitted major depressive disorder and 20 healthy subjects underwent tryptophan depletion and sham depletion in a randomized, placebo-controlled, double-blind crossover study. Serum BDNF concentrations and plasma tryptophan concentrations as well as behavioral assessments were obtained.

Results: During tryptophan depletion, BDNF levels increased in healthy volunteers. By contrast, patients with remitted major depressive disorder were unable to mount this presumed compensatory response, and BDNF levels remained low in these patients.

Conclusions: The results further substantiate the potential role of BDNF in major depressive disorder.

Recent studies have raised the possibility that the serotonergic and neurotrophic hypotheses of major depressive disorder are complementary rather than contradictory. Most pertinent for the present study are reports about lower levels of brain-derived neurotrophic factor (BDNF) in depressed patients (1, 2). These observations are noteworthy, since it has been shown that endogenous BDNF is critical for the normal development and function of central serotonin neurons as well as for the elaboration of behaviors that are mediated by brain serotonergic systems (3).

We examined the effects of tryptophan depletion and sham depletion on serum BDNF levels in unmedicated patients with remitted major depressive disorder and healthy subjects. We expected that during tryptophan depletion, serum BDNF levels would be lowered in patients with remitted major depressive disorder relative to healthy subjects and also relative to the sham depletion condition.

Method

Subjects were 27 unmedicated patients with remitted major depressive disorder (18 women and nine men; mean age=39.8 years [SD=12.7]; mean baseline Hamilton depression scale score=1.1 [SD=1.2]; mean age at onset=23.8 [SD=8.4]; mean number of previous episodes=3.6 [SD=2.6]) and 20 healthy comparison subjects (11 women and nine men; mean age=33.7 [SD=12.8]; mean baseline Hamilton depression scale score=0.7 [SD=0.8]); all were nonsmokers. Diagnosis was established according to the Structured Clinical Interview for the DSM-IV, nonpatient version. No additional lifetime diagnosis other than major depressive disorder was allowed. Subjects were medically healthy. They were entered into the study after full explanation of the purpose of the study and the study procedures, and after written consent had been obtained as approved by the NIMH Institutional Review Board. The depressed patients had been in remission (Hamilton depression scale total score <8) and had not been taking antidepressant medication for a mean of 40.4 months (SD=48.4) at the time they entered the randomized, placebo-controlled, double-blind crossover study.

Tryptophan depletion was induced by administration of capsules containing 32 g of an amino acid mixture without tryptophan; during sham depletion subjects received identical capsules containing a total of 32 g lactose. Test days were separated by at least 6 days. Serum BDNF levels and plasma tryptophan levels were obtained at baseline and 5, 7, and 24 hours after administration of the capsules on each test day. Behavioral assessments with the Hamilton depression scale were obtained at baseline and 7 and 24 hours after capsule administration.

Analysis of BDNF concentration in the plasma was performed with a ChemiKine BDNF Sandwich ELISA Kit (Chemicon, Temecula, Calif.) according to the manufacturer’s instructions with minor modifications. The samples and standards were applied in duplicate into 96-well immunoplates precoated with rabbit anti-human BDNF antibody and incubated on a shaker overnight at 4°C. After washing, biotinylated mouse anti-BDNF antibody was added and incubated overnight at 4°C. Then streptavidin-HRP was added and incubated at room temperature for 1 hour after washing. TMB substrate was added and incubated at room temperature for 15 minutes. The plate was read with a Multilabel Counter (Wallac, Finland) at 450 nm after added stop solution. The standard curve was linear from 7.8 to 500 pg/ml BDNF. The assay sensitivity is 7.8 pg/ml. The intraassay and interassay variability is within 10%.

For tryptophan level assessments, plasma was deproteinized for total tryptophan or filtered for free tryptophan measurements before it was subjected to an isocratic reversed-phase high-per-
BDNF data and plasma tryptophan concentrations were analyzed by using repeated measures ANOVA in which time and treatment were the repeated factors and group was the between-subject factor. Significant results were further examined by using Bonferroni-corrected simple effects tests. Results are presented as means and standard deviations.

Results

In patients with remitted major depressive disorder, tryptophan depletion but not sham depletion resulted in a transient lowering of BDNF levels at the 5-hour time point. In contrast, tryptophan depletion induced a significant increase in serum BDNF concentrations in healthy subjects. During sham depletion, BDNF concentrations significantly increased in the group with remitted major depressive disorder, but the increase was not significant for the comparison subjects. The treatment-by-time-by-group interaction was significant (F=5.79, df=2.87, 129.13, p<0.001). The time course of changes is described in Figure 1. No significant correlations were found between Hamilton depression scale scores and BDNF levels at any time point. Changes in BDNF levels did not differ between the 16 remitted major depressive disorder patients who showed a transient return of depressive symptoms during tryptophan depletion (defined as an increase in Hamilton depression scale total score to greater than 10) and the nine patients who remained well. As expected, tryptophan depletion lowered plasma total and free tryptophan levels (treatment-by-time interaction: F=166.75, df=2.25, 96.85, p<0.001, and F=79.99, df=2.79, 125.82, p<0.001) with no between-group differences and nadir values at the 5- and 7-hour time points.

Discussion

During tryptophan depletion, BDNF levels increased in healthy volunteers. By contrast, patients with remitted major depressive disorder did not show a similar response, and BDNF levels remained low in these patients. The transient increase in BDNF concentrations in healthy subjects during tryptophan depletion suggests a compensatory response to maintain the complex interactions between serotonergic and neurotrophic systems. It is interesting that this response appears to be dysfunctional in patients with remitted major depressive disorder, who are unable to mount this compensatory response.

We have previously postulated that antidepressants bring about their beneficial effects via both trophic and neurochemical support (4); the observations of the present study suggest that the two systems may indeed be intimately linked. Unexpectedly, increases in plasma BDNF were observed during sham depletion in patients with remitted major depressive disorder. This suggests the existence of a complex dysregulation between the systems that may be critical to the vulnerability of depression.

Previous studies have suggested that plasma BDNF may derive from both central and peripheral sources. However, findings that BDNF in the periphery crosses the blood-brain barrier by a high-capacity, saturable transport system (5), and evidence that serum BDNF levels do not differ from CSF levels, suggest that peripheral changes may also reflect central processes (6). The time course of changes is particularly relevant because acute changes in plasma...
BDNF levels in response to tryptophan depletion were measured. In this context, the onset of changes in BDNF mRNA expression in the brain increased as early as 15 minutes after stress exposure.

Baseline BDNF levels did not differ in subjects with remitted major depressive disorder versus healthy subjects. Recent studies have reported lower BDNF levels in depressed patients during a spontaneous episode of major depressive disorder. Our results suggest that it is a dysregulation of BDNF homeostasis in the face of a serotonergic perturbation that truly represents a trait vulnerability marker for depression.

References

Received Jan. 23, 2004; revisions received March 8, April 6, and April 22, 2004; accepted May 26, 2004. From the Section on Experimental Therapeutics and Pathophysiology and the Laboratory of Molecular Pathophysiology, NIMH Mood and Anxiety Disorders Program, Bethesda, Md. Address correspondence and reprint requests to Dr. Neumeister, Associate Professor of Psychiatry, Yale University School of Medicine, Molecular Imaging Program of the Clinical Neuroscience Division, 950 Campbell Ave., West Haven, CT 06516; alexander.neumeister@yale.edu (e-mail).

Dr. Charney and Dr. Manji contributed equally to this paper.

The authors thank the intramural program at NIMH and the 4 West nursing staff.

Enhanced Early Morning Salivary Cortisol in Neuroticism

Maria J. Portella, M.Sc.
Catherine J. Harmer, D.Phil.
Jonathan Flint, D.Phil.
Philip Cowen, M.D., F.R.C.Psych.
Guy M. Goodwin, D.Phil., F.R.C.Psych.

Objective: Neuroticism is a predisposing factor for major depression. The increase in salivary cortisol that follows waking provides a reliable measure of adrenocortical activity, and this response is increased in recovered depressed patients. This study compared waking cortisol levels in healthy subjects with high and low levels of neuroticism without a previous history of depression.

Method: Salivary cortisol levels were measured upon waking and at 15-minute intervals for the next hour in volunteers selected to have high (>19/23) or low (<4/23) neuroticism based on the Eysenck Personality Inventory.

Results: Subjects with high neuroticism showed significantly greater levels of salivary cortisol 30 minutes after waking, which were maintained for the next half hour.

Conclusions: Abnormalities in waking cortisol are associated with neuroticism in a way similar to those seen in major depression. Elevated waking cortisol may represent a vulnerability marker for mood disorder.

Neuroticism is a dimensional measure of an individual’s tendency to experience negative emotions that are manifested at one extreme as anxiety, depression, and moodiness and at the other as emotional stability. The heritability of neuroticism is well established (1), and some studies have suggested that there is a substantial overlap between the genetic risk factors for neuroticism and major depression (2). However, it is not known which biological correlates of neuroticism might give rise to the increased risk of clinical depression.

Major depression is often accompanied by hypothalamic-pituitary-adrenal (HPA) axis dysfunction (3). Studies have suggested that the increase in salivary cortisol that follows waking provides a reliable dynamic measure
BDNF levels in response to tryptophan depletion were measured. In this context, the onset of changes in BDNF mRNA expression in the brain increased as early as 15 minutes after stress exposure.

Baseline BDNF levels did not differ in subjects with remitted major depressive disorder versus healthy subjects. Recent studies have reported lower BDNF levels in depressed patients during a spontaneous episode of major depressive disorder. Our results suggest that it is a dysregulation of BDNF homeostasis in the face of a serotonergic perturbation that truly represents a trait vulnerability marker for depression.

Received Jan. 23, 2004; revisions received March 8, April 6, and April 22, 2004; accepted May 26, 2004. From the Section on Experimental Therapeutics and Pathophysiology and the Laboratory of Molecular Pathophysiology, NIMH Mood and Anxiety Disorders Program, Bethesda, Md. Address correspondence and reprint requests to Dr. Neumeister, Associate Professor of Psychiatry, Yale University School of Medicine, Molecular Imaging Program of the Clinical Neuroscience Division, 950 Campbell Ave., West Haven, CT 06516; alexander.neumeister@yale.edu (e-mail).

Dr. Charney and Dr. Manji contributed equally to this paper.

The authors thank the intramural program at NIMH and the 4 West nursing staff.

References

Enhanced Early Morning Salivary Cortisol in Neuroticism

Maria J. Portella, M.Sc.
Catherine J. Harmer, D.Phil.
Jonathan Flint, D.Phil.
Philip Cowen, M.D., F.R.C.Psych.
Guy M. Goodwin, D.Phil., F.R.C.Psych.

Objective: Neuroticism is a predisposing factor for major depression. The increase in salivary cortisol that follows waking provides a reliable measure of adrenocortical activity, and this response is increased in recovered depressed patients. This study compared waking cortisol levels in healthy subjects with high and low levels of neuroticism without a previous history of depression.

Method: Salivary cortisol levels were measured upon waking and at 15-minute intervals for the next hour in volunteers selected to have high (>19/23) or low (<4/23) neuroticism based on the Eysenck Personality Inventory.

Results: Subjects with high neuroticism showed significantly greater levels of salivary cortisol 30 minutes after waking, which were maintained for the next half hour.

Conclusions: Abnormalities in waking cortisol are associated with neuroticism in a way similar to those seen in major depression. Elevated waking cortisol may represent a vulnerability marker for mood disorder.

Neuroticism is a dimensional measure of an individual’s tendency to experience negative emotions that are manifested at one extreme as anxiety, depression, and moodiness and at the other as emotional stability. The heritability of neuroticism is well established (1), and some studies have suggested that there is a substantial overlap between the genetic risk factors for neuroticism and major depression (2). However, it is not known which biological correlates of neuroticism might give rise to the increased risk of clinical depression.

Major depression is often accompanied by hypothalamic-pituitary-adrenal (HPA) axis dysfunction (3). Studies have suggested that the increase in salivary cortisol that follows waking provides a reliable dynamic measure
of adrenocortical activity (4) and that this response is increased in euthymic unmedicated patients with a history of depression (5). However, it is unknown whether elevated waking cortisol is a risk factor for depression or a result of having been depressed or treated for depression in the past. The aim of this study was to examine waking cortisol in people with high or low levels of neuroticism without a history of mood disorder to test the hypothesis that high neuroticism itself is associated with altered adrenocortical regulation.

**Method**

Thirty healthy volunteers (ages 21–57 years) were selected by their neuroticism score on the Eysenck Personality Inventory. Individuals were chosen from a cohort of 20,427 families collected as part of an investigation into the genetic basis of personality (6). We contacted unrelated individuals from selected extremes: 15 subjects (seven men, age: mean=39.5, SD=14.6) with extremely high scores for neuroticism (mean=21.3, SD=1.3, range=19–23) and 15 subjects (seven men, age: mean=46.9, SD=9.0) with extremely low scores (mean=6.53, SD=1.13) were included in the analysis.

On the Structured Clinical Interview for DSM-IV, subjects were determined to be free of past or current history of axis I disorder. They had no current physical illness and had been free of medication for at least 1 month. The group with high neuroticism scores had a mean score on the Beck Depression Inventory of 9.14 (SD=4.73), and the group with low neuroticism scores had a mean score of 3.07 (SD=2.34). The study was approved by the local ethics committee. All subjects gave written informed consent.

Fasting saliva samples were collected in salivette tubes with the first sample taken immediately upon waking and continuing at 15-minute intervals for the next hour (7). After this, the subjects resumed normal activity. Subsequently, saliva samples were taken at noon, 6:00 p.m., and 10:00 p.m., with subjects avoiding food for 1 hour previously. The premenstrual week was avoided for women. Salivary cortisol was measured with an in-house double-antibody radioimmunoassay.

Salivary levels were analyzed with a two-way repeated-measures analysis of variance (ANOVA), with group and gender as-between-subject factors and time (sampling time) as the main within-subject factor and Huynh-Feldt correction (uncorrected dfs reported). Significant interactions were analyzed with post hoc t tests. The area under the curve for the first 60 minutes after waking was measured with the trapezoid method, with subtraction of baseline cortisol secretion.

**Results**

There was no difference in age between the groups (t=−1.67, df=28, p=0.10). The mean time of awakening did not differ between subjects with high neuroticism scores (7:03 a.m. ±41 minutes) and the subjects with low neuroticism scores (6:43 a.m. ±49 minutes) (t=1.15, df=28, p=0.26).

The ANOVA showed a significant group-by-time interaction (F=4.18, df=4, 104, p=0.005). Post hoc testing (Figure 1) showed a greater increase in salivary cortisol 30 minutes after waking in the group with high neuroticism scores, and this difference from subjects with low neuroticism scores was maintained for the next 30 minutes. After that, levels of cortisol were similar for both groups. There was no effect of gender (F=2.47, df=1, 26, p=0.13) or gender by time (F=0.88, df=4, 104, p=0.48).

