Editorials:

Kurt C. Stange
In This Issue: Subtle Clinical Policy

Theodore G. Ganiats
Should We Screen Patients With Viral Symptoms for HIV Disease?

Original Research:

Andrew Coco
The Cost-Effectiveness of Expanded Testing for Primary HIV Infection

Andrew Coco and Emily Kleinhans
Prevalence of Primary HIV Infection in Symptomatic Ambulatory Patients

Fiona M. Walter and Jon Emery
‘Coming Down the Line’—Patients’ Understanding of Their Family History of Common Chronic Disease

Ronald M. Epstein, Peter Franks, Cleveland G. Shields, Sean C. Meldrum, Katherine N. Miller, Thomas L. Campbell, and Kevin Fiscella
Patient-Centered Communication and Diagnostic Testing

Robin L. Kruse, David R. Mehr, Jenny T. van der Steen, Marcel E. Ooms, Richard W. Madsen, Ashley K. Sherman, Ralph B. D’Agostino, Gerrit van der Wal, and Miel W. Ribbe
Antibiotic Treatment and Survival of Nursing Home Patients With Lower Respiratory Tract Infection: A Cross-National Analysis

**Delivery of Clinical Preventive Services in Family Medicine Offices**

**Methodology:**

Cleveland G. Shields, Peter Franks, Kevin Fiscella, Sean Meldrum, and Ronald M. Epstein

**Rochester Participatory Decision-Making Scale (RPAD): Reliability and Validity**

John Scott, Alfred Tallia, Jesse C. Crosson, A. John Orzano, Christine Stroebel, Barbara DiCicco-Bloom, Dena O’Malley, Eric Shaw, and Benjamin Crabtree

**Social Network Analysis as an Analytic Tool for Interaction Patterns in Primary Care Practices**

**Systematic Reviews:**

Bruce Arroll, Steve Macgillivray, Simon Ogston, Ian Reid, Frank Sullivan, Brian Williams, and Iain Crombie

**Efficacy and Tolerability of Tricyclic Antidepressants and SSRIs Compared With Placebo for Treatment of Depression in Primary Care: A Meta-Analysis**

Kara M. Levri, Elizabeth Slaymaker, Allen Last, Julie Yeh, Jonathan Ference, Frank D’Amico, and Stephen A. Wilson

**Metformin as Treatment for Overweight and Obese Adults: A Systematic Review**

**Reflections:**

Jon O. Neher

**The Decade Dance**
On TRACK:

Kurt C. Stange

Strength from Vulnerability

Family Medicine Updates:

From the Society of Teachers of Family Medicine

STFM LAUNCHES NPI AND CALLS FOR NOMINATIONS

From the Association of Departments of Family Medicine

FAMILY MEDICINE LEGISLATIVE ADVOCACY: OUR POWERFUL MESSAGE

From the Association of Family Medicine Residency Directors

QUALITY IN RESIDENCY EDUCATION

From the North American Primary Care Research Group

RESEARCH INVOLVING LATINO POPULATIONS

From the American Academy of Family Physicians

AAFP’S POLITICAL ACTION COMMITTEE ADDS TO FAMILY PHYSICIANS’ VOICE IN WASHINGTON

From the American Board of Family Medicine

LISTENING TO THE DIPLOMATES: PHYSICIANS’ FEEDBACK ON SELF-ASSESSMENT MODULES
In This Issue: Subtle Clinical Policy

Kurt C. Stange, MD, PhD, Editor

Conflicts of interest: none reported


TESTING FOR ACUTE HIV INFECTION DEMANDS A SUBTLE CLINICAL POLICY

In this issue, 2 related analyses make the case that it is cost-effective to test for acute HIV disease among outpatients with acute viral illness symptoms. An editorialist argues that cost-effectiveness analyses should not be used to guide the care of individual patients—that "cost-effectiveness analysis is designed for policy rather than clinical use." These articles provide a chance to consider how we use data and integrative research, such as cost-effectiveness analyses, to make policy and individual decisions.

I believe that the value of cost-effectiveness analyses, such as the one in this issue, extends beyond the yes or no answers they generate. If we look past the bottom line for the general population, these analyses can inform our judgment in personalizing decisions for individuals based on what Ian McWhinney called "an acquaintance with particulars." For the decision to test for acute HIV infection among patients with viral symptoms, Coco's cost-effectiveness analysis provides data on the important general factors to consider: the patient's specific viral symptoms, the prevalence of HIV infection in outpatients with such symptoms, and the characteristics of the available tests. In the vernacular of clinical guidelines and evidence-based medicine, many will judge there to be insufficient evidence to make a general recommendation to test all primary care outpatients, with even selected viral symptoms, for acute HIV infection. Authors of a recent analysis focused on screening the general population for HIV came to a similar conclusion.

The appropriately conservative evidence-based medicine (EBM) criteria for making general recommendations mask the great subtlety of excellent clinical care. The judgment that data are insufficient to make a general policy recommendation is not the same as judging that the evidence is insufficient in specific patient populations or individuals. Too heavy-handed an application of EBM can result in the de-intellectualization and depersonalization of practice.

We need to guard against the insidious effect of a laudable focus on scientific evidence in devaluing the importance of focusing on the particulars of patients, families, communities, and local practices. This insidious effect is apparent in continuing education forums in which subtle insights not based on population studies are spoken of apologetically or sometimes not at all. The effect is seen in some clinicians trained during the EBM era, who do not feel empowered to use their intuition to go beyond classic scientific evidence to engage different ways of knowing. It is seen in a model of practice that forces the 10-minute visit for financial survival rather than allowing the time for relationships and on-the-ground knowledge to guide subtle practice. Paying for performance, when the performance is based on the tyranny of what can be measured rather than on paying attention to the particulars, has the potential to squelch the subtlety and personalization of our practice. More hopeful options are found in system transformation efforts that aim to support sophisticated individualization of care. Further guidance is found in the call by one of the founders of EBM for the unification of both evidence-based policy guidelines and evidence-based individual decision making, and in a clinical and research framework that integrates different ways of knowing.

In my practice, the cost-effectiveness analysis by Coco in this issue has led to a subtle change. I am not ready to recommend testing for acute HIV infection to even a minority of patients in my practice with viral symptoms—there is no formal policy change as a result of this study. This analysis, however, has made me look for information on the prevalence of HIV disease in my practice population and in my community. It has made me consider the possibility that certain viral symptoms may be a harbinger of acute HIV infection. I now ask more about HIV risk factors in patients and partners and will, on occasion, discuss testing for HIV infection among some patients with viral illness. I also consider both the potentially beneficial and harmful effects of even asking about the risk of HIV infection, although...
the available evidence doesn’t apply well to the situation of patients with acute viral symptoms.

Cost-effectiveness analysis can lead to general clinical rules—policy at 10,000 feet. Clinical care is provided by community-based practices at the level of individuals and families—policy on the ground. A combination of support for general strategies at 10,000 feet and freedom to implement subtle clinical policies on the ground is needed to provide care that is both effective and cost-effective.

STUDIES OF CLINICAL PHENOMENA AND PRACTICE APPROACHES

A careful qualitative study by Walter and colleagues\(^14\) in this issue identifies how patients understand the meaning of their family history of cancer, heart disease, and diabetes. The findings are important for efforts to inform patient- and family-centered risk assessment and communication.