The area under the curve for cortisol secretion was substantially greater in the group with high neuroticism scores (mean=14.92 nmol×hour/liter, SD=11.38, versus mean=5.79 nmol×hour/liter, SD=9.46) (t=2.39, df=28, p=0.02). In the subjects with high neuroticism scores, there was a correlation between the area under the curve and neuroticism score (r=0.53, p=0.04); however, Beck Depression Inventory score did not correlate with cortisol area under the curve in the volunteers with high neuroticism scores (r=0.26, p=0.37).

**Discussion**

The subjects with high neuroticism scores showed significantly greater levels of salivary cortisol between 30 and 60 minutes after waking. It is well established that waking in the morning is followed by an adrenocortical response with brief ACTH and cortisol pulses in the majority of the subjects. The greater waking increase in free cortisol found in this study is similar to that reported in recovered depressed patients (5). Also, we have recently identified a similar abnormality in unmedicated patients with acute major depression (Bhagwagar et al., unpublished). The salivary cortisol response to waking represents only a single aspect of HPA axis function; thus, we cannot conclude that neuroticism is associated with generally increased HPA axis activity. However, our observations are consistent with a greater ACTH response to waking or increased sensitivity of the adrenal gland to ACTH stimulation in subjects with high neuroticism scores.

Our group had no past or current history of depression, suggesting that increased waking cortisol is a risk factor for, rather than a consequence of, depression. It is possible that this abnormal response is inherited as part of the trait of neuroticism and, like high neuroticism scores, waking

**FIGURE 1. Mean Diurnal Variation in Salivary Cortisol Levels in Subjects With High or Low Neuroticism Scores**

![Graph showing mean diurnal variation in salivary cortisol levels](http://ajp.psychiatryonline.org)
salivary cortisol levels show significant heritability (8). Another possibility, however, is that the elevated waking cortisol levels might be a consequence of experiencing more subjective stress. Nonetheless, our data suggest that elevated morning cortisol levels can exist in the absence of major depression. In this respect, our findings could resemble the abnormal HPA axis activity seen in the first-degree relatives of depressed patients who have not experienced depression (9). Longitudinal prospective studies initiated before the onset of depression are needed to untangle the relationship between neuroticism, cortisol hypersecretion, and depression.

Received Jan. 15, 2004; revisions received March 18 and April 20, 2004; accepted May 14, 2004. From the Department of Psychiatry, Warneford Hospital, University of Oxford. Address correspondence and reprint requests to Dr. Harmer, Department of Psychiatry, Warneford Hospital, Neurosciences Bldg., Oxford OX3 7JX, U.K.; catherine.harmer@psych.ox.ac.uk (e-mail).

Funded by the Medical Research Council and Wellcome Trust.

Ms. Portella was supported by the Departament d’Universitats, Recerca i Societat de la Informació-DURSI (Generalitat de Catalunya), through grant 2002FI 00756.

The authors thank Alison Reed for technical assistance.

References

Letters to the Editor

Neurotoxicity Associated With Free Valproic Acid

To the Editor: Effective treatment of bipolar disorder with valproic acid is associated with serum levels of 45–125 µg/ml. Lower levels are associated with reduced clinical response, and higher levels with adverse events (1). Toxicity above 100 µg/ml is related to saturation of protein binding sites and nonlinear increases in free valproate (2). Although increased unbound valproic acid has been associated with toxicity in hypoalbuminemic patients (3), there are no reports, to our knowledge, of toxicity associated with free valproate in psychiatric patients. We report the case of a patient who experienced toxicity with increased unbound valproic acid and normal serum valproate and albumin.

Mr. A was a 49-year-old single Caucasian man who was admitted to a state hospital with a diagnosis of schizoaffective disorder, manic. His other disorders included type II diabetes mellitus and hyperlipidemia that were treated with metformin, 850 mg b.i.d., atorvastatin, 20 mg at bedtime, and regular insulin coverage. Vivid delusions and hallucinations were refractory to treatment with multiple antipsychotic agents. Agitation, irritability, and mood swings led to the addition of divalproex sodium. A regimen of 56 mg/day of perphenazine, 2000 mg/day of extended-release divalproex sodium, 6 mg/day of lorazepam, 50 mg/day of sertraline, and 2 mg/day of benzotropine did not lead to symptom abatement or, initially, altered cognition or sensorium. After 2 months, the hospital staff noted an ataxic gait, confusion, and slurred speech. Reducing Mr. A’s perphenazine dose yielded no benefit. Tapering his lorazepam and benzotropine led to modest sensorium clearing; however, after 5–6 days, the staff noted increasing confusion, disorientation, and ataxia. His serum valproic acid level was 86.1 µg/ml (normal range=50–125), his albumin level was 3.8 g/dl (normal range=3.2–5.2), and his liver function tests, ammonia levels, and platelet count were within normal limits. His free valproic acid was found to be 15 µg/ml (normal range=5–10), so his dose of extended-release divalproex sodium was reduced to 1000 mg/day, and his sensorium returned to normal in approximately 3 days.

“Normal” serum valproic acid ranges anticipate typical protein binding; however, atypical saturation kinetics could result in increased free valproic acid disparate from that suggested by serum total concentrations (3, 4). No specific factors appeared to account for the patient’s unexpectedly high free valproic acid level. His medical conditions and medications are not known to influence valproate binding. Our tapering of lorazepam (the levels of which may increase when valproate is added) did not resolve his cognitive impairments, which did respond to reduced valproate. Measurement of free valproate should be considered in patients with unexplained altered cognition, even when protein levels are apparently normal.

References

David I. Mayerhoff, M.D.
Jeffry Nurenberg, M.D.
Snehal Shah, R.P.
Steven J. Schleifer, M.D.
Greystone Park, N.J.

Multiple Endocrine Neoplasia Type 1 Presenting as Psychosis

To the Editor: The syndrome of multiple endocrine neoplasia type 1 is an autosomal-dominant inherited disease affecting several endocrine organs (1). The affected organs include the pituitary gland, the parathyroid glands, and the endocrine pancreas. The multiple endocrine neoplasia type 1 gene is a tumor-suppressor gene located on chromosome 11q. The psychiatric symptoms associated with multiple endocrine neoplasia type 1 have not yet been reported. We report the first case of our knowledge of a patient with multiple endocrine neoplasia type 1 presenting as psychosis.

Ms. A was a 59-year-old Asian woman who had been hospitalized for approximately 3 years. Her first psychotic episode, characterized by auditory and visual hallucinations, delusions, and catatonia, occurred at age 44. She had been hospitalized at ages 47, 50, 53, and 54 for similar psychotic episodes, each persisting for about 3 months. At her most recent hospitalization, her psychiatric symptoms were not alleviated by any antipsychotics and mood stabilizers; only lithium was slightly effective. Her serum levels of gastrin, insulin, and glucagon were normal. A subsequent partial pancreatectomy did not relieve her psychiatric symptoms. Three months after the operation, a high level of serum parathyroid hormone was found during an examination for osteoporosis. These two extraordinary endocrine findings led us to suspect multiple endocrine neoplasia type 1. Magnetic resonance imaging and ultrasonography revealed a microadenoma of the anterior pituitary gland and a swelling of the parathyroid gland. The diagnosis was confirmed by detection of a deficient of the heterozygous 357del4 multiple endocrine neoplasia type 1 gene on exon 2 chromosome 11q (2). The serum concentrations of prolactin, growth hormone, and adrenocorticotropic hormone were normal. The microadenoma of the anterior pituitary gland was nonfunctioning and did not need treatment in particular. The parathyroid swelling necessitated surgical treatment. A high level of serum parathyroid hormone recovered within normal limits after the partial parathyroidectomy; however, the psychiatric symptoms did not improve.

We investigated the profile of hereditary disposition of Ms. A’s family. Two of her three descendants could be ex-
manifestations of our patient could be linked with multiple endocrine neoplasia type 1. We concluded that the psychiatric symptoms differed from those typical of schizophrenia. These clinical features seemed to be one of the psychiatric manifestations of multiple endocrine neoplasia type 1. We concluded that the psychiatric manifestations of our patient could be linked with multiple endocrine neoplasia type 1.

**References**


**SHINSUKE KITO, M.D.**

**TOKU NAKAJIMA, M.D., Ph.D.**

**HIROSHI YAMADERA, M.D., Ph.D.**

**YOSHIIKOKOGA, M.D., Ph.D.**

Tokyo, Japan

**SHINJI KOSUGI, Ph.D.**

**NORITAKA HAI, Ph.D.**

Kyoto, Japan

---

**Liepmann's Phenomenon During Benzodiazepine Withdrawal**

**To the Editor:** Liepmann’s phenomenon has been described in the context of alcohol-related delirium (1). We present a case of Liepmann’s phenomenon apparently occurring with excessive use of diazepines.

Mr. A, a 44-year-old man, was admitted to the gastroenterology unit of Kitasato University for general fatigue of 1 week’s duration. He had been followed since he was 29 for ulcerative colitis and had undergone a total proctocolectomy with an ileal pouch and anastomosis at 37 years of age. On admission, he was alert, and the results of a physical examination, routine blood tests, and X-ray radiographs of his chest and abdomen were within normal limits. Brotizolam, 0.25 mg/day, was prescribed for insomnia.

On the second hospital day, a colorectal fiberoscopy revealed no substantial worsening of the mucosal lesions. At 3:00 a.m. on the fourth day, Mr. A began to wander about, saying, “Many strange little people are walking around,” “I’m coming home,” and “I must meet an appointed user now.” His restlessness alternated with sleeping every few minutes. A diagnosis of delirium was made, and haloperidol, 5 mg, was injected intramuscularly at 5:30 a.m. and had little effect on him. At 11:00 a.m., when a consultant psychiatrist gently closed Mr. A’s eyes and asked if he could see birds, Mr. A replied, “That’s right, I can see birds ([Hontoda, tori ga mieru](http://ajp.psychiatryonline.org)).” He also exhibited Liepmann’s phenomenon (1) in relation to a whale (i.e., he said that he could see a whale when the psychiatrist closed his eyes and asked if he could see a whale). A brain computed tomography scan revealed normal findings, and an EEG showed a low-voltage fast pattern with no paroxysmal discharges.

At 11:30 a.m., Mr. A’s father informed his psychiatrist by telephone that Mr. A consumed “too much alcohol and hypnotics” every night, but later, the “too much alcohol” was confirmed to be 350 to 500 ml of beer. At the time, Mr. A was unable to answer when asked whether he used any hypnotics. At around 2:00 p.m., while exhibiting Liepmann’s phenomenon in regard to an airplane, he repeated the names of the psychiatrists whom he usually consulted. The psychiatrists were contacted and informed us that Mr. A was taking 2.4 mg/day of alprazolam, 1.5 mg/day of etizolam, 0.5 mg/day of triazolam, 2 mg/day of estazolam, 0.25 mg/day of brotizolam, 2 mg/day of flurazepam, and 10 mg/day of zolpidem.

Oral diazepam, 20 mg over 24 hours, and drip infusion of flunitrazepam, 2 mg at night, was started. Since then, Mr. A has not exhibited Liepmann’s phenomenon. This delirious episode resolved in 5 days. His dose of diazepam was reduced to zero in 8 weeks. The episode of delirium may have been attributable to withdrawal from excessive use of sedative drugs. Liepmann’s phenomenon in Mr. A was observed exclusively during the delirium.

To our knowledge, few reports, other than the report by Miura et al. (2) on withdrawal from meprobamate (3000 mg/day) have described Liepmann’s phenomenon in conditions besides alcoholism. However, this case clearly demonstrates that it is necessary to be alert to the “concealed” or possible use of excessive diazepines underlying Liepmann’s phenomenon.

**References**


**MASANORI SAITO, M.D., Ph.D.**

**YOSHIKI MATSUI, M.D., Ph.D.**

**HITOSHI MIYAOKA, M.D., Ph.D.**

Kanagawa, Japan

---

**Increase in Risperidone Plasma Level With Lamotrigine**

**To the Editor:** A combination of clozapine and risperidone is effective in treating patients with schizophrenia who are unresponsive to other atypical antipsychotics or monotherapy with clozapine (1). Nevertheless, there are patients who respond only partially or even fail to respond to this combination. Saba et al. (2) and Dursun et al. (3) reported on patients who showed a substantial improvement of persistent positive symptoms when lamotrigine was added to clozapine therapy. There is evidence supporting the positive effects of lamotrigine based on its glutamate excess-release inhibition (4).

Because of these interesting reports, we decided to supplement the clozapine-risperidone combination of Ms. A, a 26-year-old inpatient who suffers from therapy-resistant schizophrenia with imperative auditory hallucinations, with lamotrigine. She had been taking clozapine, 550 mg/day, for 5 years and risperidone, 8 mg/day, for 4 weeks and had only responded partially. Her plasma levels of ris-
peridine (55–70 ng/ml) and clozapine (800–1100 ng/ml) were stable. Her lamotrigine dose was titrated up to 250 mg/day in steps of 25 mg per week. After Ms. A had been taking 175 mg/day of lamotrigine for 5 days, her plasma level was 5 mg/liter, and her risperidone plasma level was 69 ng/ml. We further increased her dose of lamotrigine to 200 mg/day. Her risperidone level rose to 284 ng/ml, and in follow-up measures, it showed a value of 412 ng/ml. Her clozapine level rose to 1300 ng/ml. Ms. A did not have any intoxication symptoms. Because we did not assume any connection of increased plasma level of risperidone with lamotrigine, we heightened the dose to 225 mg/day. The next measurement of both plasma levels indicated an exorbitant increase of risperidone plasma level, up to 412 ng/ml. Ms. A complained of dizziness and tiredness. We quickly reduced the dose of risperidone to 2 mg/day and withdrew the drug 1 week later. An overdose of risperidone was unlikely since it was taken under supervision.

Metabolism of risperidone occurs mainly in the liver and is dependent mostly on cytochrome P450 isoenzyme CYP 2D6. Lamotrigine does not inhibit CYP 2D6. It is eliminated by the kidneys after glucuronidation in the liver. Until now, we have had no explanation for the increase of the risperidone plasma level during concomitant therapy with clozapine and lamotrigine. Clinicians should be aware of this effect.

References
3. Dursun SM, McIntosh D, Milliken H: Clozapine plus lamotrigine in treatment-resistant schizophrenia (letter). Arch Gen Psychiatry 1999; 56:950

SARA DORIS BIENENTREU, M.D.
KLAUS-THOMAS HELMUT KRONMÜLLER, M.D.
Heidelberg, Germany

Neuroleptic Malignant Syndrome Induced by Quetiapine and Fluvoxamine

TO THE EDITOR: Neuroleptic malignant syndrome is induced less by atypical antipsychotics than by conventional antipsychotics, and one particular atypical antipsychotic, quetiapine, rarely causes neuroleptic malignant syndrome (1). Selective serotonin reuptake inhibitors (SSRIs) are prescribed worldwide for the treatment of depression, and SSRIs and atypical antipsychotics are often prescribed together as augmentation therapy for depression. Here, we describe a male patient who suffered from severe neuroleptic malignant syndrome induced by concomitant administration of quetiapine, 150 mg/day, and fluvoxamine, 100 mg/day, although he had not shown any side effects with monotherapy with quetiapine, 150 mg/day, or fluvoxamine, 150 mg/day.

Mr. A, a 57-year-old man who was diagnosed with major depression at age 56, according to DSM-IV criteria, had been treated with fluvoxamine, 150 mg/day, for 1 year. During the year of treatment, no remarkable side effects were apparent, and after his condition had fully remitted, fluvoxamine was tapered off. Five months later, he presented with irritation and agitation, and risperidone was prescribed for these symptoms. Because of drug-induced extrapyramidal symptoms, risperidone was subsequently replaced by quetiapine, 150 mg/day, which was effective for his irritation and left him free of extrapyramidal symptoms. However, 2 months later, he increasingly developed a depressive mood and inhibition. Thus, fluvoxamine, 50 mg/day, was prescribed, in addition to quetiapine, and fluvoxamine was increased to 100 mg/day 1 week later. On the 10th day of concomitant administration of quetiapine and fluvoxamine, Mr. A stopped eating and drinking and developed muscle rigidity. On day 13, he was admitted to our hospital. At this time, he had a high temperature, severe extrapyramidal symptoms, high blood pressure, and tachycardia and was falling into a stupor. Laboratory tests showed elevation of his creatinine phosphokinase (7,500 IU/liter) and leukocyte (1.3 × 10^9/liter) levels. We stopped all psychotropic drugs and immediately started infusion of dantrolene, under a diagnosis of neuroleptic malignant syndrome, and Mr. A’s symptoms improved gradually. However, on the fourth day of admission, he developed complications with acute pneumonia and respiratory failure and was transported to the intensive care unit and was treated with infusions of dantrolene and antibiotics. Three weeks later, he recovered from respiratory failure and neuroleptic malignant syndrome, and his laboratory measurements returned to normal.

In this case, concomitant administration of quetiapine and fluvoxamine caused neuroleptic malignant syndrome, although each drug alone did not cause any side effects. Since the doses of quetiapine and fluvoxamine were relatively low and these drugs are metabolized by different cytochrome P450 subtypes (2), induction of neuroleptic malignant syndrome was probably not due to an increase in the quetiapine and/or fluvoxamine concentrations. Hence, in this case, neuroleptic malignant syndrome may have been caused by a dopamine-serotonin disequilibrium (3), which was induced by concomitant quetiapine and fluvoxamine.

References

RYOHEI MATSUMOTO, M.D.
YURINOSUKE KITABAYASHI, M.D., Ph.D.
YASUHITO NAKATOMI, M.D.
HIDETO TSUCHIDA, M.D., Ph.D.
KENJI FUKUI, M.D., Ph.D.
Kyoto, Japan
Elevated Plasma Ghrelin Levels in Night-Eating Syndrome

To the Editor: Night-eating syndrome is characterized by repetitive awakening because of hunger and excessive eating. In humans, food intake appears to be regulated reciprocally by the anorexigenic leptin and the orexigenic ghrelin. The physiological nocturnal leptin surge is blunted in night-eating syndrome. Ghrelin levels are changed in eating disorders, with elevation in anorexia and Prader-Willi syndrome and blunting in obesity (2). In night-eating syndrome, ghrelin levels are unknown so far. We report the course of ghrelin plasma levels during the treatment of night-eating syndrome.

Ms. A, a 27-year-old woman, reported that the first symptoms of night-eating syndrome appeared 1 year before she underwent our psychiatric examination. As in previous reports, administration of a selective serotonin reuptake inhibitor (citalopram, 100 mg/day) was started (3). A complete remission was obtained within 8 weeks. Medication was continued for another 8 weeks before withdrawal. About 8 weeks later, Ms. A described a relapse of night-eating syndrome.

Her plasma ghrelin concentrations were measured by radioimmunoassay (Phoenix Pharmaceuticals, Belmont, Calif.) at three times (before treatment, 8 weeks later [during drug treatment and full remission], and finally, after relapse). Specimens were collected on all three occasions: every 20 minutes between 10 p.m. and 7:00 a.m. by a long catheter.

Mean nocturnal ghrelin concentrations (10 p.m. to 7 a.m.) were 1051 pg/ml before treatment, 977 pg/ml 8 weeks later (during drug treatment and full remission), and 1013 pg/ml after relapse. The body mass index (kg/m²) of Ms. A was about 23.5 at all examinations. As control subjects we investigated three healthy women, 22, 23, and 32 years old, with body mass indexes of 21.9, 22.4, and 20.8 kg/m². Their mean ghrelin concentrations were 372, 402, and 338 pg/ml, respectively.

This preliminary observation suggests that nocturnal ghrelin levels are enhanced in night-eating syndrome. Despite the known sleep-promoting capacity of ghrelin, its elevated levels might disrupt sleep because of hunger at night. Similarly, a higher dose of exogenous ghrelin caused nocturnal eating in a normal subject, whereas a lower dose promoted sleep (4). It appears that elevated ghrelin concentration is not state-dependent and may reflect a marker of vulnerability for developing night-eating syndrome.

For this study, written informed consent and human subjects research committee approval were obtained.

References

Mania in a Boy Treated With Modafinil for Narcolepsy

To the Editor: Modafinil is the first-line treatment for narcolepsy. It may also improve mood in narcoleptic patients (1). However, psychostimulants may exacerbate psychotic symptoms in psychotic patients (2). Cases of psychosis have also been reported during psychostimulant abuse (3) and during abuse of prescribed drug in narcoleptic patients (4) but not following medical use. Here we report the case of a boy with narcolepsy.

Albert was a 17-year-old boy who was diagnosed with narcolepsy at age 14. He was first prescribed modafinil, 400 mg/day for 1 year, switched to methylphenidate, 40 mg/day for 2 years, then returned to modafinil, 400 mg/day. The switching was because of complaints of irritability and of a lack of efficacy for sleep attacks. Albert then experienced flight of ideas, sexual excitation, and increased irritability. These manic symptoms resulted in friction with family members and a fight for which he could have been put on trial. Then, free of psychostimulant treatment, Albert was described as sad, anhedonic, and withdrawn. Following reintroduction of modafinil, the same manic symptoms reoccurred. After a meeting with a judge, Albert experienced self-referential thinking and suspiciousness. Later, a full manic episode developed within 3 days, including insomnia, tachypsychia, logorrhea, psychomotor agitation, and mood-incongruent psychosis. There was no grandiosity but delusion of persecution, based on auditory hallucinations (his uncle reproaching him for his past sexual behavior), complex visual hallucinations (a vampire hiding in his bedroom and trying to bite him), and a feeling of being talked to through the television. Albert was hospitalized, and the modafinil was stopped. The mania required pharmacological treatment that started after written consent was obtained from both Albert and his parents.

These mood symptoms seem time-related to psychostimulant administration and interruption. Exposure lasted for only 3 years, but discontinuation and reintroduction might have lowered the manic threshold. Contrary to previous reports of psychosis induced by psychostimulant abuse (3), the patient showed no trend toward dose escalation. This could be the first report of mania under a therapeutic dose of modafinil. The symptoms were compatible with psychostimulant-induced psychosis. Although an independent psychiatric disorder cannot be ruled out, we suggest a careful psychiatric monitoring of patients receiving modafinil and other psychostimulants for the treatment of narcolepsy.

References
Clozapine-Induced Agranulocytosis After 11 Years of Treatment

TO THE EDITOR: Clozapine can cause life-threatening agranulocytosis in up to 0.8% of patients treated with this medication (1). This limits the use of clozapine and mandates regular hematological monitoring. The risk of blood dyscrasia is highest in the initial 6–18 weeks but has been reported after years of treatment (2). We present the case of a patient with schizophrenia who developed clozapine-induced agranulocytosis after 11 years of pharmacotherapy.

Mr. A, a 46-year-old Hispanic man, was diagnosed with chronic schizophrenia in the 1970s and had only partial response to various antipsychotic medications until clozapine was initiated 11 years ago. He improved significantly while taking clozapine, 675 mg/day. Recently, his psychosis worsened. At one of his routine biweekly hematological screenings, his WBC count was 1,300/mm³, his neutrophil count was 12%, and his bands were 2% (bands are immature neutrophils that increase when infection is present; normal ranges: WBC count=4,800–10,800/mm³, neutrophil count=42%–75%, and bands=0%–5%). That prompted a referral to the hospital. During the workup, a urinary tract infection was documented, despite an absence of clinical symptoms or signs of infection. In the previous 4 months, his WBC count had fluctuated between 2,800/mm³ to 5,000/mm³, with granulocyte counts in the normal range, and he had three periods documented when his WBC counts were below 4,000/mm³ (normal WBC count range=4,000–12,000/mm³). These leukopenias lasted 26, 22, and 5 days each and spontaneously resolved without changes in clozapine dosing.

Mr. A was hospitalized with neutropenic precautions, and clozapine was discontinued. Because of his mental status deterioration, aripiprazole, 15 mg/day, was started orally on day 4. Although his WBC count had risen to 1,600/mm³, one 480-mg dose of recombinant granulocyte colony stimulating factor was administered on the same day. On day 6, upper gastrointestinal bleeding occurred. An endoscopy revealed gastric ulcers that were cauterized. On day 10, a urinary tract infection was treated with trimethoprim-sulfamethoxazole. Mr. A’s WBC count gradually normalized to 5,300/mm³ (neutrophil count of 45.7%) on day 14 and remained normal throughout his hospitalization. He became more organized after 3 weeks of aripiprazole, 40 mg/day, but he did not regain his previous level of functioning. The risk for bone marrow suppression precluded restarting clozapine.

Clozapine can cause life-threatening agranulocytosis, which mandates weekly monitoring of the CBC during the first 6 months of treatment and biweekly monitoring thereafter. Neutropenia has been documented after 2.5 years of pharmacotherapy (2), and agranulocytosis has been reported after 17 months of treatment (3). In our patient, bone marrow suppression developed after 11 years of otherwise uncomplicated successful treatment with clozapine. Early detection of agranulocytosis reduces mortality (4). Although it involved a single case, this report suggests the importance of continued monitoring of CBC in clozapine-treated patients, even after many years of uncomplicated use.

References
1. Alvir JMJ, Lieberman JA, Safferman AZ, Schwimmer JL, Schaaf (2). Finally, in a recent clinical trial by Goldberg and colleagues (3), one case of mania was reported in a patient with a personal history of bipolar depression while being treated with pramipexole. Here, we describe a case of mania in a patient with Parkinson's disease given pramipexole and ropinirole who had no personal or family history of bipolar disorder.

Pramipexole, Ropinirole, and Mania in Parkinson's Disease

TO THE EDITOR: Dopamine receptor agonists, such as pramipexole and ropinirole, are a safe and effective initial therapy for mild to moderate Parkinson's disease. There are at least three lines of evidence to suggest that this class of drugs may also be related to mood symptoms. First, at the clinical level, besides ameliorating motor symptoms, pramipexole has shown antidepressant effects in Parkinson's disease, in major depression, and in treatment-resistant unipolar and bipolar depression. Next, at the basic science level, pramipexole and ropinirole are novel dopamine receptor agonists with a high affinity for all dopamine D 2 subfamily receptors and show highest affinity for the D 3 receptor subtype (1). The antidepressant effect of pramipexole and ropinirole may be related to a resensitization or potentiation of the D 2/D 3 receptors in the mesolimbic system, a region relevant to mood regulation (2). Finally, in a recent clinical trial by Goldberg and colleagues (3), one case of mania was reported in a patient with a personal history of bipolar depression while being treated with pramipexole. Here, we describe a case of mania in a patient with Parkinson's disease given pramipexole and ropinirole who had no personal or family history of bipolar disorder.

Ms. A was a 37-year-old white woman with a 4-year history of Parkinson's disease. Her family history revealed a paternal grandmother with a single major depressive episode and a sibling with anorexia nervosa. Her Parkinson's disease symptoms had been treated with levodopa and

LETTERS TO THE EDITOR

http://ajp.psychiatryonline.org

Am J Psychiatry 162:4, April 2005

814
Mr. A, a 35-year-old Japanese man without a previous psychiatric history, was seen with perceptual disturbances. One month before his evaluation, he had stopped using 5-MeO-DIPT because of a so-called bad trip—anxiety, palpitations, auditory oversensitiveness, and visual distortion—after six or seven times using between 15 mg and 30 mg of 5-MeO-DIPT over 5 months. He was bisexual and had used the drug to enhance intercourse with a male partner. A few days before his evaluation, after the announcement of his father’s diagnosis of a brain tumor, his 5-MeO-DIPT-induced phenomena of a “bad trip” returned, although he had not taken 5-MeO-DIPT.

There was no evidence of CNS infection or organic brain disease. Amphetamine was not detected in Mr. A’s urine. He was not clinically depressed. Schizophrenia-like symptoms, such as delusions or auditory hallucinations, were not present. He was given oral risperidone, 1 mg/day. Within 3 days, his perceptual disturbances remarkably decreased, and 7 days later, they had almost completely disappeared. Given his clinical features and history of drug ingestion, we made a diagnosis of hallucinogen-persisting perception disorder induced by 5-MeO-DIPT. Mr. A was discharged 1 month later. Although this medication was maintained for 4 months and then terminated, he has had no relapse.

The disturbances of serotonergic function may be a factor in hallucinogen-persisting perception disorder, although the pathophysiology remains unclear. Regarding the treatment of LSD-induced flashbacks, the choice of medication is still controversial. The use of various agents, including neuroleptics, serotonin reuptake inhibitors (SSRIs), anticonvulsants, and benzodiazepines, has met with limited success (2). Some researchers report that risperidone, which is a serotonin-dopamine antagonist, exacerbates symptoms of hallucinogen-persisting perception disorder (3). Others note that SSRIs exacerbate flashbacks (4). In this case, his perceptual disturbance symptoms responded to risperidone treatment.

For public safety, 5-MeO-DIPT is a controlled substance in several countries. However, it is available in many areas, and the patient obtained it through the Internet quite easily. We are concerned that the abuse of 5-MeO-DIPT may be more widespread than previously thought. We believe that studies are needed to verify the relationship between 5-MeO-DIPT and hallucinogen-persisting perception disorder and to call public attention to the toxicity of 5-MeO-DIPT.

References

5-Methoxy-N,N-Diisopropyltryptamine-Induced Flashbacks

TO THE EDITOR: The drug 5-methoxy-N,N-diisopropyltryptamine (5-MeO-DIPT) has hallucinogenic and mild euphoric properties, similar to those of other tryptamine compounds (1). Hallucinogen-persisting perception disorder is characterized by the transient recurrence of perceptual symptoms experienced while intoxicated with the hallucinogen, often called “flashbacks.” LSD-induced flashbacks are well known. It is also reported that hallucinogen-persisting perception disorder is induced by cannabis and methamphetamine. However, to our knowledge, there are no published reports of 5-MeO-DIPT-induced hallucinogen-persisting perception disorder.

Abhay Singh, M.B.B.S.
Robert Althoff, M.D., Ph.D.
R. Jared Martineau, M.D.
James Jacobson, M.D.
Burlington, Vt.