Epstein and colleagues\(^15\) used a covert standardized patient to evaluate physicians’ patient-centered communication. They find that greater patient-centeredness is associated with lower rates of diagnostic testing and greater health care expenditures, but also with increased visit length. Without reimbursement systems that recognize this trade-off, the systems and patient benefits of a patient-centered approach are paid for by the primary care clinician and are thus disincentivized.

An interesting study of pneumonia treatment in nursing home patients takes advantage of the natural experiment of intercountry practice differences.\(^16\) The authors find that patients’ health status appears to be more important than the aggressiveness of antibiotic treatment in mortality among nursing home residents with lower respiratory tract infection.

In an intensive mixed methods study of Midwestern family medicine practices, Crabtree and colleagues\(^17\) examine the place of clinical preventive services among competing demands for provision of acute and chronic illness care. They find diverse practice approaches that focus on the clinical encounter, and they identify factors that may be useful in designing practice-individualized improvement strategies.

METHODOLOGY STUDIES, SYSTEMATIC REVIEWS AND REFLECTIONS

Two methodology studies show the value of a new measurement tool and of a novel application of a promising analysis technique. Shields and colleagues\(^18\) describe the development and validation of the theory-based Rochester Participatory Decision Making Scale. In another study, social network analysis\(^19\) is applied to characterize 2 family medicine practices. Previously used in public and community health settings, the authors find this technique to be valuable in quantitatively analyzing interaction patterns and in understanding differences between practices.

One systematic review finds evidence for common clinical practices that have been understudied in the primary care setting, whereas another review finds that the available data do not support an emerging clinical practice. Arroll and colleagues\(^20\) find that both tricyclic antidepressants and selective serotonin reuptake inhibitors are effective for treatment in primary care. This information is important because previous systematic analyses have not focused on the primary care setting. These authors find that the relatively low doses of tricyclic antidepressants sometimes used in primary care may be effective. The systematic review by Levri and colleagues\(^21\) finds insufficient evidence to support using metformin as treatment of overweight or obese adults without diabetes, or for treatment of nondiabetic women with polycystic ovary syndrome.

In an essay portraying a challenging patient, Neher\(^22\) shows how uncovering a remote traumatic event helps to understand past missed opportunities while requiring great care to manage the resulting anger and its potential consequences. Interestingly, the online discussion\(^23\) of 2 articles from the last issue of Annals\(^24,25\) raises related issues about the importance and potential difficulties of uncovering past emotional traumas.

The impact of the research and essays in this issue will be enhanced by your interpretation. We invite all Annals readers to participate in the online discussion at http://www.AnnFamMed.org.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/386.

References

Should We Screen Patients With Viral Symptoms for HIV Disease?

Theodore G. Ganiats, MD
Health Outcomes Assessment Program, University of California, San Diego, La Jolla, Calif

I
n this issue, Coco addresses a vitally important question: Is it cost-effective to test for acute HIV infection among outpatients complaining of viral symptoms and at least 1 risk factor for HIV infection? If so, what is the appropriate method for such testing? In a companion article Coco and Kleinhaus address a key factor in answering these questions: What is the prevalence of acute HIV infection in 13- to 54-year-old ambulatory patients who have any of 17 viral symptoms?

The author offers a convincing justification for this analysis: primary HIV infection is a major factor in the HIV epidemic, and most patients become symptomatic and seek care but are seldom tested or have HIV diagnosed. Thus, we miss an opportunity to intervene early. He determines the cost-effectiveness of expanding testing for primary HIV infection to a large cohort of outpatients.

How can we best use this impressive work? Although Coco’s conclusions are based on a model, we cannot wait for a randomized controlled trial to confirm his findings, because there will likely never be such a trial addressing this question. Are the results valid? How should this research affect the care we provide our patients, ie, how should we apply the results tomorrow? These complex questions cut to the core of cost-effectiveness analysis. Cost-effectiveness analysis is a tool, and like other tools we use in medicine, we need to know its strengths and limitations—how we can best use it to for the betterment of our patients.

Cost-effectiveness analysis is designed to assist decision makers, specifically health policy makers. There are both ethical and pragmatic reasons why we should be cautious in applying cost-effectiveness analyses, such as Coco’s study, directly to our clinical practice.

First, in all but the simplest of decisions, a decision model cannot include all the elements we would consider important. Not only are some the variables unknown, but many values are not in the model. For example, how much do patients fear the needle for the blood draw? How do patients value or fear the concept of HIV infection? Without answers to these and other questions, the model is incomplete. Results of the model are therefore similar to other elements of a decision process, such as considerations of politics, ethics, and justice for policy decisions; and history, physical, and laboratory information for clinical decisions. They inform but do not dictate the decision.

Second, what assumptions does the analyst make regarding the variables? Coco assumes, for example, that the life expectancy of the patient is 39.5 years, a good average for his model, as well as a good assumption for the policy maker trying to decide a health benefits package. His results, however, will not apply to either a 20-year old college student or a 70-year old retired librarian. Coco explicitly acknowledges this limitation, stating the impact of the report, not in clinical terms, but in policy terms: “Expanded testing for primary HIV infection … may be a sound expenditure of health care resources.” In other words, the primary decision makers who will use his findings are not patients or clinicians at the point of care but policy makers who set the framework from which we provide care.

A third key issue relates to the uncertainty inherent in any complex decision. Cost-effectiveness analysis not only makes explicit the assumptions related to these uncertainties, but uses sensitivity analysis to addresses the question, How sensitive are the results of the analysis to uncertainty in the variables? There are several types of uncertainty in this cost-effectiveness analysis. For some variables, eg, the cost of a return

Conflict of interest: none reported

CORRESPONDING AUTHOR
Theodore G. Ganiats, MD
Health Outcomes Assessment Program
Room 204, Stein Clinical Research Bldg
University of California San Diego
9500 Gilman Dr
La Jolla, CA 92093-0628
tganiats@ucsd.edu
office visit, the value for a given person is known, but there is a range of values in the population. For other variables, the individual patient has a single value, but it is not knowable. For example, it is not possible to know whether any given patient will be lost to follow-up, but a population average can be determined. Finally, there are times when we have no data on the variable, eg, the sensitivity and specificity of the p24 antigen EIA. Coco’s Table 1 lists all the key variables, the baseline estimate, and the range used in the sensitivity analysis, along with references to support the assumptions, thus allowing readers to make their own judgments.

To their credit, Coco and Kleinhans went the extra yard to estimate a key variable for this analysis: the national prevalence of primary HIV infection in patients visiting ambulatory settings with fever, rash, or sore throat, and a diagnosis consistent with an acute viral illness. The additional study, which is published in a sister article, is a beautiful example of how one can inform cost-effectiveness analysis with data from large data sets, specifically the National Ambulatory Medical Care Survey, the National Hospital Ambulatory Medical Survey, and data from the Centers for Disease Control and Prevention. The prevalence of HIV infection in patients with viral symptoms seeking care in an ambulatory practice is the key variable in the cost-effectiveness analysis and, therefore, in the decision to test for HIV infection. The limited data on the variability of the prevalence of acute HIV infection among ambulatory patients with many constellations of viral symptoms is a major factor that policy makers and clinicians should consider before applying Coco’s cost-effectiveness analysis to specific settings and to specific types of patients and clinical presentations.