References

Abhay Singh, M.B.B.S.
Robert Althoff, M.D., Ph.D.
R. Jared Martineau, M.D.
James Jacobson, M.D.
Burlington, Vt.
Rate of Tardive Dyskinesia in Hospitalized Patients

TO THE EDITOR: In the era of typical antipsychotic medications, tardive dyskinesia was a common and often undiagnosed disorder (1). The newer atypical antipsychotic medications hold the promise of a lower prevalence of movement disorders. We conducted a study whose goal was to determine if this promise has translated from the systematic research to a naturalistic sample.

Evaluations were completed for 162 patients from Central State Hospital in Petersburg, Va. The sample consisted of chronically ill, treatment-resistant adult patients, most of whom had psychotic disorders (70%) or bipolar disorder (9%) based on DSM-IV criteria. The majority (94%) of patients were taking antipsychotic medication. Of this subset, 25% were taking a typical antipsychotic but not an atypical antipsychotic, 52% were taking an atypical antipsychotic but not a typical antipsychotic, and 23% were taking both. Each patient was rated with the Involuntary Movement Scale (2). The primary rater (D.E.R.) was trained by the developers of this scale and achieved a high level of reliability with respect to gold standard ratings (intraclass correlation coefficient [ICC]=0.96 for parkinsonism global rating, p<0.001; ICC=0.79 for dyskinesia global rating, p<0.001). A diagnosis of parkinsonism was defined as having a rating of 2 (mild) or higher on the global rating scale. Tardive dyskinesia was diagnosed by the criteria of Schooler and Kane (3).

Per hospital policy, almost all of these patients had a previous examination for dyskinesia by their attending physician with the Abnormal Involuntary Movement Scale (4). These data were used to compare the prevalence of diagnoses of tardive dyskinesia between the attending physicians and the authors of the current study.

Tardive dyskinesia was present in 40% of the patients (40% of the patients taking only typical antipsychotics, 39% of the patients taking only atypical antipsychotics, and 47% of the patients taking typical and atypical antipsychotics). Of the patients with tardive dyskinesia, 57% had not been diagnosed previously with tardive dyskinesia. Of the patients newly diagnosed with tardive dyskinesia, 48% were mildly ill, 48% were moderately ill, and 4% were severely ill.

For patients taking any antipsychotic medication, 49% had parkinsonism (67% of the patients taking only typical antipsychotics, 34% of the patients taking only atypical antipsychotics, and 60% of the patients taking typical and atypical antipsychotics). Parkinsonism was uncommonly mentioned as an adverse effect, and no rating scale was used by the attending physicians to assess parkinsonism.

A substantial number of patients had tardive dyskinesia or parkinsonism. The high prevalence of newly diagnosed cases of tardive dyskinesia suggests a continuing need for better identification of these disorders and the opportunity for further preventing or treating these disorders.

References

DAVID E. ROSS, M.D.
MICHELE THOMAS, PHARM.D., B.C.P.P.
MARIA BOOTH, R.N., C.S.
MICHAEL WEINBORN, PH.D.
Midlothian, Va.

Do Smokers Who Commit Suicide Have High Blood Levels of Nicotine?

TO THE EDITOR: Cigarette smoking is associated with a higher risk for suicide and attempted suicide (1, 2). In addition, an association between cigarette smoking and suicidal behavior across major psychiatric disorders may be related to lower brain serotonin function in smokers with depression (2). Unfortunately, no information is available concerning nicotine and cotinine levels in the body fluids of smokers who committed suicide.

We examined the nicotine and cotinine levels in the blood and urine of 36 forensic autopsy cases with no obvious putrefaction that were handled between October 2002 and March 2004. Our cases consisted of eight smokers who committed suicide (six men, 51–76 years of age, and two women, 54–79 years of age), eight smokers who did not commit suicide (seven men, 36–79 years of age, and one woman, 54 years of age), and 20 nonsmokers (15 men, 19–97 years of age, and five women, 62–82 years of age). One suicide case was present in the nonsmoking group. None of the suicide smokers consumed nicotine preparations or tobacco leaves.

Eight suicide smokers had high levels of nicotine and cotinine in their blood (mean=115 ng/ml, SD=49, and mean=405 ng/ml, SD=291, respectively) and urine (mean=1940 ng/ml, SD=2340, and mean=1170 ng/ml, SD=1570, respectively). In contrast, eight nonsmoker smokers had lower levels of nicotine and cotinine in their blood (mean=30.1 ng/ml, SD=17.7, and mean=122 ng/ml, SD=65, respectively) and urine (mean=383 ng/ml, SD=417, and mean=170 ng/ml, SD=86, respectively). Blood nicotine levels in our eight suicide smokers were significantly higher than in nonsuicide smokers (t=4.61, df=14, p=0.0004). Although urine nicotine level and blood and urine cotinine levels were higher in suicide than nonsuicide smokers, they did not achieve a level of statistical significance. In six of 20 nonsmokers, nicotine and/or cotinine originating from passive smoking were detected. Neither nicotine nor cotinine was detected in the remaining 14 cases.

Ura et al. (3) reported that habitual smokers with no psychiatric disorders who consumed 26 or more cigarettes per day had blood nicotine levels of 20.0–680.0 ng/ml (mean=35.2 ng/ml, N=34) in a free smoking experiment. Blood nicotine levels measured in our eight nonsuicide smokers were similar to their results.

Our data, although they are preliminary and limited in number, strongly suggest that a marked increase in cigarette smoking in persons with psychiatric disorders may be a sign of an imminent suicide attempt. Consequently, the smoking status of psychiatric patients may serve as a clinical sign of...
their mental status and provide an early warning sign of a possible suicide attempt.

References

FUMIO MORIYA, Ph.D.
YOSHIKI HASHIMOTO, M.D.
Kochi, Japan

Dementia With Lewy Bodies, Visual Hallucinations, and Medications

To the Editor: In their recent study, Clive G. Ballard, M.R.C.Psych., M.D., et al. (1) “confirm” high frequencies of visual hallucinations and delusions in dementia with Lewy bodies and also conclude that visual hallucinations are significantly more persistent in this disorder than in Alzheimer’s disease. Although extensive clinical evaluations were performed before death, the authors do not report the medication status of their patients. The impact of dopaminergic drugs on the mental state of demented parkinsonian patients should not be ignored. It is interesting that 66% of the patients with dementia with Lewy bodies in this study had visual hallucinations. A prior meta-analysis of dementia with Lewy bodies reported noted that 68% of the patients with dementia with Lewy bodies receiving dopaminergic drugs had visual hallucinations, but only about half that rate was found in medication-free patients (2). Dr. Ballard et al. may be prematurely attributing visual hallucinations to the pathological process of dementia with Lewy bodies per se rather than to an epiphenomenon, i.e., medication status. A review of their patients’ medications could shed light on this question.

References

MICHAEL SERBY, M.D.
STEVEN C. SAMUELS, M.D.
New York, N.Y.

Child Psychopharmacology, Effect Sizes, and the Big Bang

To the Editor: We read with interest the article by Karen Dineen Wagner, M.D., Ph.D., et al. (1) in the June issue. In their study comparing citalopram to placebo, we were surprised to find the authors reporting an overall effect size of 2.9. With the commonly cited criteria set forth by Cohen, effect sizes can be considered trivial (<0.2), small (0.2 to <0.5), moderate (0.5 to 0.8), or large (>0.80).

By these metrics, the reported effect size can be characterized as gargantuan, big bang-worthy. The value does not appear to be a benign typographical error for “0.29,” given that “2.9” appears twice. An accurate effect size cannot be manually calculated with the information provided in the article. However, in order to arrive at the effect size of 2.9, it can be estimated that a pooled standard deviation of the change score of 2.1 would have been required. Such a narrow standard deviation of the change score seems improbable (a manual calculation with the Ns and standard deviations in the article yields a value of 15.6, for an effect size of 0.4). Moreover, such a low standard deviation of the change score would suggest uniformity in response that is far from consistent with comparable studies.

We surmise one of two possibilities. The first is that a simple arithmetic mistake occurred and was not picked up, despite otherwise meticulous attention to detail. A trickster decimal point may be to blame, and a demoted effect size of 0.29 may gain in honesty what it loses in the sex appeal of an inflated 2.9 status. A smaller effect size seems more plausible, and not only because a meta-analysis of 33 trials of selective serotonin reuptake inhibitors (SSRIs) for the treatment of adult depression (2) arrived at a pooled effect size of 0.4 but because the current study, although statistically significant, was not that clinically impressive. Only 36% of the patients treated with citalopram responded, compared to 24% of those with placebo (for a lukewarm number needed to treat of 8). These results, while modest, are respectable in their own right and nothing to sneeze at in a clinical area that has been short on proven therapeutic options. But a majestic sequoia of 2.9 they are not.

Alternatively, the authors may have used a different definition or formula to calculate the effect size. This would be unfortunate because the basic job description of an effect size is to facilitate communication among investigators and across measures. The gargantuan 2.9 becomes an unfortunate jar-scr juice of nails against the chalkboard: it robs from the melody of welcome that this timely contribution otherwise merits.

References

ANDRÉS MARTIN, M.D., M.P.H.
WALTER S. GILLIAM, Ph.D.
New Haven, Conn.
JEFFREY Q. BOSTIC, M.D., Ed.D.
Boston, Mass.
JOSEPH M. REY, M.D.
Sydney, Australia

Am J Psychiatry 162:4, April 2005
http://ajp.psychiatryonline.org

LETTERS TO THE EDITOR
ized clinical trial reporting has been described in the CONSORT statement (1); one of its recommendations is to de-
scribe the flow of the subjects in the study (number screened, proportion randomly assigned, etc.). Dr. Wagner and col-
leagues did not report the proportion of subjects who were
excluded from the random assignment after the single-blind
period. This information is critical because a placebo run-in
period might help to “wash out” nonspecific responders, al-
lowing sharper evaluation of treatment-specific effects as
shown in some pharmacotherapy studies (2).

An additional concern is the elicitation method used for
adverse events at a time when the safety of SSRIs in youth
has been called into question (3). The adverse events were:
“reported by patients or observed by investigators” (Wagner et
al., p. 1080). The reliability of this practice is questionable
because some adverse events, even very severe ones, could
neither be reported by the patient nor observed by the investiga-
tor and would need to be specifically assessed (4).

Finally, it is somewhat surprising that the authors do not
compare their results with those of another trial, involving
244 adolescents (13–18-year-olds), that showed no evidence
of efficacy of citalopram compared to placebo and a higher
level of self-harm (16 [12.9%] of 124 versus nine [7.5%] of 120)
in the citalopram group compared to the placebo group (5).
Although these data were not available to the public until De-
ceMBER 2003, one would expect that the authors, some of
whom are employed by the company that produces citalo-
pram in the United States and financed the study, had access
to this information.

References
R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the qual-
ity of reporting of randomized controlled trials: the CONSORT
statement. JAMA 1996; 276:637–639
2. Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Car-
mody T, Rintelmann J: A double-blind, randomized, placebo-
controlled trial of fluoxetine in children and adolescents with
depression. Arch Gen Psychiatry 1997; 54:1031–1037
3. US Food and Drug Administration: Antidepressant Use in Chil-
drug/antidepressants/default.htm
JS, Levine J, Fried J, Abikoff H, Zito JM, McCracken JT, Findling
RL, Robinson J, Cooper TB, Davies M, Varipatis E, Labellarte MJ,
Scallion L, Walkup JT, Capasso L, Rosengarten J: Review of safety
assessment methods used in pediatric psychopharmacology. J
5. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Bod-
dington E: Selective serotonin reuptake inhibitors in childhood
depression: systematic review of published versus unpub-

RÉMY P. BARBE, M.D.
Pittsburgh, Pa.

TO THE EDITOR: We read with interest the study by Dr. Wagner
et al. We have a number of concerns about this study. In the
Method section, it is not clear how the patients were re-
cruited. One is also left in the dark about the method of ran-
dom assignment and if the random assignment list was con-
cealed. The authors also give no indication of how they

arrived at the sample size and if a power calculation was done.
Given the recent concerns about the risk of suicidal thoughts
and behaviors in children treated with SSRIs, this study could
have attempted to shed additional light on the subject. The
authors called the analysis of data an intent-to-treat analysis,
although four patients who were lost to follow-up were ex-
cluded. In a true intent-to-treat analysis, all patients are ana-
alyzed in the groups to which they were initially assigned,
regardless of whether they received the treatment or not. We
consider the use of the term “intent-to-treat” in this context
misleading.

Dropouts from the study have been accounted for by using
the last observation carried forward. Treatment response in
depression is frequently followed by a subsequent return
to original or baseline values on a scale such that the last ob-
servation carried forward may be an unduly optimistic esti-
mate. The classification of dropouts as treatment failures is
based on safer assumptions than the last observation car-
rried forward.

Our greatest concern is with the results and conclusions
drawn. There is no table showing the results in detail. The au-
thors have only stated that 36% of citalopram-treated patients
met the criteria for response, compared to 24% of patients re-
cieving placebo. This response rate, while in itself marginal
compared to other studies of antidepressants, does not in it-
self show that citalopram is better than placebo.

We calculated the absolute benefit increase of using citalo-
pram as 0.12 (95% confidence interval [CI]=−0.015 to 0.255).
The relative benefit increase that could be attributed to citalo-
pram was 50% (95% CI=−135% to 6%). The odds ratio, i.e., the
odds of improving while taking citalopram compared to pla-
cebo was 1.75 (95% CI 0.92 to 3.43). The number needed to
treat, i.e., the number of children who need to be treated with
citalopram for one additional positive outcome was eight
(95% CI=4 to infinity). None of these shows that citalopram is
any better than placebo.

We would argue that the authors did not provide sufficient
evidence to support their claim that “citalopram produces a
statistically and clinically significant reduction in depressive
symptoms in children and adolescents” (p. 1082). We are sur-
prised that the most respected psychiatric journal in the
world published a study that is misleading to its readers in the
extreme.

MANU MATHEWS, M.D., M.R.C.PSYCH.
BABATUNDE ADETUNJI, M.D., F.A.S.A.M.
JOANNE MATHEWS, M.D.
BIJU BASIL, M.D.
VINU GEORGE, M.D.
KUMAR BUDUR, M.D.
Cleveland, Ohio
SHINY ABRAHAM, M.D.
Kottayam, India

Dr. Wagner and Colleagues Reply

TO THE EDITOR: Dr. Mathews and colleagues request further in-
formation about the randomized, placebo-controlled trial of
citalopram for treatment of depression in children and ado-
lescents. Randomization was on a 1:1 basis and was stratified
by age group. The random assignment list was concealed
from the investigators, which is fundamental to the claim that

http://ajp.psychiatryonline.org
Am J Psychiatry 162:4, April 2005
the study was performed under double-blind conditions. The protocol-specified population for all efficacy analyses, defined as the “intent-to-treat” population, included all patients who received at least one dose of double-blind study medication and had at least one postbaseline efficacy assessment. The analyses we presented in the manuscript were not only conventional in nature; they were, in fact, defined a priori. The justification for defining this population for the efficacy analyses is that the primary analysis was the change from baseline, therefore requiring a postbaseline assessment.

Although recently a mixed-model approach has gained some currency for the analysis of efficacy in antidepressant trials, the last-observation-carried-forward method of analysis has always been conventionally considered the most conservative method of analysis. Certainly this was the case when the study protocol was being developed. In escitalopram trials in adult patients, last-observation-carried-forward analyses minimize the treatment effect that is demonstrated by observed-cases analyses of the patients who actually remain in treatment (1, 2). These analyses are considered more conservative than observed-cases analyses for acute treatment antidepressant studies because the onset of antidepressant effect is typically delayed for up to several weeks. Therefore, the last observation of patients who discontinue active treatment prematurely is not likely to capture the full potential antidepressant effect.

Regarding suicidality, it is helpful to note that the manuscript states clearly that no serious adverse events were observed in the trial for citalopram-treated patients. At the time the manuscript was developed, reviewed, and revised, it was not considered necessary to comment further on this topic.

Dr. Martin and colleagues inquire about the value of 2.9, which was calculated as the quotient of the least square mean, divided by the common standard error of the mean for each treatment group. With Cohen’s method, the effect size was 0.32.

In response to Dr. Barbe’s questions about the methods of this randomized clinical trial for the treatment of depressed children and adolescents, there were 75 subjects who were screened but not randomly assigned. The method for elicitation of adverse events was chosen because it was the accepted standard at the time the study was designed for multicenter, industry-sponsored clinical trials in juvenile depression.

It may be considered premature to compare the results of this trial with unpublished data from the results of a study that has not undergone the peer-review process. Once the investigators involved in the European citalopram adolescent depression study publish the results in a peer-reviewed journal, it will be possible to compare their study population, methods, and results with our study with appropriate scientific rigor.

We believe that the results of our study, which demonstrated a significant difference between citalopram and placebo beginning at week 1, is clinically meaningful, particularly at a time when there have been so few antidepressants shown to have superiority to placebo for depressed children.

References

KAREN DINEEN WAGNER, M.D., PH.D.
ADELAIDE S. ROBB, M.D.
ROBERT L. FINDLING, M.D.
JIANGQING JIN, PH.D.
Galveston, Tex.

Reprints are not available; however, Letters to the Editor can be downloaded at http://ajp.psychiatryonline.org.
Ethics, Values, and Religion


Ethics is a topic of increasing salience in modern society. The multiplicity of theoretical perspectives and the diversity of practical ethical dilemmas in modern society make ethics a topic of increasing complexity as well.

Professor Aiken’s text seeks to bring coherence to our understanding of moral aspects of 21st century life experience and clinical practice. The result is a unique and expansive text, which, unfortunately, does not have sufficient theoretical or empirical grounding to represent academically rigorous scholarship in the field. Many readers will nevertheless enjoy this book for its sheer breadth and eclecticism.

The first chapter of the two-part book covers the conceptual basis of ethics as informed by insights drawn from religion, behavioral science, and philosophy. The next chapter focuses on moral development, discussing a range of perspectives from psychoanalysis to Piaget and from the “Defining Issues Test” to the moral impact of different parenting styles. Subsequent chapters address differences in moral conduct of individuals and groups as reflected in social scientific data and key issues in education such as the “Character Education Manifesto” and fundamental values of academic integrity.

The second part of the book focuses on ethical issues, principles, and practices in different applied contexts. The chapters include bioethics, research ethics, business ethics, ethics in law and government, media ethics, sports and sexual ethics, and environmental ethics and international relations. Aiken treats a wide set of intriguing issues such as euthanasia, research misconduct, ethics and war, professionalism in the military, and “cyberethics.”

The book reads like a long commentary rather than a formal analysis—Aiken freely offers his perspectives throughout the text. For example, in the chapter on moral education he states,

During the past few decades, violence, murder and suicide, precocious sexual behavior, and other immoral and illegal behaviors have continued without appreciable reduction among American children and adolescents. The decline in ethics and morality has, however, not gone unnoticed.

Assertions of this nature make the text interesting, provocative, and readable. Subsequent authors (building on this text) will want to develop the empirical and conceptual substantiation needed to support—or refute—Aiken’s observations.

The glossary at the end of the book is extensive. For instance, Aiken provides a mix of definitions of colloquial terms (“cooking data” and “bait-and-switch”), scientific and clinical concepts (“inheritable genetic modification,” “attention deficit hyperactivity disorder” and “ego ideal”), legal terms (“usury” and “antitrust legislation”), cross-cultural beliefs and religious ideals (“Dukkha,” “Jen,” “jihad,” and “Diaspora”), as well as subtle philosophical concepts (“categorical imperative” and “deontology”). A strength of the book is special features such as visually accessible tables and the series of questions and URLs for web sites with further information at the end of each chapter.

In sum, this book is unusual for its ambitious scope and atypical style. The writing is very clear and readable, and the structure of the text is evident. Some readers may take issue with its editorial style and with Aiken’s interpretation and application of the concepts he articulates. It is best viewed as a long, off-beat “treatise,” as it is accurately described on the back cover. It is a document that challenges readers to think in a more eclectic and interdisciplinary manner, and it will leave them with greater awareness of just how much ethical issues influence us personally and professionally on a daily basis.

RICHARD H. GIBSON, M.D.
LAURA WEISS ROBERTS, M.D., M.A.
Milwaukee, Wis.


Mainstream psychology is once again joining dynamic psychiatry and transpersonal psychology in the scientific study of character and its development. This book is the first progress report of the effort by positive psychologists to develop a “classification of the sanities.” The authors provide a list of traits that can be studied in future work on well-being. However, an adequate positive psychology of character must show how anyone can learn to feel good, not just some people with special personality traits. Therefore, the development of a taxonomy of well-being requires consideration of character development within a comprehensive psychobiological model of personality, as is done elsewhere (1).

The present volume is largely composed of well-informed reviews of 24 personality traits that are called “character strengths” because each is asserted to be socially valued in its own right and also provides one of many alternative paths to virtue and well-being. Each trait is considered in a separate chapter along with psychometric evidence documenting its reliable measurement. Each trait is defined behaviorally and illustrated by a representative case. For example, being asked for advice by others is taken as a measure of a wise perspective, so the late advice-columnist Ann Landers is suggested as a paragon of wisdom. Not everyone will agree with such choices, but simple behavioral measures do bring concepts of virtue down to a practical level of everyday experience that is understood and approached by many people.

The focus on common behaviors facilitates reliable measurement and gives hope that many people are on the path to well-being. Such hope is limited, however, by the facts that not everyone has the strengths listed and that some putative strengths also have disadvantages. For example, curiosity (novelty seeking) can certainly have its disadvantages, such as greater risk of abuse of illegal substances. Their paragon of
curiosity is John Lilly, a psychologist notorious for substance abuse, as depicted in the movie *Altered States*.

Peterson and Seligman are more convincing in their challenge to the relativistic assumption that values depend largely on cultural conventions. They suggest that a small number of specific personality traits have been consistently valued as adaptive in all humans regardless of culture. Their important conclusion is supported by a review of virtues recognized in each of the great world cultures (China, India, Greece). For example, their list of six core virtues is a slight rearrangement of the seven virtues of the Christian tradition: temperance (e.g., modesty, self-control), justice (e.g., fairness, citizenship), courage (e.g., fortitude, bravery), practical wisdom, humanity (e.g., charity), and transcendence (e.g., hope, faith).

Peterson and Seligman are self-proclaimed agnostics who specifically deny any faith in the divine. To accommodate the prominent role of faith in the happy life, they recognize it as an alternative to hope on the path to their overarching concept of transcendence. However, empirical findings show that the character traits that measure faith, hope, and charity are all interdependent and synergistic in making a person feel good (1). This casts doubt on the agnostic view that hope and faith are simply alternative paths to well-being. When faced with adversity or death, we realize that the moral level of intellect and virtue recognized by agnostics is hard and incomplete—doomed to heroic struggle and despair, as acknowledged by Freud and Erikson. Without faith, hope and love cannot reach a transcendental level that is spontaneous and unconditional.

The major accomplishment of this book is in showing that empirically minded humanists can measure character strengths and virtues in a rigorous scientific manner. Peterson and Seligman are forthright in stating that their theoretical perspective has not been adequate to produce a taxonomy in which specific configurations of traits predict a person's level of well-being. Their materialistic worldview proves inadequate to provide a theoretical account of transcendental phenomena like creative gifts and the contemplative experiences that are the foundation for faith and spirituality (1).

Although Peterson and Seligman are agnostics, they have now observed in their own factor analytic work that spiritual faith is a major dimension of character independent of hopeful self-directedness and charitable cooperativeness. Their finding confirms earlier psychometric work showing that spirituality is an important dimension of character that contributes to well-being (1–3). I hope that the authors' integrity and open-minded humility will serve as an inspiration for other empirically minded humanists to evaluate the adequacy of their own worldviews, no matter what conclusions they may reach.

**References**


C. ROBERT CLONINGER, M.D.

*St. Louis, Mo.*

### “Are You There Alone?” The Unspeakable Crime of Andrea Yates, by Suzanne O’Malley.


When a mother kills her children, how much does mental illness matter when the mother’s guilt is judged in the courtroom? The case of Andrea Yates, who drowned her five children on June 21, 2001, suggests that in some cases the verdict falls before the trial starts. Although abundant evidence exists to prove that Ms. Yates suffered severe mental illness in the 2 years before and at the time of the tragedy, psychosis and delusional hopelessness were not enough for her to be judged not guilty by reason of insanity in court.

The case took an unexpected turn recently when the trial court’s verdict was overturned on appeal. Although the appeals court’s reasoning focused on an error by the testifying forensic psychiatrist, it is a reasonable inference that the court’s ruling was based on the assumption that, other things being equal, the jury was at a tipping point. Given the facts presented, for the jury to have been at a tipping point can be understood as a reflection of a folk psychology whereby people are predisposed by the horror of an act itself to use judgmental heuristics. It is thus no wonder that Andrea Yates’s acts are understood more easily as bad rather than mad, regardless of the fact pattern.

The puzzling story of Andrea Yates has now received a much needed recounting from journalist Suzanne O’Malley. “Are You There Alone?” is a heartfelt account of the events that led to the tragic deaths of Noah, John, Paul, Luke, and Mary Yates. O’Malley argues that psychosis with manic features, combined with medical mismanagement, stressful circumstances, and religious obsessions masking delusions, resulted in the tragedy. Her reading of the health records presents Andrea Yates’s treatment as a litany of misdiagnoses, poor treatment, wrong medications, and the role of the health insurance company rather than the clinician as the key decision maker. Nonetheless, despite being fragmented and confusing, the medical records documented that Andrea Yates suffered serious psychotic illness and delusions before and after she drowned her children. Mentally ill or not, however, she appeared to admit to knowing that what she did was legally wrong in videotaped interviews shown in court, and the death-qualified jury found her guilty and sane according to Texas laws.

The verdict will continue toward further appeal and a potential retrial or plea bargain. O’Malley’s account gives rise to questions on which a potential appeal ruling or any retrial could turn. One such question is, How valid are videotaped interviews for forensic purposes with psychotic individuals? Especially when the psychoses of those individuals before they committed the acts in question included that they were being videotaped! Moreover, by the time the videos were shot, Andrea Yates had already been repeatedly interviewed. In her aloneness with the terror of psychosis, with her delusions masking guilt and grief over her abhorrent deed and unimaginable loss, might she not seek nonverbal cues and guidance for how to maintain connection? We do not read that there was any serious exploration as to whether, in her suffering, she might have had a natural need to turn her interviewers into unwitting directors to absolve her of an otherwise unbearable confrontation with the horror.
Although forensic psychiatrists are trained to examine accused persons such as Andrea Yates for feigning madness, it is far more difficult to detect the accused feigning badness or filling in the blanks as we might expect them to. Some accused would rather present themselves as bad than mad, more terrified of the aloneness of the latter than the legal consequences of the former. In this instance, if a trained, thoughtful, and experienced forensic psychiatrist could, as any human being might, become confused in the heat of cross-examination between what he was told and his observations, then is it not as likely that Andrea Yates, in the midst of the unbearable grief that the death of her children brought to the surface, might have become confused between what she imagined she was supposed by society to say to the videotape-directing interviewer and what she actually remembered? Neither Andrea Yates nor Dr. Park Dietz should be scapegoated for the failures of the mental health and medicolegal systems.

O’Malley succeeds in providing detailed, memorable descriptions of the horror, and she explicates formerly mysterious issues of the religious influences of Mr. Woroniecki, the role of Randy Yates, and the political and financial aspects of the trial. Psychiatric ethics courses can use “Are You There Alone?” to raise haunting questions regarding the injustice of a social and medical system where psychotic patients feel they need to present themselves as bad rather than mad.

BEATA ZOLOVSKA
Boston, Mass.

HAROLD J. BURSZTAJN, M.D.
Cambridge, Mass.


Dr. Favazza has taken on the daunting task of reviewing the interface between the Christian religious belief system, especially the Bible, and psychiatry. As a clinical psychiatrist and a seminary graduate, I was eager to read this book. I was engaged by some aspects of the book but disappointed by others.

The Bible has been used, misused, and abused throughout the centuries. It has been the basis for great good, but unfortunately also for evil: to justify slavery, wars, persecutions (the persecuted having become the persecutors), financial exploitation, environmental destruction, subjugation of women, and denial of citizenship rights to homosexuals. These abuses happen because the Bible is a powerful tool that can be wielded by greedy, arrogant, bigoted, and otherwise sinful human beings to support their own agendas. We human beings, present company included, are all sinners (Romans 3:23), a religious principle that is empirically verifiable.

Chapter 4, on homosexuality, is among the better sections of PsychoBible. What one sees in the Bible is often directly related to one’s preexisting opinions, and the issue of homosexuality is perhaps the most telling example. Dr. Favazza identifies the first-century context of pagan temple prostitution in which Paul’s letters were written and demonstrates how those whose political agenda is against homosexual rights lift these passages out of context. People who are negatively predisposed toward homosexual rights tend to see in Paul’s letters a blanket condemnation of all homosexual behavior. Those who are predisposed to support homosexual rights, on the other hand, tend to see a more limited condemnation of the fornication and temple prostitution practices that were widespread at that time.

PsychoBible contains some good discussions of the ideas of Freud, Jung, and others pertaining to religion. There is a great deal of helpful historical background information on the compilation of the Old and New Testaments. There is coverage of healing through prayer and interesting topics in neurotheology, such as out-of-body experiences.

Those who claim that the Bible is entirely without error (“verbal and plenary inspiration”) are themselves demonstrably mistaken, but to call such people “stubborn” and “lazy,” as Dr. Favazza does (p. 324), seems inappropriate for a scholarly work. Bias in the opposite direction (the “minimalist” position) assumes that everything in the Bible is incorrect or is a culturally motivated distortion until proven otherwise. Dr. Favazza cites the Jesus Seminar (a group of minimalist scholars) repeatedly and approvingly but fails to note that its findings are arrived at on the basis of serial speculative assumptions. Between these two extremes can be found more moderate and evidence-based ways to consider the accuracy of the Bible.