Given the necessary limitations these assumptions place on the generalizability to our patients, does this analysis deserve space in the Annals of Family Medicine? The answer is a resounding yes. Coco’s paired articles deserve close evaluation by clinicians, researchers, and policy makers. Clinicians will see that screening for HIV infection in those with viral symptoms will often be cost-effective, though it may be premature to implement such screening into practice. Researchers now have new areas of investigation to provide better data for this cost-effectiveness model. Policy makers now have an excellent study to help inform their decisions related to screening patients with viral symptoms for acute HIV disease. Given the current state of our knowledge and based on Coco’s analysis, screening for HIV infection with p25 antigen EIA in those with acute viral symptoms should be viewed as a valid use of resources, and consideration should be given to developing policies supporting this practice.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/389.

Key words: HIV infections/prevention & control; primary HIV infection/epidemiology; cost-benefit analysis, mass screening; delivery of health care; health services research

Submitted August 2, 2005; accepted August 30, 2005.

References
The Cost-Effectiveness of Expanded Testing for Primary HIV Infection

Andrew Coco, MD, MS
Healthcare Research Center, Lancaster General Hospital, Lancaster, Penn

ABSTRACT

PURPOSE Primary infection with the human immunodeficiency virus (HIV) is a major factor in the HIV epidemic. Most patients become symptomatic and seek care, but seldom are they tested or is their condition diagnosed. The objectives of this study are to determine whether it is cost-effective to expand testing for primary HIV infection to a larger cohort of patients, and, if so, which diagnostic assay is most cost-effective.

METHODS We undertook a cost-effectiveness analysis of testing a hypothetical cohort of more than 3 million outpatients with fever and other viral symptoms regardless of HIV risk factors using 3 diagnostic assays: p24 antigen enzyme immunoassay (EIA), HIV-1 RNA assay, and third-generation HIV-1 EIA. Antiretroviral therapy was started when the CD4 cell count decreased to 350/µL. Outcome measures were the incremental cost-effectiveness of the diagnostic assays, number of cases identified, cases avoided in sexual partners, and threshold prevalence. For sensitivity analyses, we used $50,000 as the threshold for cost-effectiveness.

RESULTS At the baseline prevalence of 0.66%, p24 antigen EIA testing was the most cost-effective option at a cost of $30,800 per quality-adjusted life-year gained when compared with no testing. There were 17,054 cases identified, and infection was avoided in 435 partners. Probabilistic sensitivity analysis, in which the estimates for all variables are varied simultaneously, determined that expanded testing with p24 antigen EIA compared with no testing had a 67% probability of being cost-effective at the baseline prevalence and a 71% probability at a prevalence of 1%.

CONCLUSIONS Expanded testing for primary HIV infection with p24 antigen EIA may be a sound expenditure of health care resources.


INTRODUCTION

Primary human immunodeficiency virus (HIV) infection is a transient, symptomatic illness characterized by high HIV-1 RNA levels before an effective immune response develops. More than 90% of cases go undiagnosed even though up to 90% of patients have symptoms and seek medical care. Primary HIV infection plays a major role in the HIV epidemic.

Diagnosing primary HIV infection allows for an important opportunity to interrupt HIV transmission, because persons in this stage of HIV infection can be sources of new infections. First, they are in a state of heightened infectivity because they have high levels of viremia, often with a molecule count exceeding 1,000,000/µL. Additionally, because affected persons are unaware of the diagnosis, they might not practice preventive sexual behavior. Many patients who learn of being HIV-infected adopt behaviors that can reduce the risk for transmitting HIV. In addition to the public health benefits of early diagnosis, patients can also benefit by being observed for immune function deterioration and given timely initiation of antiretroviral therapy based on clinical practice guideline recommendations.

There are numerous reasons why 90% of cases of primary HIV infec-

Conflict of interest: none reported

CORRESPONDING AUTHOR
Andrew Coco, MD, MS
Director, Healthcare Research Center
Comprehensive Care Clinic for HIV Disease
Family and Community Medicine
Lancaster General Hospital
555 N Duke St
Lancaster, PA 17604-3555
ascoco@lancastergeneral.org
tion are undiagnosed. Primarily, the symptom complex is similar to that of influenza and other nonspecific viral illnesses. Several studies have concluded that no symptom is sufficiently sensitive or specific to allow for targeted testing. Second, it is difficult to determine who has a high probability of primary HIV infection when evaluating risk factors alone. Some patients may be unaware of their risky behavior. Health care workers can further contribute to the problem by failing to ask about risk factors, or they might feel unqualified to manage issues associated with HIV infection.

Confusion about which test to order can also contribute to the low detection rate of primary HIV infection. Sensitivity, specificity, and cost limit the available options. At the initial stages of infection, HIV-1 antibody tests are nonreactive, and the diagnosis of primary HIV infection is made by either the p24 core antigen test or HIV-1 RNA assays (viral load). The p24 core antigen test is more specific and less expensive than the HIV-1 RNA assays, but it is less sensitive in detecting cases. The HIV-1 RNA assays are highly sensitive, but they are more expensive and have decreased specificity, which can result in a higher false-positive rate when used in groups that have a low disease prevalence. Another option is the third-generation HIV-1 antibody test, which is able to detect immunoglobulin (Ig) M as well as IgG antibodies. The Panel on Clinical Practices for Treatment of HIV Infection through the US Department of Health and Human Services (DHHS) recommends testing with the HIV-1 RNA assay when acute infection is suspected and risk factors are present but advises against testing lower risk populations because of concerns about false-positive diagnoses. Given the low percentage of primary HIV infection diagnoses, however, it is unlikely that many clinicians follow these guidelines and do any testing.

Because of the differences in characteristics and costs of the available diagnostic assays, a cost-effectiveness analysis of expanded testing of persons with viral symptoms at varying prevalence rates could be useful for policy development and guiding clinical practice. The objectives of this study were to perform an incremental cost-effectiveness analysis of expanded testing of persons with viral symptoms and at least 1 risk factor using these 3 diagnostic assays for primary HIV infection and to determine the lowest prevalence at which expanded testing is cost-effective.

METHODS

Study Design

The study was an incremental cost-effectiveness analysis of 3 tests for primary HIV infection using a decision analytical model. A strategy of no testing was used for baseline comparison, because little testing for this diagnosis is currently done. The analysis included the following outcome measures of expanded testing for primary HIV infection: (1) increased quality-adjusted life-years (QALYs) in the patient as a result of starting antiretroviral therapy before profound deterioration in immune function occurs; and (2) prevention of HIV infection in the patient's sexual partners. Several potential outcome measures were not included: (1) benefits of immunizations (hepatitis A and B and Streptococcus pneumoniae), cervical cancer screening, and tuberculosis screening; (2) prevention of HIV infection in needle-sharing contacts of injection drug users; (3) avoidance of costly diagnostic workups and hospitalization for those with more severe primary HIV infection symptoms; and (4) benefits of detecting cases of chronic HIV infection. Starting antiretroviral treatment during the acute phase of HIV was not included because data showing a benefit for structured treatment interruptions in preserving natural immunity were lacking. The harms from expanded testing included in the analysis were (1) decreased quality of life caused by the anxiety resulting from a false-positive diagnosis, and (2) decreased quality of life associated with being aware of being HIV infected.

The time frame for the study was 39.9 years, the average life expectancy of the mean age (39.5 years) of the hypothetical cohort being tested. The analysis adopted a societal perspective, including all costs and health effects, except for work loss and transportation, which were considered negligible compared with laboratory testing, visit costs, and lifetime medical costs for treatment of HIV disease. The analysis also provided information from the perspective of a third party payer by determining the cost per case of primary HIV infection diagnosed. We conducted sensitivity analyses to determine the stability of the results using reasonable variations in the data and assumptions.

Target Population

A national estimate of the prevalence of primary HIV infection in symptomatic ambulatory patients regardless of risk factors is published in this issue of Annals. The estimated primary HIV infection prevalence for patients with a fever was 0.66% (95% confidence interval [CI], 0.53%-0.92%), for those with a rash it was 0.56% (95% CI, 0.35%-0.94%), and for those with a sore throat it was 0.13% (95% CI, 0.10%-0.19%). The highest of these rates (0.66%) was used as the baseline for the analysis. To make population projections about the number of cases diagnosed, number of infections avoided, and program costs, we assumed that 20,000, or 50% of the 40,000 annually newly infected patients annually, developed symptoms and sought care. At the baseline prevalence of 0.66%, 3,030,303 patients would need to be tested in the model.
The prevalence of primary HIV infection has also been directly measured in a study of patients at an urban urgent care center who complained of viral symptoms and had at least 1 HIV risk factor. The prevalence in this study was 1% (95% CI, 0.1%-1.9%). This prevalence was used to explore how a wider range of values affected the analysis.

Diagnostic Tests
Three tests with considerably different characteristics are available: (1) HIV-1 RNA assay (by polymerase chain reaction, branched-chain DNA, or transcription-mediated amplification, which is 100% sensitive but has a false-positive rate of 2% and is more expensive; (2) p24 antigen enzyme immunosorbent assay (EIA), which is less sensitive but has a false-positive rate of almost 0% and a moderate cost, and (3) third-generation HIV-1 EIA, which is less sensitive and specific than the other 2 tests, but costs the least. This test, also known as an HIV-1 antigen sandwich or combitest, is capable of detecting both IgG and IgM antibodies. Standard HIV-1 antibody EIA tests detect only IgG antibody and are much less sensitive for primary HIV infection. A rapid test (OraQuick HIV rapid test) became available in late 2002 that could identify HIV-infected persons earlier and decrease the failure rate for return visits for test results. This test was not included in the analysis because we lacked data in the setting of primary HIV infection testing.

Study Model
The study model is a decision analytical model (Figure 1) in which a hypothetical cohort of 3,030,303 million patients with fever and other viral symptoms consistent with primary HIV infection, regardless of risk factors, were tested by either the p24 antigen EIA, an HIV-1 RNA assay, a third-generation HIV-1 EIA, or not tested at all. In the model sufficient serum was collected initially so that a standard HIV-1 antibody test and, if needed, a confirmatory Western blot reflexively test could be done for those with positive initial tests. All patients were scheduled for a return visit in 14 days for test results and posttest counseling; a projected follow-up rate was 69%. Patients with a positive test had follow-up HIV clinical care arranged, including transportation, if necessary. After testing, patients were categorized through Bayesian revision according to test result: true positive, false positive, true negative, or false negative.

Event Pathways
Patients in the true-positive category who continued care were seen for an initial visit at 1 month, at which time the standard HIV antibody test and Western blot assay would be positive. At this visit, an initial battery of laboratory tests (Table 1), as well as a CD4 cell count and a quantitative HIV-1 RNA assay, was ordered. The cost of genotype resistance testing was not included in the baseline laboratory work, but it was added, as reflected in the higher range of values used in the sensitivity analysis (Table 1). Continued care entailed a return visit, a CD4 cell count determination, and a quantitative HIV-1 RNA assay every 4 months. Surveillance visits continued for 6.2 years, the mean duration of time it takes from diagnosis for the CD4 cell count reach 350/µL. At this stage of infection,
3-drug antiretroviral therapy was given according to Department of Health and Human Services guidelines, and patients incurred lifetime medical costs and survival times based on estimates from a mathematical simulation model of HIV disease.7,23 Additionally, some patients in this pathway altered their sexual behavior and thus avoided infecting their sexual partner during the several weeks of primary HIV infection symptoms. Partners who avoided infection avoided incurring lifetime HIV medical costs and reduced quality of life. Future behavior changes outside the symptomatic primary HIV infection period were not included in the analysis. Patients that failed to return for follow-up had HIV infection diagnosed at a later stage and incurred reduced lifetime medical costs and survival times based on an extrapolation of data from previous estimates of stage of disease when diagnosed, lifetime medical costs after starting combination antiretroviral therapy, and QALYs of survival.14,23

Patients in the false-positive category were assumed Table 1. Summary of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline Estimate</th>
<th>Range Used in Sensitivity Analysis</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs ($)</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>p24 antigen EIA</td>
<td>24.65</td>
<td>12.33-49.30</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>HIV-1 RNA assay</td>
<td>118.89</td>
<td>59.45-237.78</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Third-generation HIV-1 EIA</td>
<td>19.17</td>
<td>9.59-38.34</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Western blot assay</td>
<td>27.05</td>
<td>13.53-54.10</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>CD4 cell count per microliter</td>
<td>90</td>
<td>45-180</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Initial battery of laboratory tests (new diagnosis)*</td>
<td>254</td>
<td>200-614</td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Expanded testing program costs</td>
<td>101.47</td>
<td>51-203</td>
<td>MMWR20</td>
</tr>
<tr>
<td>Discounted lifetime medical costs (diagnosed with PHI and antiretroviral therapy started at CD4 cell count of 350/µL)</td>
<td>95,800</td>
<td>47,900-191,600</td>
<td>Freedberg et al23</td>
</tr>
<tr>
<td>Discounted lifetime medical costs (PHI not diagnosed and antiretroviral therapy started when HIV diagnosed)</td>
<td>88,100</td>
<td>44,050-176,200</td>
<td>Freedberg et al23</td>
</tr>
<tr>
<td>Return visit</td>
<td>52.53</td>
<td>40-67.86</td>
<td>Kaplan &amp; Anderson24</td>
</tr>
<tr>
<td>Test characteristics</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>p24 antigen EIA, sensitivity</td>
<td>0.887</td>
<td>0.770-0.957</td>
<td>Hecht et al,8 Daar et al9</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9996</td>
<td>0.9950-0.9999</td>
<td>Hecht et al,8 Daar et al9</td>
</tr>
<tr>
<td>HIV-1 RNA, sensitivity</td>
<td>1.000</td>
<td>—</td>
<td>Hecht et al,8 Daar et al9</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.980</td>
<td>0.950-0.999</td>
<td>Hecht et al,8 Daar et al9</td>
</tr>
<tr>
<td>Third-generation HIV-1 EIA, sensitivity</td>
<td>0.790</td>
<td>0.600-0.920</td>
<td>Hecht et al8</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.970</td>
<td>0.930-0.990</td>
<td>Hecht et al8</td>
</tr>
<tr>
<td>Probability of indeterminate Western blot</td>
<td>0.000004</td>
<td>—</td>
<td>Kleinman et al22</td>
</tr>
<tr>
<td>Prevalence factors (%)</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Patients lost to follow-up</td>
<td>31</td>
<td>16-62</td>
<td>MMWR21</td>
</tr>
<tr>
<td>Prevalence in screened population</td>
<td>0.66</td>
<td>0.53-0.92</td>
<td>Coco &amp; Kleinhans16</td>
</tr>
<tr>
<td>Sexual transmission factors</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Patients who change behavior to avoid infecting sexual partner</td>
<td>50</td>
<td>0-96</td>
<td>MMWR8</td>
</tr>
<tr>
<td>Patients that are sexually active</td>
<td>50</td>
<td>25-85</td>
<td>MMWR8</td>
</tr>
<tr>
<td>Infectivity (probability of sexual transmission during PHI period)</td>
<td>15</td>
<td>0-30</td>
<td>Yerly et al,25 Pilcher et al26</td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>Asymptomatic HIV infection</td>
<td>0.937</td>
<td>0.926-0.949</td>
<td>Schackman et al27</td>
</tr>
<tr>
<td>Anxiety while waiting for confirmatory test results for patients with a positive screen</td>
<td>0.682</td>
<td>0.400-0.800</td>
<td>Kaplan &amp; Anderson24</td>
</tr>
<tr>
<td>Quality-adjusted life-expectancy (discounted), years</td>
<td></td>
<td></td>
<td>Medicare fee schedule</td>
</tr>
<tr>
<td>No PHI</td>
<td>24</td>
<td>—</td>
<td>NCHS15</td>
</tr>
<tr>
<td>Positive screening result, no PHI</td>
<td>23.9735</td>
<td>23.950-23.983</td>
<td>NCHS,15 Kaplan &amp; Anderson24</td>
</tr>
<tr>
<td>PHI diagnosed at screening with follow-up care and antiretroviral treatment started at CD4 cell count of 350/µL</td>
<td>11.9</td>
<td>11.832-11.952</td>
<td>Freedberg et al23</td>
</tr>
<tr>
<td>PHI not diagnosed at screening or lost to care with antiretroviral treatment started when HIV diagnosed</td>
<td>11</td>
<td>—</td>
<td>Freedberg et al23</td>
</tr>
</tbody>
</table>