PsychoBible is weak on matters of theology, and there seems to be little understanding of the nuances of basic Christian beliefs, such as the Trinity and the Atonement. Martin Luther’s contributions are given a few scant lines, while his negative aspects (he was anti-Semitic, had bowel problems, and used vulgar language) get extended coverage. One does not come away with a balanced understanding of Luther. Several pages are spent recounting the hypothesis, accepted by almost no one, that Jesus faked his own death and resurrection (pp. 69–74).

Dr. Favazza makes several dismissive statements for which he gives no supporting evidence. One example is that it is “doubtful” that Luke was either a physician or a companion of Paul (p. 7). Although it is true that the attribution of authors’ names to the gospels was done decades after their writing, there is a case to be made that Luke, the first-century companion of Paul, did in fact provide the source material on which the present Gospel of Luke is based. Another example is the statement that Isaiah’s Old Testament prophecy of the Suffering Servant “certainly was describing Israel” (p. 26) as opposed to the future Christ.

PsychoBible is written throughout with a tone of flippancy. For example, “Praise the Lord and Pass the Medication” (p. 243) is the title of one chapter. Flippancy might make the book entertaining but also conveys a message about how the author regards this subject matter. There are some notable misspellings (Roman emperors “Troyan,” p. 48, and “Domition,” p. 29) and incorrect uses of theological terms (“divine entities” would not include “angels, saints, and demons” [p. 22]; monotheism and henotheism are confused [p. 20]). Much of the material in the Notes section would have been better incorporated into the main body of the text. There is an extensive bibliography, but citations in the text are sometimes inadequate. I tried to track down the original source page numbers for a quotation from St. Augustine cited on pages 15 and 330 but came to a dead end.

I recommend this book, but not without reservations, to discerning readers who are interested in human behavior and
Christianity. There is much here that is useful, but PsychoBible is not without flaws. Informed readers should expect to find some areas in which they will disagree with Dr. Favazza.

OWEN D. BUCK, M.D., M.S., M.T.S.

Lewiston, Me.


This is an unusual book to be reviewed—unusual in that it is essentially a monograph that presents some of the results of Dr. Stevenson’s lifelong researches into the evidence for reincarnation, and unusual in dealing with a topic—reincarnation—that is a very rare focus in psychiatric discussion, especially in these days of increased interest in neuropsychologic and molecular genetic approaches to complex human behaviors, including psychopathology. Accordingly, background information is warranted to show how this particular book fits into the study of human behavior and, by extension, psychiatry, and how this monograph from Dr. Stevenson’s research career is a product of his lifelong interest in the paranormal—extrasensory experiences and kindred phenomena.

Ian Stevenson was born in Canada, the son of a Scottish lawyer and his wife, who had an interest in psychic phenomena. Stevenson studied medicine at St. Andrews in Scotland and at McGill in Montreal. During the late 1940s, early in his medical career, he did research in psychosomatic medicine at New York Hospital, where Dr. H.G. Wolfe led a group investigating effects of life stress and its concomitant emotion on patients’ organ systems. Because of his discontent with the then current psychosomatic interpretations of his colleagues, Stevenson moved into psychoanalytic training in 1951, graduating from the Washington Psychoanalytic Institute in 1958. In 1957 he became a child psychiatrist at the University of Virginia Hospital at Charlottesville.

Early in his career at Charlottesville he became aware of the lack of scientific method behind Freudian psychoanalytic hypotheses such as the “assertion that a person’s later character depends almost exclusively on the events of infancy” (1) and the lack of interest in other concepts of unconscious mental processes “current in the early 20th Century (including Pierre Janet, Morton Prince, William James, C.G. Jung, and FWH Myers)” (1). These dissatisfactions with the prevailing psychiatric interpretations of personality led Stevenson in the early 1950s to read systematically in the literature of theosophy and psychic research. He became more attracted to and involved with psychic research because of its approach to both spontaneous psychic phenomena and laboratory methods to demonstrate phenomena such as telepathy and clairvoyance. For a variety of reasons Stevenson chose the “study of psychic experience—those that occur spontaneously in everyday life” (1). Stevenson’s approach to the question of reincarnation was to evaluate the evidence for it and devise a protocol for the gathering of further evidence to delineate the phenomena of human behaviors, which suggested that some aspects of human personality might survive apparent death and manifest themselves in the living. Stevenson described his general approach (for which the book under review provides a specific example) as follows:

In the study of spontaneous paranormal phenomena we must usually interview and cross-question informants about events that have happened before we arrive on the scene. In principle, the methods are those that lawyers use in reconstructing a crime and historians use in understanding the past. Once we have the best account possible of the events in question, we consider one by one the alternative explanations and try to eliminate them until only the single most probable one remains. Then we try with further observations to confirm or reject the initially preferred explanation. In addition, we search through series of apparently similar phenomena for recurrent features that may provide clues to causative conditions and processes of occurrence. (1)

After careful review of available phenomena that had suggested the possibility of reincarnation, Stevenson, following the methodology of early psychic researchers (Gurney et al. in 1886 and Myers in 1903), devised a protocol for recovery and evaluation of memories of apparent previous lives, a process Stevenson described in 1977 (2). This paradigm for investigation focused on spontaneous cases suggestive of reincarnation that were described in young children. Why young children? Because young children should be less likely to be exposed to information about life details of a dead individual who is reincarnated.

A brief description of a typical case of the reincarnation type would show the following features: 1) Starting in years 2–4, the child spontaneously narrates details of a previous life. 2) Volume and clarity of statements from the child increase until ages 5–6, when the child talks less about them. 3) By age 8, remarks about previous life generally cease. 4) Unexpected behavior unusual for child but concordant with behavior of deceased person occur, e.g., phobias for guns or special interests and appetites. 5) In many cases the child has a birthmark or congenital deformity that corresponds in location and appearance to fatal wounds on the body of the previous personality. A high number of reincarnated personalities report violent death, which the child alludes to. 6) In some cultures the individual who “reincarnates” predicts his or her next reincarnation and may appear in a dream to the expectant mother of the child to announce an intention to reincarnate in the baby. 7) After the age of 10 these child subjects usually develop normally.

Stevenson has followed this information-gathering protocol since the early days of his worldwide travels to investigate spontaneous cases suggestive of reincarnation, which have been published over the years as separate volumes covering different cultures and a book, now in its second edition, which summarizes his work to the present (3). The present monograph, European Cases of the Reincarnation Type, fits into this series of publications (4–8).

As the author states in his preface to this book, there are three purposes for the present publication: 1) To show that cases of the type described above occur in European cultures, where fewer individuals believe in reincarnation than in Asia. 2) To show that essential features of these cases are similar to those found in Asia, Africa, and Northwest North American native tribes. 3) To show that some of the cases reported provide evidence of information transmitted by means outside of normal communication—extrasensory perception being one
and reincarnation another. Stevenson also hopes that this book will stimulate the reporting and investigating of more cases of this type from Western cultures.

The purposes just described above explain very well the contents of the monograph under review. It is important to the concept of reincarnation or whatever the collection of behaviors and physical manifestations of reincarnation that there should be a substantial element of invariance in their manifestation. Focus on a syndrome of behaviors in childhood, when environmental or verbally acquired information from the culture is a less likely explanation of similarities between the child and the allegedly reincarnated person, is one approach to defining an elusive set of behaviors. Further evidence of invariance is to discover that vastly different cultures appear to share with very similar behavioral and physical manifestations in these phenomena—as in the case of birthmarks symbolizing injuries in the previous personality or the apparent high incidence of violent unexpected death in many of those who are reincarnated. This particular book discusses in great detail cases suggestive of reincarnation in Western European cultures and points out features similar to those reported from Asia, Africa, and North American Indian tribes. The similarities of cases from culture to culture suggest some organizing process—Stevenson proposes a paranormal process such as reincarnation or extrasensory perception. The quality and amount of information given by the child regarding the previous personality and its independence from usual cultural and environmental sources of this information are essential to the argument that the phenomena are uniquely caused—not coincidence and not by information provided by adults talking about a deceased individual.

In the general discussion of the book (p. 252), Stevenson states that the European cases presented in this book are “much weaker in evidence of a paranormal process than the cases found in Asia, especially in India and Sri Lanka.” He goes on to quote a study that examined 799 cases from six different countries and found that the European cases showed much lower scores for presence of a paranormal process involvement in the information available to the child.

The intensive study of these children has revealed several provocative findings, such as the apparent increase in death due to violence in those who reincarnate and the startling correspondence found between birth marks on the child and similar marks or distinguishing features present on the body of the reincarnated personality during their lifetime, such as wounds, injuries, and other stigmata. These physical types of findings are detailed in Where Reincarnation and Biology Intersect (9), but examples are present in the current monograph. Coincidences like these remind us that there remains a host of phenomena to be described and studied, as Dr. Stevenson has done for most of his research lifetime. The present book provides an introduction to an exciting range of such phenomena and furnishes an inspiring example of application of a painstaking protocol to sift facts from fancy. Ian Stevenson, now in his ninth decade, is looking forward to publishing a monograph on reincarnation type cases from the United States (personal communication), and one wishes him both the time and energy to continue this task.

References
4. Stevenson I: Twenty Cases Suggestive of Reincarnation. Charlottesville, University of Virginia Press, 1974
5. Stevenson I: Cases of the Reincarnation Type, vol I: Ten Cases in India. Charlottesville, University of Virginia Press, 1975
7. Stevenson I: Cases of the Reincarnation Type, vol III: Twelve Cases in Lebanon and Turkey. Charlottesville, University of Virginia Press, 1983

REMI J. CADORET, M.D.
Iowa City, Iowa

Psychoanalysis


We are very indebted to Martin Bergmann and the Other Press for the second in a series of conferences sponsored by the Psychoanalytic Research and Development Fund of New York. The first of these was titled The Hartmann Era (1), and the new one is of just as high a quality. What makes it even more remarkable is that although Bergmann presents a long and clear expository essay at the beginning and runs the symposium that follows later with a firm hand, this meeting was held on his 90th birthday!

The book itself is divided into four parts and is well done. The first part reviews a whole series of famous dissidents in the history of psychoanalysis, including Adler, Jung, Rank, Ferenczi, Horney, Reich, Rado, Klein, Fromm, Lacan, Kohut, Fairbairn, Winnicott, Bowlby, and other important individuals and symposia where controversy flared up. Bergmann manages to do this in 109 pages.

The second part contains contributions prepared by a number of prominent psychoanalysts, some of them my favorite authors. The contributions vary in quality and often contain a great deal of high-level speculation. Some authors write directly to the point, while others digress. There are seven prepared contributions, plus yet another article by Bergmann.

The third part consists of an open meeting held on February 14, 2003, Bergmann’s 90th birthday. In this section the individuals who offered prepared contributions in the second section discuss with each other the problems of dissidence...
and controversy in the history of psychoanalysis. This discussion was apparently taped and edited and is presented here.

The book remains good to the very end because there is a fourth section that consists of an essay by Wallerstein, who could not attend the meeting because of a sudden illness in his family. He was allowed to see the transcript of part 3 and asked to make a response to the symposium.

There is so much material packed into this book on the subject of psychoanalysis, both theory and practice, that it almost constitutes a complete textbook on current and previous important issues in the field. The audience for the book would consist primarily of psychoanalysts and psychoanalysts in training as well as any psychiatrists who are interested in the history of psychoanalysis or in the many conflicting psychoanalytic theories that prevail today.

Bergmann asks, “Can we learn something about the nature of psychoanalysis itself from the dissidents?” (p. 3), and he hopes that this set of papers and the subsequent discussion will be helpful to students who have to “deal with the fact that psychoanalysis is no longer a monolithic movement” (p. 3). Bergmann points out that especially the early dissidents were told that they simply had “resistance” and, “It gave the psychoanalytic group a sense of coherence and contempt for the dissidents” (p. 5). He then devotes the rest of his opening essay to a review of the story of each of the dissidents. He believes that they were all charismatic, “because the capacity to surround oneself with admiring students is a sine qua non for a successful dissident” (p. 65), and he relates how the dissident theories gradually developed a greater and greater rejection of Freud’s entire edifice.

He claims that except for Jung’s psychology, “no other school of psychoanalysis represents as radical a break from Freud’s theory, philosophy, and technique of treatment as does Heinz Kohut’s self psychology” (p. 73). He sees an important root of dissidence in a training analysis when “the hostility toward one’s own analyst, projected on Freud, is greater than the gratitude for what the analyst, with all her or his shortcomings, did achieve” (p. 78), and usually this appears after the training analysis is finished and a self-analysis is carried on. The prime examples of this, according to Bergmann, are the two so-called analyses of Mr. Z: the first is a report of Kohut’s analysis and the second is a report of Kohut’s subsequent self-analysis. It is clear, as Bergmann points out, that “classical psychoanalysis came to an end with World War II” (p. 92). There are no more classical analysts, and papers in the current literature that tend to attack the so-called classical analyst are setting up a straw person.

There is a lot of discussion in the prepared presentations in part 2 about the multiple conceptions of psychoanalysis, the plurality of theories, and whether this is an acceptable situation or one that needs to be transcended. Green presents a very nice review of why psychoanalysis has fragmented. He concludes that the compromise in the field today, a “pretense of tolerance, search for willy-nilly common sharings that are not very convincing and appear as life jackets to avoid sinking” (p. 126), prevents the collapse of the entire field. Kernberg, one of the most active participants in the meeting, repeats his already published arguments that psychoanalysis has created a serious problem for itself “by accepting, indeed enacting, its reputation as isolationist, elitist, and biased against empirical research” (p. 136). Scharff maintains that there is a major difference between the theory of Melanie Klein and British object relations theory, although they are frequently lumped together. Wallerstein brings up the book review by Bachant and Richards (2) in which they divided what they called “psychoanalytic metatheories” into five groups of adherents: 1) those who wish for a common ground, 2) those who advocate a multimodal approach to the phenomena of psychoanalysis, 3) those who want one total composite psychoanalytic theory, 4) a group that wishes to throw out metapsychology altogether, and 5) those who see the variant metapsychologies as falling into either drive-structural theory or object relational theory. This latter constitutes a dichotomy that contrasts the human ego as a pleasure-seeking organ with that of the human ego as an object-seeking organ, pursuing human relatedness.

The issue of whether psychoanalysis is a one-person psychology and, more recently, a two-person psychology is discussed by a number of the authors. The more recent two-person psychology approach focuses on the interactional and intersubjective quality of the analytic encounter. Bergmann, in his prepared presentation in part 2, concludes that dissidence is so unwelcome because it affects “the idealization of psychoanalysis that many need to continue the profession” (p. 251).

Part 3 is probably the most long-winded and somewhat repetitive section of the book, but the topics covered are extremely important to anyone interested in psychoanalysis. There is a great deal of discussion about the vehemence with which dissidents are attacked, ostracized, and, often, asked to leave. There is a lot of discussion of Ferenczi, who is often called the mother of psychoanalysis just as Freud was the father. Blum points out, “It was Ferenczi who really began a new focus on the analyst’s mind, the analyst influencing the analytic process with bilateral, unconscious communications” (p. 296). Ferenczi’s views are being taken increasingly seriously at the present time, although when they were first put forward he was ostracized and even accused of being mentally and physically ill.