EIA = enzyme immunosorbent assay; HIV = human immunodeficiency virus; PHI = primary HIV infection; CBC = complete blood count; G6PD = glucose-6-phosphate dehydrogenase; CMV = cytomegalovirus; RPR = rapid plasma reagin; PPD = purified protein derivative (tuberculin); NAAT = nucleic acid amplification test.

* Initial battery of laboratory tests includes: CBC, chemistry panel, G6PD assay, toxoplasmosis titer, CMV titer, RPR, PPD skin test, viral hepatitis panel, lipid panel, urinalysis, chest radiograph, urine NAAT for gonorrhea and chlamydia.
to have intermediate results on Western blot assay. Patients in this pathway were observed for 3 months with 2 follow-up visits. Testing confirmed a negative result after 2 standard HIV antibody tests, 3 Western blot assays, and a quantitative HIV-1 RNA assay.

Patients in the false-negative category also had a brief follow-up visit with an HIV counselor to be informed of the negative results.

Patients that were not tested reported symptoms at later stages of infection, as did those patients in the true-positive category who were lost to follow-up.

The primary outcome measures were the incremental cost per QALY gained, the number of cases of primary HIV infection identified, and the threshold prevalence at which expanded testing had a cost per QALY of less than $50,000, a value that is generally considered to be a cost-effectiveness threshold for a single patient. This analysis includes both patients and their partners; the ramifications of this approach are discussed below. Other outcomes were the number of cases of infection avoided through changes in sexual behavior, the number of false-positive diagnoses, the number of false-negative diagnoses, and the cost per case diagnosed. The model was programmed using decision analysis software (TreeAgePro [version 2004], TreeAge Software Inc, Williamstown, Mass).

**Decreased Transmission Through Change in Sexual Behavior**

From a public health perspective, determining which patients are in this phase of infection can decrease their high-risk behavior and have a substantial impact on subsequent transmission. In an unpublished study from the Centers for Disease Control and Prevention (CDC) of 1,363 HIV-infected men and women, among the 69% who were sexually active during the preceding 12 months, 78% to 96% used a condom at their most recent anal or vaginal intercourse with a known HIV-negative partner, and 52%-86% reported condom use with a partner of unknown serostatus. The analysis used lower rates of 50% for both estimates. To determine a rate of transmission, a study of 197 persons with documented primary HIV infection showed, through gene sequencing and contact tracing, that transmission occurred at the time of primary HIV infection in 30%. In keeping with a bias against expanded testing, the analysis decreased this rate by one half to 15%, because HIV can be transmitted during the presymptomatic phase of primary HIV infection.

**Data on Costs**

The costs of conducting an HIV-testing program were obtained from a state-funded program in Massachusetts that offered HIV counseling, testing, and referral to 3,068 patients entering 1 of 4 hospital-associated urgent care centers. Included in this program were the cost of HIV counselors and intake nurses who arranged for telephone follow-up, visits to homeless shelters, and travel vouchers to bring positive patients into care. The costs of laboratory tests were obtained from the Medicare fee schedule for Lancaster, Pa, for 2002. Clinical visit costs were obtained from a national survey of physician’s office charges. Lifetime medical costs incurred after the initiation of 3-drug antiretroviral therapy at CD4 cell counts of 350/µL and lower were based on an extrapolation of data from a computer-simulated model of HIV-infected persons. Costs were converted to 2002 dollars by the medical care component of the Consumer Price Index. Future costs were discounted at a rate of 3%.

**Data on Health-Related Quality of Life**

Quality of life for asymptomatic HIV infection was obtained from a national probability sample of HIV-infected adults. Utility values for the mental anguish resulting from waiting for confirmatory tests of a positive screening test were derived from the Quality of Well-Being index. QALYs for patients starting 3-drug antiretroviral therapy at CD4 cell counts of 350/µL or less until time of death were based on an extrapolation of data from previous estimates of age and stage of disease presentation and QALYs of survival after initiation of combination antiretroviral therapy. Future gains in quality of life were discounted at a rate of 3%.

**RESULTS**

**Base Case Analysis**

The base case analysis results are displayed in Table 2. Expanded testing with the p24 antigen EIA test had the lowest incremental cost-effectiveness ratio of the 3 expanded testing strategies. Under base case assumptions, the cost per QALY for the p24 antigen EIA was $30,800 compared with no testing. Testing with an HIV-1 RNA assay or the third-generation HIV-1 EIA was dominated, or evaluated as inferior, by the p24 EIA strategy.

The p24 antigen EIA strategy, because of the high specificity of the test, resulted in significantly fewer false-positive diagnoses: 1,127 compared with 90,257 and 59,169 for the third-generation EIA and HIV-1
RNA assay, respectively (Table 2). The HIV-1 RNA assay, because of the high sensitivity of the test, identified 2,946 more cases of primary HIV infection and allowed for the avoidance of 66 more cases of infected partners compared with the p24 antigen EIA. Additionally, there were no false-negative cases with this option compared with 3,012 with the p24 antigen EIA option. The analysis also calculated costs from a third party payer perspective as testing cost per case identified. The p24 antigen EIA option had the lowest cost per case identified ($29,090).