Several of the participants point out that the attempt to do psychoanalytic treatment with the now ubiquitous borderline patient has forced analysts to reconsider their theories and has itself by necessity spawned a number of new theoretical and technical ideas modifying the standard techniques. Blum warns,

> We have a kind of fragmentation or dissolution of psychoanalysis with all kinds of burgeoning theories and “every flower should blossom.” Every theory can be viewed as potentially valid or equally valid. And that’s a danger on the other side of monolithic orthodoxy. (p. 340)

This is not theoretical pluralism with awareness and investigation of differences but is what is now known as eclecticism, a stance that all of the authors oppose.

Kernberg warns of the danger of oversimplification and reductionism every time some new discovery in neurobiology is brought into psychoanalytic theory, even though psychoanalysts are good friends with neuroscientists. Blum points out that “institutes are spending less and less time teaching Freud—reading Freud’s papers, discussing those papers,
speaking of the core of psychoanalysis” (p. 348), and the participants agree. Green labels as a core concept of psychoanalysis that the unconscious leads us to the drives and is a product of them. He concludes that dissidents can be identified in the things that they refuse: “We can refuse it massively like Adler and Jung—or subtly with object relationships, even more with attachment theory, and even more with intersubjective theory” (p. 354).

In the final, short section of the book, Wallerstein depicts the core of psychoanalysis as the concepts of the unconscious—hidden workings of the mind that powerfully mold our thoughts, feelings, and behaviors—and of psychic determinism—“What comes before helping determine what comes after” (p. 363). He points out that the various theories are not yet amenable to scientific testing and argues that clinical experience will never lead to definitive answers. He makes a strong plea for empirical psychoanalytic research.

References

RICHARD D. CHESSICK, M.D., Ph.D.
Evanston, Ill.


The marriage between psychiatry and psychoanalysis has been increasingly strained during the past few decades, as psychiatry has become more and more biologically oriented and psychoanalytic treatment is seen as incompatible with the “hurry up” emphasis of managed care. Although threatened, the marriage has survived because patients treated with medicines still need to talk to someone. Psychiatrists today must rely on the language of psychoanalysis to have conversations with and about patients, to roughly estimate the extent of clinical change whatever its cause, and to understand what is going on in their patients’ minds. Psychiatrist readers of the Journal who are devoted to the third of these concerns—understanding—will find this book stimulating and valuable. It is not easy reading but contains many vivid case illustrations of treatments that were failing for no obvious reason, a dismaying situation all too familiar to all of us. Readers from all perspectives will appreciate the clarity with which Goldberg depicts the unexpected treatment twists and turns demonstrating remedies for a variety of clinical impasses.

Supported by his many years of clinical experience and his lucid understanding of previous mental models that depend on evidence, validity, or explanatory power, Goldberg proceeds to identify in detail the therapeutic stance he finds most rationally justified. I will try to convey his approach through a brief discussion of one dualism that he decries, the split between the domains of self and object, which influences the ways therapeutic interventions are conceptualized and used during psychoanalysis and psychotherapy.

When we think about our minds, we do so in images that have spatial dimensions. For instance, there is the image of the iceberg, whose part above the water line represents the area of consciousness and whose large invisible portion hidden below represents the unconscious. Or there is the image depicting the relationship of ego to id as that of a rider (ego) attempting to dominate a strong wild horse (the unconscious id). The mind is conceptualized as a sort of encapsulated bag with the subjective portion inside and the external objective world outside. The upper region in the bag is the realm of consciousness and is split off from the lower region, the world of the unconscious. Or if we imagine one bag to represent myself as the subject, we can imagine that you (the nonself other) is a different bag designated as an object. When Goldberg inspects the model of child development of the object relations school, he notes that for them, the subject’s bag contains a psychological world of representations that is distinct and different from the real external perceptual world. According to that theory, the universe of perceptions and people in the object bag can be taken into the subject bag (internalized) throughout a person’s lifetime.

Goldberg argues that a quite different concept of the mind emerges if the mind is thought of more as a psychological activity than an interaction of spatially located containers. The idea of the mind as activity almost completely dissolves the polarizing split between the self and the object that persists in object relations theory and is fully compatible with the analytic findings of Heinz Kohut, the philosophical position of Martin Heidegger (1), as described in an appendix of the book, and most major philosophers today (e.g., John Searle [2]).

In this line of thinking, the developing child’s relations with the mother and father can be viewed as a series of influential meaning-making emotional transactions or enactments (conscious and unconscious), all of which will determine the coloration of the adult personality and its vicissitudes. Such a perspective does not see maturation as a progressive series of internalizations during which the real perceptual and emotional world of significant people set up camp in the subject’s mind but, rather, as the growth of the child’s mind-body-self that can never be separated from its sustaining love objects. Goldberg writes that “a self extends beyond an individual’s skin, and the mind extends beyond an individual’s skull” (p. 124).

Interpretation during analysis and psychodynamic therapy now becomes primarily a means to detect, review, and modify meanings. “Meaning” refers to the myriad back-and-forth emotional and verbal transactions between patient and therapist that have relevance for symptoms, maladaptive defenses, and behaviors, as well as for personality strengths. However, Goldberg maintains—contrary to the late Stephen Mitchell (3)—that the analyst is in a privileged, usually optimal position to identify meaning and to recognize reality dis-
tortions. "All patients need both to be understood...but they need to have such understanding explained to them" (p. 194). The therapist, too, may never be sure of the truth, while mired in profound misunderstanding of the patient's communications for sustained time periods. Therefore, the therapist must continue to “persistently puzzle” with the patient searching for new comprehension of both past and present sources of patients' suffering. Derrida, the enemy of finality in attributing meaning, while not cited in the book, would approve of this stance were he still alive.

To act according to the precept of placing the welfare of the patient above their own self-interest, therapists may temporarily have to break rules promulgated by followers of doctrinaire schools of psychological theory. The therapeutic space must not be one where “anything goes,” but one in which “everything matters,” because the therapist's activity must always have the goal of detecting meaning and enhancing understanding. (Exploitation of the patient is of course always unethical.)

Viewed from this perspective, the mind is not identical with the brain because activities of the brain or its organic deficits only rarely correspond at all with the mind's activities or with its aberrations that characterize psychopathology. (An exception may be seen, the author indicates, in cases of antisocial personality disorder, which may have an association with reduced prefrontal gray matter of the brain. Such rare patients cannot profit from "talking therapies.") Goldberg opposes any "hybrid mixture of brain and person" that we try to cure. This therapeutic posture might support some brain-mind dualism, in disagreement with Kandel's conclusion (“Psychoanalysis is, in the best sense, a part of biology”), which appeared in the Journal as a response to the fractious debate (4) regarding his theories (5).

Freud used the interpretive method to lift repressions or to move emotional content out from the dark unconscious regions of the mind across the repression barrier to the bright conscious world where rational decisions could be made. Most of us, however, have long suspected that the idea of “where Id was let Ego be” is not a good explanation of why and how Freud's “talking cure” works. Goldberg finds Freud's treatment approach often inadequate and feels it produces inevitable misunderstanding, a term that is found in the title of the book. Interpretation leading to insight by patients into childhood sources of their psychopathology used to be considered the single most important curative factor in psychoanalytic treatment. This may still be so, but only if the concept is updated and clarified in Goldberg's manner.

The author's conceptualization of interpretation leading to insight uses the image of a “hermeneutic circle,” the repeated cycle of understanding-misunderstanding-understanding, kept in motion by new reality developments or frustrations that challenge the most recent understandings. It begins with the therapist's initial understanding of the patient's symptoms and neurotic suffering and proceeds with both of them beginning to recognize remaining misunderstandings. Through interpretation, a new understanding may occur with some insight into the previous defective psychological configurations, which then leads to recognition of remaining or novel misunderstandings. The cycle may have to continue even after the ending of treatment through mobilization of the patient's autoanalytic function or further treatment. Dur-

ing almost all therapy outside of formal analysis, interpretation must be liberally melded with other therapeutic activities (i.e., soothing, clarification, abreaction, suggestion, and self-esteem maintenance) described by the author in his first major book, Models of the Mind (6), written with John Gedo.

This last and I think best of the 26 books with which Goldberg has been associated illustrates the many difficulties and dead ends that analytic therapists and their patients must surmount while traveling around segments of the "hermeneutic circle." The book’s mustering of evidence is always firmly rooted in the author's lifetime of listening to many, many patients with empathy. He seems to retain a remarkable objectivity, but he often expresses his ideas with a self-deprecatory, ironic humor. I had a few uproarious laughs while reading chapter 12, "Me and Max," which is a terrific introduction for students seeking to master the most critical elements of doing competent psychotherapy.

Psychiatrists who only write medicine prescriptions and have little or no interest in talking with patients should ignore this book. Those of us who continue to believe that sensitive emotional communication is crucial for understanding and facilitating clinical change (with or without the use of psychoactive agents) must give this book considerable careful attention. In grappling with these matters Goldberg defines the process of insight more extensively than any writer since Freud. In doing so he is no longer fulfilling the mandate to apply and amplify the brilliant discoveries of Heinz Kohut but truly becomes "his own man," capable of fundamental original contributions to depth psychology.

References
4. Olds DD; Rifkin A; Moore DP; Berman LH; Shevrin H; Hierholzer R; Alarcon RD: Kandel's challenge to psychoanalysts and reply of ER Kandel (letters). Am J Psychiatry 1999; 156:662–666

PETER BARGLOW, M.D.
Berkeley, Calif.


This is an excellent text. Dr. Horowitz displays impressive skills as a theoretician, a clinician, and an author. His language moves flawlessly from abstract, philosophical, and theoretical constructs, to complex statistical formulae, to concrete, substantive case examples. We read of the integrative forces of social, cognitive, motivational, and affective systems...
of human experience as they affect the development of personality and psychopathology.

Early chapters of the book outline an interpersonal approach to psychotherapy, with a focus on interpersonal motives in human experience. In particular, the concepts of “communal” and “agentic” motives are described as basic to interpersonal theory. Extended discussions of different disorders from an interpersonal perspective are then presented in later chapters.

A visual representation of interpersonal motives is provided throughout the book in the form of a graph whose characteristics evolve with the author’s developing discussion. This graph illustrates the relationship of different interpersonal motives to one another, to research findings, and to different mental disorders. As a pictorial image of interpersonal functioning, the graph provides the reader with a practical and unambiguous cognitive tool to grasp the main thesis of the text.

The author also uses literary and visual art to illustrate his themes. He describes a scene from Tennessee Williams’ A Streetcar Named Desire as reflective of frustrated goal achievement. He also portrays the life of the artist Edvard Munch as an example of schizoid personality disorder, noting the interpersonal terror of engulfment that emerges even in the artist’s paintings. These illustrations from the literary and visual art world only add to the readability and interest of the text.

Although informative and well-written, the text’s commentary on attributional style could be extended, particularly the discussion regarding the emotions of guilt and shame. Recent research on these emotions has noted important distinctions between them that the text does not consider. A fuller treatment of these issues would improve this section of the work and further highlight important interpersonal implications of attributional style.

The book could also benefit from greater clarification of the interpersonal motives in the experience of schizophrenia. The author does a nice job of describing research regarding communication deviance, expressed emotion, and affective style in the families of people with schizophrenia. Greater theoretical clarification as to the manner in which these might involve interpersonal motives, whether “communal” or “agentic,” would continue to strengthen the theme of the book.

Overall, Dr. Horowitz has produced a fine text and has succeeded in the attempt to illustrate the power and significance of interpersonal motives in the development and maintenance of psychopathology. His book would be an excellent choice for the graduate student training to become a psychotherapist. It would also make a solid and important addition to the library of the seasoned clinician or the professional researcher.

MARcia WEBB, Ph.D.
Seattle, Wash.


One evening many years ago, when I was a second-postgraduate-year resident on call for inpatient admissions, I looked up from my chart at the nurses’ station to see a young woman being dragged onto the unit by her boyfriend and his mother. I use the word “dragged” quite literally: the young woman—I’ll call her Laura—was held up on either side by her two companions as they moved her down the hallway. Sometimes her feet made small shuffling movements, but mainly her legs and arms were rigid and her face was in a fixed grimace. Laura was only 20 years old and had given birth to her first baby 2 weeks earlier. Over the past few days she had become mute and akinetic to the point that she was no longer eating or drinking. This was her first hospital admission, although she had seen a psychiatrist in the past for an episode of agitation where she had run into traffic and started disrobing. She carried no specific diagnosis and had been treated briefly with antipsychotic medications.

As a young resident, I was fascinated by Laura. She was close to my own age, an attractive person with shoulder-length dark hair (now unkempt) who had won medals on her high school track team. When I stood near her to examine her—she could not sit down—she did not respond to any of my questions, but kept the same fixed stare, the same odd grimace, although from time to time I thought that I saw a twitch around her mouth when I made an especially kind comment to her.

This was the mid-1980s, an era of not only 4-week hospital stays but psychodynamically oriented nursing staff, who informed me that Laura’s mutism, unresponsiveness, posturing, waxy flexibility, and stereotypic hand movements reflected her murderous ambivalence toward her newborn daughter. Luckily, my senior resident was Allan Reiss (now a prominent behavioral neurogenetic and neuroimaging researcher), and my attending was Irvin Yalom, who always encouraged his trainees to think and work independently. I started to read as much as I could about catatonia, and after a few unsuccessful attempts with other medications, Allan and I decided to try lorazepam. After a single dose of 2 mg, Laura’s limbs began to move, her face relaxed, and she talked! Once she responded to the benzodiazepine, we added lithium carbonate, and within 2 weeks Laura returned home to care for her baby and Allan and I wrote the case up for publication (1).

Flash forward almost two decades: I am holding the 2003 book Catatonia: A Clinician’s Guide to Diagnosis and Treatment. As I start to leaf through it, I ask myself, How germane is this topic to the average practicing psychiatrist? At the same time, memories of Laura’s case come flooding back to me; she is reflected in almost every single chapter. A few minutes later, I am persuaded: Drs. Fink and Taylor have marshaled an impressive amount of both scholarly and clinical evidence in support of the following axioms:

1. Catatonic features are not rare phenomena but occur in about 10% of acutely hospitalized psychiatric and neuropsychiatric patients.
2. Catatonic features are easily recognizable, if one is trained to recognize them.
3. Catatonia is concomitant with a range of brain disorders, most commonly (as in Laura’s case) serious mood disorders.
4. Catatonia is highly responsive to treatment but is tragically underrecognized and thus undertreated, causing
great (and unnecessary) morbidity and suffering for patients, as well as high costs to society.

Indeed, the authors’ goal is, through this book, to raise an awareness of catatonic syndromes in psychiatrists and neurologists who work in acute care settings (inpatient units, emergency rooms, consultation services).