**Sensitivity Analyses**

**Threshold Analysis of Prevalence of Primary HIV Infection**

To determine the lowest primary HIV infection prevalence at which expanded testing was cost-effective at the usual standard of $50,000 per QALY, we performed a 1-way sensitivity analysis (keeping all other variables constant) using the 95% CI range of the directly measured prevalence estimate (0.1%-1.9%). Expanded testing with the p24 antigen EIA exceeded this standard when compared with no testing at a prevalence of less than 0.35%. The incremental cost-effectiveness ratios for other prevalence rates reported in the literature, comparing expanded testing of primary HIV infection using the p24 antigen EIA with no testing, showed the following results: at a prevalence of 1% (urban patients with risk factors) costs were $23,000 per QALY; at a prevalence of 0.56% (patients with a rash and other viral symptoms) costs were $35,000 per QALY; at a prevalence of 0.13% (patients with a sore throat and other viral symptoms) costs were $129,000 per QALY.

**Other 1-Way Sensitivity Analyses**

The baseline results remained fairly stable when compared with other 1-way analyses using the variable ranges in Table 1. The p24 antigen EIA remained the most cost-effective option, with the cost per QALY, ranging from $15,500 to $50,600 when compared with no testing. The HIV-1 RNA assay continued to be inferior or have a cost per QALY of greater than $100,000 compared with the p24 antigen EIA, and the third-generation HIV-1 EIA remained inferior. The results of the key variables that had the largest impact on the baseline results are shown in Table 3.

**Multiway Sensitivity Analyses**

We performed a Monte Carlo simulation (probabilistic sensitivity analysis) in which the values for all the variables listed in Table 1 were simultaneously varied. We entered each variable as a probability distribution based on reported 95% CIs, when available, or as a reasonable range as indicated in the third column in the table. The log-normal distribution was assumed for cost variables and the beta distribution for probability and utility variables. We randomly selected new values from within each of the probability distributions during each of 100,000 iterations and calculated 95% likelihood comparisons of the strategies.

---

**Table 2. Cost, Effectiveness, and Incremental Cost-Effectiveness of Expanded Testing for Primary HIV Infection of 3,030,303 Hypothetical Patients at a Prevalence of 0.66% With Third-Generation HIV-1 EIA, p24 Antigen EIA, and HIV-1 RNA Assay**

<table>
<thead>
<tr>
<th>Variable</th>
<th>No Testing</th>
<th>Third-Generation HIV-1 EIA</th>
<th>p24 Antigen EIA</th>
<th>HIV-1 RNA Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost (millions), $</td>
<td>1,762.1</td>
<td>2,233.6</td>
<td>2,258.2</td>
<td>2,561.8</td>
</tr>
<tr>
<td>Incremental cost (millions), $</td>
<td>—</td>
<td>471.5</td>
<td>24.6</td>
<td>303.6</td>
</tr>
<tr>
<td>Effectiveness (thousands) QALYs</td>
<td>69,710.0</td>
<td>69,720.8</td>
<td>69,726.1</td>
<td>69,725.8</td>
</tr>
<tr>
<td>Incremental effectiveness, QALYs</td>
<td>—</td>
<td>10,800</td>
<td>5,300</td>
<td>(300)</td>
</tr>
<tr>
<td>Effectiveness, No.</td>
<td>—</td>
<td>15,803</td>
<td>17,054</td>
<td>20,000</td>
</tr>
<tr>
<td>Primary HIV infection cases diagnosed</td>
<td>20,000</td>
<td>4,899</td>
<td>5,287</td>
<td>6,200</td>
</tr>
<tr>
<td>Primary HIV infection cases lost to care (31% of those diagnosed)</td>
<td>—</td>
<td>90,257</td>
<td>1,127</td>
<td>59,169</td>
</tr>
<tr>
<td>False-positive diagnoses</td>
<td>—</td>
<td>2,924</td>
<td>3,012</td>
<td>0</td>
</tr>
<tr>
<td>False-negative diagnoses</td>
<td>—</td>
<td>403</td>
<td>435</td>
<td>501</td>
</tr>
<tr>
<td>Cases avoided per behavior change</td>
<td>—</td>
<td>29,836</td>
<td>29,090</td>
<td>39,985</td>
</tr>
<tr>
<td>Cost-effectiveness, $</td>
<td>—</td>
<td>Dominated*</td>
<td>30,800</td>
<td>Dominated*</td>
</tr>
<tr>
<td>Testing cost per case identified</td>
<td>—</td>
<td>Dominated*</td>
<td>30,800</td>
<td>Dominated*</td>
</tr>
<tr>
<td>Incremental cost per quality-adjusted year of life gained</td>
<td>—</td>
<td>Dominated*</td>
<td>30,800</td>
<td>Dominated*</td>
</tr>
</tbody>
</table>

Note: each column is compared with the one to the left.

HIV = human immunodeficiency virus; EIA = enzyme immunosorbent assay; QALYs = quality-adjusted life-years.

* Dominated means this option cost more and was less effective than other options.
The p24 antigen EIA testing strategy had a 19% probability of being dominant (more effective and less costly), a 48% probability of having an a cost per QALY of less than $50,000, and a 33% probability of having a cost per QALY of more than $50,000 compared with no testing. There was a 0% chance of it being less effective. The HIV-1 RNA assay, when compared with the p24 antigen EIA, had a 3% probability of being dominant or having a cost per QALY of less than $50,000, a 44% probability of having a cost per QALY of more than $50,000, and a 53% probability of being inferior (less effective and more costly). Figures 2 and 3 show how these comparison probabilities changed with varying the prevalence.

**DISCUSSION**

This study is the first to determine the cost-effectiveness of expanded testing for HIV infection in the acute phase. Using conservative assumptions, expanded testing of patients with fever, other viral symptoms, and at least 1 risk factor with the p24 antigen EIA had 67% probability of being superior or having a cost per QALY of less than $50,000 compared with no testing. Because of the high specificity of the p24 antigen EIA, false-positive diagnoses would be limited to 1,127 of the 3,030,303 persons tested in the model. Expanded testing would result in short-term avoidance of 435 infections in sexual partners. Although not modeled, early diagnosis of HIV infection could also allow for long-term decreased transmission, because patients would be aware of their communicable status for many more years. Also sexual partners

---

### Table 3. Changes in Incremental Cost per Quality-Adjusted Life-Year in Key 1-Way Sensitivity Analyses

1. Doubling the cost of lifetime medical care for patients being observed to CD4 cell counts of 350/µL or seeking care at later stages of infection ($88,100/$95,800 to $176,050/$191,600) increased the cost of expanded testing with the p24 antigen EIA from $30,800 to $34,100 compared with no testing.
2. Doubling the expanded testing and counseling enrollment program costs ($101.47 to $203), increased the cost of the p24 antigen EIA testing option to $49,800 compared with no testing.
3. Increasing the specificity of the HIV-1 RNA assay (0.98 to 0.999) decreased the cost of the HIV-1 RNA assay option to $142,000 compared with the p24 antigen EIA option.
4. Assuming no benefit to sexual partners of patients with PHI, ie, no cases avoided through changes in behavior, increased the cost of the p24 antigen EIA to $50,600 when compared with no testing.