Certainly, they have succeeded. Catatonia is a gem of a book. It is written in a clear, accessible, yet authoritative style by two clinician-scientists who have dedicated their careers to improving our recognition, understanding, and treatment of this syndrome. The book is exceptionally well organized, with bold, succinct chapters and sections (“The Signs of Catatonia Are Identifiable,” “Features Demonstrated in Examination,” “Conditions in Which Catatonia Is Expressed”). The authors’ command of the literature—from historically significant citations on clinical features such as Kahlbaum (1874), Leonhard (1942), and Abrams and Taylor (1976), to reviews of dynamic neuroimaging studies and neurotransmitter hypotheses—is truly impressive and sets a new standard for the field.

Despite the high level of the scholarship, this book most definitely is not aimed at academic “talking heads.” There are multiple patient vignettes in each chapter—more than 50 in all. The appendix contains a description of how to examine a patient for catatonic features as well as a catatonia rating scale. A complete differential diagnosis of catatonic syndromes is presented, including conditions that may be mistaken for catatonia (stupor, obsessive-compulsive disorder, stiff-person syndrome). The chapter on treatment contains multiple tables summarizing such clinically important facts as prognostic factors, the management steps for malignant catatonia (neuroleptic malignant syndrome), and factors affecting relapse.

Despite this impressive clinical knowledge base, the neuroscience of catatonia remains nonspecific and somewhat unsatisfying. The authors do a masterful job of synthesizing a large and complex literature and propose that catatonia, a disorder of motor system output, arises from disconnection between frontal circuitry and perceptual-integrating brain systems. Although undoubtedly accurate in its broad terms, this model does not explain why some patients show verbigeration (which must involve Broca’s area), others show ambi-tendency (supplementary motor area?), and still others show impulsivity or combative-ness (orbitofrontal cortex?). Likewise, the authors never fully explain how their model of catatonia contrasts with a distinct clinical disorder of complex motor output such as Tourette’s syndrome.

None of this matters to Laura. She called me this morning to tell me proudly that her daughter (who began working full-time at age 16), bought a condominium last month single-handedly, at the age of 19! Laura herself has been stable for more than two decades on a regimen of lithium, lorazepam, and an antipsychotic, has raised two daughters, and has never needed another hospitalization. With this book in the hands of acute care clinicians, there will be many more outcomes like hers.

Reference

SOPHIA VINOGRADOV, M.D.
San Francisco, Calif.


Dante Cicchetti from the University of Rochester and Elaine Walker of Emory University provide an extensive collection of chapters that present established as well as hypothetical ideas on the ultimate biological view of human behavior and psychopathology. This book is an ambitious effort to explain and weave together the many different disciplines of neuroscience, neural development, and psychopathology to create an understanding of the intimate relationships among these complex areas. The editors are successful in this effort, in my opinion.

As one would expect, this is not a quick read, and it is not going to be an essential text on the shelves of many clinicians. In fact, its real-life clinical utility will probably be minimal at best. However, for those with a serious interest in understanding the proposed biological views of normal and aberrant development of the nervous system and its intimate connections to other integrated systems of the body and to the ultimate expression of what is diagnosed behaviorally, this can be a fascinating book.

The book itself developed out of a series of papers from conferences on neurodevelopment, neurosciences, and developmental psychopathology. Fifty international experts in their fields have contributed 21 chapters divided into four sections. Part 1 discusses human neurodevelopmental processes in the prenatal, perinatal, and postnatal periods and the attendant risks for adult mental disorders. Part 2 focuses on animal models of development of the nervous system as they relate to psychopathology. Part 3 includes 16 chapters involving developmental processes that illustrate high-risk conditions and mental disorders, including personality disorders, the neurobiology of mood disorders in children and adolescents, traumatic stress disorders, and psychotic disorders.

Briefly looking at three of the chapters will help to understand the broader focus of the book. The first chapter, “Principles of Neurobehavioral Teratology,” presents a brief overview of current teratology and 10 basic principles of research in this area. These principles include possible mechanisms of effect, likely agents, exposure, dose-response relationships, and timing effects. The chapter ends, as many do, with a section titled Future Directions. Chapter 9, “Early Orbitofrontal-Limbic Dysfunction and Autism,” discusses aspects of social and other behaviors, links them to neural structures and circuits, and proposes related deficits in autisticform behaviors. Finally, chapter 21, “Neurohormonal Aspects of the Development of Psychotic Disorders,” is a broadly written chapter that includes, among other topics, important background in-

As someone whose interests span clinical, research, and policy issues in behavioral health services, I am often called upon to meet with students and trainees who are interested in policy and administration. I also codirect a seminar series on social and community psychiatry for psychiatric residents and frequently present lectures to medical or public health students on the “mental health system” and related policy issues. As such, I was looking forward to identifying a single text that I could share with interested students and use as a backbone for the seminar. Mental Health Services: A Public Health Perspective has a number of positive attributes, but, overall, it does not have the coverage and focus to meet that need.

The text has 20 chapters organized into four parts: Service Delivery Issues, Selected Populations at Risk (children and adolescents, adults, older adults, substance abuse), Special Issues, and Managing Mental Health Systems. The book’s greatest strengths are quality of the contributors (all are major leaders in each of the domains selected) and the broad range of disciplines represented. The greatest weaknesses are the lack of a cohesive vision and insufficient linkages among the chapters. As a result, a student would not really get the big picture. To try to remedy this often encountered problem in multiauthored texts, the editors had most authors include a section at the end of each chapter titled Implications for Mental Health Services. However, no consistent framework is discernible for how the authors determined the implications, and the sections often seem tacked on.

Another problem that limits the applicability to non-mental-health disciplines is that there is very little on the nature of treatment for behavioral conditions and their effectiveness (especially for those who are not seriously mentally ill). Also, some clinical vignettes might help nonclinicians get a more concrete sense of how practices and policies affect individuals. The epidemiology chapters are all excellent reviews of the literature, but they provide very little connection to services and policy issues per se. As a text for nonepidemiologists, the book might have included a section on how these surveys are typically conducted, warts and all. There is a chapter titled “Co-Occurring Disorders,” but most other chapters ignore this phenomenon. For example, the chapter on criminal justice says almost nothing about substance abuse.

There are some outstanding chapters that I would certainly offer to trainees with specific interest: “Mental Health Disability Law,” “Policy and Services Delivery,” “Child Mental Health Policy,” “Mental Health Policy and Aging,” and the chapter on evaluation are all comprehensive and superb.

Finally, despite the subtitle, A Public Health Perspective, that perspective does not appear to be well represented in the text. For example, epidemiologic data strongly demonstrate that the largest proportion of mental health care is provided in the primary care sector, yet there is no chapter focused on that issue.

HAROLD ALAN PINCUS, M.D.
Omaha, Neb.


This book describes the authors’ approach to the creation, implementation, and evaluation of mental health services in schools among diverse cultures. It begins by pointing out that 20% of children and adolescents in the United States have symptoms of a diagnosable mental disorder and that childhood neuropsychiatric disorders will rise proportionately by more than 50% internationally by the year 2020 to become one of the five most common causes of childhood morbidity, mortality, and disability in the world. The authors describe the pivotal role that schools can have in many different cultures and how the integration of mental health services in schools can provide the starting point toward the prevention and treatment of mental health problems in children and adolescents.

The book presents the participatory culture-specific intervention model, describes its foundations, and introduces it with illustrations of the authors’ work in the field; it also describes detailed procedures for implementing the model and discusses future directions in research along with questions raised by the model. The participatory culture-specific intervention model is an interdisciplinary model based on applied anthropology and school psychology. The authors emphasize the participatory nature of the model throughout the book and the importance of engaging the stakeholders (children, families, teachers, school staff, the larger community, etc.) as partners in the process. They also stress the importance of the application of ethnographic research methods and an ecological perspective as part of the model.

The authors illustrate the participatory culture-specific intervention model with their work from the Sri Lanka Mental Health Project, where they developed school-based mental health services in the urban community of Kandy. It was particularly interesting to read about the struggles and re-

MARK H. FLEISHER, M.D.
Omaha, Neb.

http://ajp.psychiatryonline.org

Am J Psychiatry 162:4, April 2005
wards of the project taking shape over the years and the interdisciplinary nature of it, with the participation of sociologists, government officials, psychiatrists, anthropologists, and other specialists from both Sri Lanka and the United States.

The first six of the 11 phases of the participatory culture-specific intervention model are devoted to research and include developing the model, building up relationships with the community, and working with the cultural brokers. The following five phases deal with issues of intervention and include steps such as the implementation of the program, evaluation of the program, and continuation and extension of the program. There is also a partnership component that is central to the model and that encompasses all the other phases. The book describes each of the phases in detail with numerous examples from their own work along with that of others in the field of school psychology. The authors stress the critical role of the cultural broker in forming partnerships with the stakeholders and how the broker serves as an expert and interpreter of the culture and a liaison between the researchers and the community.

This book is very thorough and convincingly demonstrates the usefulness of the participatory culture-specific intervention model. It is written clearly and contains numerous examples, appendices, and references. The book is primarily intended for school psychologists, but psychiatrists and other mental health professionals working in the school system and in the implementation of school-based mental health services will also find it rewarding and thought-provoking.

Israel Katz, M.D.
Albert C. Gaw, M.D.
San Francisco, Calif.


This manual for research and outcome measures in health and human services is an important topic today for clinicians, teachers, researchers, and administrators in public health. Its main objective was “to bridge and augment health and human services with scientific research inquiry.” It is intended to affect clinical practice, mainly through improving research and program evaluation, but it reaches for too many audiences—health care workers, administrators, and health services researchers. There are more than 100 chapters and 1,000 pages.

The first few chapters in section 1 provide an overview for the book (e.g., background, how to implement procedures, performance indicators). Section 2 reviews research ethics and grant applications and would be of use to administrators, junior faculty, or driven trainees embarking on a research career. Clinicians may use it for reference. Section 3 is squarely focused on the title with regard to how evidence-based issues apply to diagnosis, interventions, and outcomes. It appears more useful for researchers and administrators than clinicians. Some childhood and adult mental health disorders are covered as exemplars (e.g., attention deficit hyperactivity disorder, depression, and anxiety). For some topics, only a single evidence-based approach is addressed (e.g., cognitive behavior therapy for posttraumatic stress disorder).

Section 4 is well organized and covers epidemiology and public health research. Once again, certain populations are targeted as exemplars (e.g., Northern Plains Indians). Sections 5 and 6 focus on conceptualization and operationalization of clinical assessment, providing a brief overview of measures, mainly those used in nonmedicine sectors. The overview presented here is not as well organized, detailed, or clinically applicable for medicine and psychiatric settings as can be found in other books (1).

Section 7 is very useful for program evaluation, and section 10 complements it by addressing continued quality improvement—both are helpful for administrators and/or researchers. Sections 8 and 9 provide qualitative and quantitative research exemplars, respectively. The exemplars are generally public health populations (e.g., HIV, cancer prevention, drug courts). The sections highlight some clinical issues in the process of demonstrating how one would administratively evaluate populations and settings. Things to do are well outlined.

We have several suggestions for the next edition. First, we would group the chapters on how to conduct health services research as one section of the book, with applications of these techniques in another portion. There is, for example, a wonderful chapter on constructing validated scales that is buried after a chapter about measuring fatherhood propensities. Second, our sense is that the chapter authors were encouraged to provide a topic review and then to discuss their own research. The weakness in this technique is that it does not provide a sense of where this research falls in a continuum of research methods or link it to other related research methods. The examples override the approach, but the examples make the techniques real. Third, the chapters could benefit from an outline, learner objectives, and a standardized format. The headings are easy to see, although the print could be bigger. The tables could be more helpful. Fourth, contributions should be considered from health services researchers in pediatrics, internal medicine, and medical informatics, who have significantly contributed to advances in qualitative and quantitative health services research methods. This would include a discussion of issues in large database research, educational research techniques that affect patients, or a variety of other standard health services research techniques. However, some community-based interventions are discussed. Fifth, the authors only briefly discuss problems that they ran into during their research and common mistakes that were made. This would have been very valuable to other researchers and public policy workers. Finally, the link to applied clinical practice is not very strong, nor is the assertion of the authors that cost-savings can be achieved with these measures. Health services research and evidence-based practice are costly to conduct, but they assure accountability and allow effective programs to flourish while eliminating pet projects without evidence of benefit.

This is a reasonable reference book. The text, intended as comprehensive, is actually a compendium of interesting chapters that address research, grants, and the administrative foundation of evidence-based medicine in public health settings. In the attempt to enhance practice-based medicine, it is a good hands-on, “how-we-did-it” book for researchers, ad-
ministrators, and clinicians charged with evaluating or initiating clinical programs. Academic physicians working in public health might find this useful, but most physicians and clinicians will not find it immediately applicable.

Reference


DONALD M. HILTY, M.D.
MALATHI SRINIVASAN, M.D.
Sacramento, Calif.

Corrections

In the article "fMRI of Response to Nicotine During a Smooth Pursuit Eye Movement Task in Schizophrenia” by Jason R. Tregellas, Ph.D., et al., in the February issue (Am J Psychiatry 2005; 162:391–393), the first two sentences in the footnote to Figure 1 (p. 392) should read as follows: Figure shows differences in blood-oxygen-level-dependent (BOLD) responses between before and after nicotine compared with differences in BOLD responses before and after placebo; t values based on a second-level paired t test. Areas where nicotine showed less activation during the smooth pursuit eye movement task following nicotine administration compared with placebo were the right hippocampus ($x=30, y=-15, z=-14$) and bilateral parietal eye fields ($right x=36, y=-62, z=36$; $left x=-42, y=-62, z=34$).

In the article “A Multidimensional Meta-Analysis of Psychotherapy for PTSD” (Am J Psychiatry 2005; 162: 214–227) by Rebekah Bradley, Ph.D., et al., the improvement effect size data for cognitive processing and exposure for the study by Resick et al. were transposed. For cognitive processing, the improvement effect sizes should be 2.95 for pre- versus posttreatment and 2.75 for treatment versus control condition. The respective effect sizes for exposure should be 2.16 and 1.92.

An error appeared in the article “A Videotape Intervention to Enhance the Informed Consent Process for Medical and Psychiatric Treatment Research” by Donna A. Wirshing, M.D., et al. in the January issue (Am J Psychiatry 2005; 162: 186–188). In the first paragraph of the Conclusions section (p. 188), the fourth line should read as follows: Current symptom severity did not seem to matter, but the participating clinical trials involved maintenance treatments of stabilized schizophrenia patients.

In the article “Comparison of Cortical 5-HT2A Receptor Binding in Bulimia Nervosa Patients and Healthy Volunteers” (Am J Psychiatry 2004; 161:1916–1918) by Ingeborg Goethals, M.D., et al., the authors inappropriately acknowledged Koen Van Laere, M.D., Ph.D., Dr.Sc., for comments on the manuscript prior to publication. Dr. Van Laere never received a manuscript version of the study nor did he submit comments.

Reprints are not available; however, Book Forum reviews can be downloaded at http://ajp.psychiatryonline.org.