EIA = enzyme immunosorbent assay; HIV = human immunodeficiency virus; PHI = primary HIV infection.
who avoid infection would not transmit the infection to others. Consequently, expanded testing for primary HIV infection could have a strong impact on curtailing the HIV epidemic and contribute greatly to achieving the CDC goal of reducing the annual number of HIV infections by 50% per year by 2005.\textsuperscript{33}

Several recent studies have addressed with varying results the issue of screening for HIV, not primary HIV infection, in the general population.\textsuperscript{34,35} One study found that 1-time screening of the general population with a prevalence of 0.1\% would cost $113,000 per QALY, whereas another analysis found that screening populations with a prevalence as low as 0.05\% had a cost per QALY of $50,000 when including costs and benefits for partners. It is difficult to compare these results with this analysis of expanded testing for primary HIV infection because of the different screening costs and benefits to partners.

Our analysis has several limitations. The estimates of lifetime medical costs after initiation of antiretroviral therapy were derived from a study that used 1998 dollars for cost estimates.\textsuperscript{23} Although these costs were adjusted to 2002 dollars using the medical care component of the consumer price index, it is possible that the cost of antiretroviral therapy has increased faster than that of other medical costs. Again, however, doubling these costs through sensitivity analysis did not appreciably alter the incremental cost-effectiveness ratios of the screening options, because without any expanded testing program, HIV infection would be diagnosed at a later stage and would incur similar costs when treatment was started (Table 3).

The analysis was also limited by using test characteristics derived from previous studies that included patients who were recruited when they were outside the symptomatic phase of primary HIV infection.\textsuperscript{8,9} The sensitivity of the p24 antigen EIA, in particular, could have been underestimated in these studies. Our analysis was based on screening patients who sought care because of symptoms, which have an average duration of 14 days.\textsuperscript{2} In a previous study, 100\% of 20 patients with primary HIV infection tested within 1 week of symptom onset had positive p24 antigen EIA results, but they had negative results 3 weeks after symptom onset.\textsuperscript{36} A higher sensitivity estimate for the p24 antigen EIA would have resulted not only in a more favorable incremental cost-effectiveness ratio compared with no testing but also more cases detected and fewer false-negative diagnoses. Additionally, newer tests, such as fourth-generation HIV-1 EIAs and p24 antigen signal-amplification-boosted EIA of heat-denatured plasma, are now available that may perform better in screening for primary HIV infection, but have yet to be evaluated.\textsuperscript{37,38}

The baseline prevalence estimate was derived from an analysis of a national outpatient database of patients with fever and other symptoms and diagnoses consistent with primary HIV infection.\textsuperscript{16} That estimate was based on assumptions that may not reflect the actual number of patients with primary HIV infection seeking medical care. Until more accurate data are available on the prevalence of primary HIV infection in low-risk populations, caution should be exercised in developing expanded testing policy based on these results. This analysis, however, showed that expanded testing for primary HIV infection has a 60\% probability of being cost-effective at a prevalence of 0.35\%, or almost one half of the baseline rate.

Finally, cost-effectiveness analyses typically apply the $50,000 per QALY standard to individual patients. This analysis combined the QALYs of the patients being tested and their sexual partners who avoided infection because of the communicable nature of HIV infection to show the full impact of an expanded testing program. In 1-way sensitivity analysis it was shown that without preventing transmission to sexual partners, the p24 antigen EIA had a cost per QALY of $50,600 compared with no testing (Table 3).

Using reasonable assumptions, the analysis has shown that expanded testing for primary HIV infection with the p24 antigen EIA has a high probability of being cost-effective. The cost-effectiveness of expanded testing for primary HIV infection in populations with a prevalence of 0.66\% or greater compares favorably with accepted screening programs, such as colon cancer screening, annual Papanicolaou smears in HIV-infected women, and breast cancer screening.\textsuperscript{39-41} Expanded testing for primary HIV infection may be a sound expenditure of health care resources and could have an impact on curtailing the HIV epidemic in the United States.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/391.

Key words: HIV infections/prevention & control; cost-benefit analysis; mass screening; delivery of health care; health services research


References


Prevalence of Primary HIV Infection in Symptomatic Ambulatory Patients

Andrew Coco, MD, MS
Emily Kleinhans

1Health Research Center, Lancaster General Hospital, Lancaster, Penn
2Goucher College, Baltimore, Md

ABSTRACT

PURPOSE Recognizing primary human immunodeficiency virus (HIV) infection is important for public health. The prevalence in outpatient settings is largely unknown but would be useful in developing testing guidelines. The objective of this study is to estimate the national prevalence of primary HIV infection in symptomatic ambulatory patients regardless of risk factors.

METHODS Patients 13 to 54 years old with each of 17 primary HIV infection symptoms, as well as other reported reasons for their visit consistent with primary HIV infection, were identified from the 2000 National Ambulatory Medical Care and National Hospital Ambulatory Medical Surveys to provide the denominator for the prevalence estimate. These survey data can be extrapolated to represent 90% of all US ambulatory care visits, including those to physician’s offices, emergency departments, and hospital clinics. Patients with symptoms and diagnoses inconsistent with a viral illness were excluded. The estimate for the numerator was derived from Centers for Disease Control and Prevention estimates and the medical literature.

RESULTS Patients complaining of fever and other visit reasons consistent with primary HIV infection had a disease prevalence of 0.66% (0.57%-1.02%), those with rash had a prevalence of 0.50% (0.31%-0.82%), and those with pharyngitis had a prevalence of 0.16% (0.11%-0.22%). Patients with other symptoms represented numbers of visits insufficient for reliable estimates of their prevalence.

CONCLUSIONS These estimates of the prevalence of primary HIV infection in ambulatory patients with fever, rash, and pharyngitis can aid with development of clinical testing guidelines and clinical decisions around testing for acute HIV infection.


INTRODUCTION

Ninety percent of the estimated 40,000 new HIV infections in the United States each year are associated with the acute HIV syndrome. One to 4 weeks after infection, patients with the acute HIV syndrome experience symptoms of a viral illness, such as fever, fatigue, pharyngitis, myalgias, rash, and weight loss, as well as other nonspecific symptoms. Most of these patients seek medical care, but seldom do they have primary HIV infection diagnosed.

The diagnosis of primary HIV infection has important clinical and public health implications. Patients who do not have their condition diagnosed at this early stage of infection will often seek care much later for acquired immunodeficiency syndrome (AIDS), when treatment may not be as effective. Diagnosis of primary HIV infection represents an important opportunity to prevent transmission to others, because in the early stages of the disease, patients have high levels of viremia, which, coupled with a lack of awareness of their diagnosis, can lead to transmission through sexual activity.

The prevalence of primary HIV infection in patients visiting outpatient facilities in the United States is largely unknown. One recent prospective
study determined a prevalence of 1% for patients with viral symptoms and at least 1 risk factor seeking care at an urban urgent care center, a setting in which a high prevalence of HIV infection is expected.\textsuperscript{10} Other studies have concluded that no symptom is sufficiently sensitive and specific to allow for targeted testing.\textsuperscript{4,5} It is difficult to determine which persons have a substantial probability of primary HIV infection when evaluating risk factors alone.\textsuperscript{11} Furthermore, patients may not be aware of their risky behavior.\textsuperscript{12}

An estimate of the national prevalence of primary HIV infection in patients seeking care at ambulatory settings, regardless of risk factors, would be important for clinical care in developing testing guidelines. Our objective for this study was to calculate a national estimate of the prevalence of primary HIV infection for patients visiting physician’s offices, emergency departments, and hospital outpatient clinics with symptoms consistent with primary HIV infection regardless of risk factors.

METHODS
This study is an analysis of the 2000 National Ambulatory Medical Care Survey (NAMCS) and the 2000 National Hospital Ambulatory Medical Survey (NHAMCS) to identify ambulatory patients with symptoms and diagnoses consistent with primary HIV infection to develop a denominator for a national ambulatory prevalence estimate. The numerator for the prevalence was derived from the Centers for Disease Control and Prevention (CDC) estimate of annual new cases of HIV infection and modified by reports in the medical literature regarding the percentage of patients newly infected with HIV who have specific symptoms, as well the percentage that seek medical care during the phase of acute infection.

Target Population
The target population is US patients with symptoms of primary HIV infection seeking care at physician offices, hospital emergency departments, and hospital outpatient clinics. This population is represented in the 2000 NAMCS and 2000 NHAMCS. These surveys are national probability samples of nonfederal office-based physicians, hospital emergency departments, and hospital outpatient departments.\textsuperscript{13,14} The NAMCS includes only visits to physician offices. The NHAMCS includes 2 separate databases: 1 for emergency departments and 1 for hospital outpatient clinics. Both surveys are conducted annually by the National Centers for Health Statistics and use a multistage probability sample design that is stratified and weighted to allow for population estimates of annual visits. Selected entities complete a survey form for visits during a randomly selected 1-week period. Combined, the surveys represent about 90% of US ambulatory visits with the exception of federal, veterans, and military outpatient facilities. The analysis is restricted to patients aged 13 through 54 years, inclusive, as they account for 95.2% of new HIV cases based on surveillance data from 25 states.\textsuperscript{15}

Prevalence Estimate: Denominator
A list of 17 symptoms consistent with primary HIV infection was generated from 3 US studies by including any symptom reported by more than 25% of patients with a diagnosis of primary HIV infection (Table 1).\textsuperscript{2,4,5} Both the NAMCS and NAMHCS include variables for up to 3 reasons for the visit. The NAMCS and the hospital outpatient clinic file in the NHAMCS contain a variable that allows for designation of acute visits. Emergency department visits in the NHAMCS are assumed to be for acute problems. The databases were analyzed separately for each of the 17 primary HIV infection symptoms listed in Table 1 to generate the total number of annual patients with an acute problem and each symptom as 1 of the 3 reasons for their visit. The total number of patients reporting each symptom was refined to include only those for whom the other 2 of their top 3

<table>
<thead>
<tr>
<th>Table 1. Primary HIV Infection Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptom Reported in the Literature</strong></td>
</tr>
<tr>
<td>1. Fever</td>
</tr>
<tr>
<td>2. Pharyngitis</td>
</tr>
<tr>
<td>3. Lymphadenitis</td>
</tr>
<tr>
<td>4. Myalgia</td>
</tr>
<tr>
<td>5. Arthralgia</td>
</tr>
<tr>
<td>6. Fatigue</td>
</tr>
<tr>
<td>7. Night sweats</td>
</tr>
<tr>
<td>8. Nausea</td>
</tr>
<tr>
<td>9. Vomiting</td>
</tr>
<tr>
<td>10. Diarrhea</td>
</tr>
<tr>
<td>11. Rash</td>
</tr>
<tr>
<td>12. Weight loss</td>
</tr>
<tr>
<td>13. Oral ulcer</td>
</tr>
<tr>
<td>14. Headache</td>
</tr>
<tr>
<td>15. Neck stiffness</td>
</tr>
<tr>
<td>16. Loss of appetite</td>
</tr>
<tr>
<td>17. Malaise</td>
</tr>
</tbody>
</table>

Note: Symptoms reported by more than 25% of patients in Schacker et al,\textsuperscript{7} Hecht et al,\textsuperscript{8} and Daar et al.\textsuperscript{1} HIV = human immunodeficiency virus; NAMCS = National Ambulatory Medical Care Survey; NHAMCS = National Hospital Ambulatory Medical Care Survey.
Prevalence of Primary HIV Infection

symptoms were also consistent with primary HIV infection. For example, patients reporting fever as their first reason for the visit would be included in the estimate for that symptom if their second and third reasons for the visit were fatigue or pharyngitis or other symptoms consistent with primary HIV infection as listed in Table 1. Whereas, if fever was their second or third reason for the visit, then the first and third or first and second other 2 reasons for the visit would need to be for symptoms consistent with primary HIV infection. A few other visit reasons not listed in Table 1 but considered consistent with a viral illness were also considered inclusion variables. These were enlarged glands, difficulty swallowing (as a symptom of throat pain), and unspecified pain.

NAMCS and NHAMCS contain data for up to 3 diagnoses for each visit. Primary diagnoses, but not secondary or tertiary ones, considered inconsistent with any of the constellation of primary HIV infection symptoms in Table 1 were excluded to further define the estimate.

The data were analyzed with the Statistical Export and Tabulation System (SETS 2.0), US National Center for Health Statistics. Relative standard errors, the measure of the sampling variability that occurs by chance, were calculated based on National Center for Health Statistics’ coefficients and formulas as reported in the NAMCS and NHAMCS.

Prevalence Estimate: Numerator
The numerator for disease prevalence was derived from the CDC estimate of the annual number of new HIV infections of 40,000 per year. In addition to the 13- to 54-year age restrictions (95.2% of new HIV diagnoses), this number was further decreased by 3 other factors. First, the 40,000 new HIV infections estimate was reduced by 10%, because the NAMCS and NHAMCS represent only 90% of US annual ambulatory visits. Second, an estimate of the percentage of patients with primary HIV infection reporting each symptom was derived from 127 patients who had primary HIV infection diagnosed from 3 US studies. Percentage estimates were calculated from all patients with primary HIV infection, regardless of whether acute HIV syndrome was reported, and included 95% confidence intervals. Third, the numerator estimate needed to be decreased by the percentage of patients with primary HIV infection who sought medical treatment. Only 2 studies, both of predominately urban gay men, reported findings for patients with primary HIV infection symptoms who were seeking care. Both studies reported rates close to 90%. Our analysis used the lower rate of 50% because of limited data on care-seeking rates for patients from other segments of the population, such as persons of low socioeconomic status, who may be less likely to seek medical care.

RESULTS
Prevalence Denominator Estimates
The annual number of patients with symptoms of fever, rash, and pharyngitis, as well as other symptoms and diagnoses consistent with primary HIV infection symptoms, is displayed in Supplemental Tables 1, 2, and 3 (available online only at http://www.annfammed.org/cgi/content/full/3/5/400/DC1). Data detailing exclusionary diagnoses and symptoms in the 3 practice settings represented by NAMCS and NHAMCS are available from the authors. The analysis of the other 14 symptoms from Table 1 showed an inadequate number of patient visits to allow for accurate annual estimates. The combined total number of patient visits for fever and other primary HIV infection symptoms and diagnoses consistent with an acute viral illness from these 3 settings was 2,175,551; there were 6,415,111 such patient visits for pharyngitis and 1,344,060 visits for patients with a rash.

The analysis provided data on which diagnoses were most likely to cause patients to report primary HIV infection symptoms. For example, 46% of patients visiting physician’s offices with complaints of fever and other primary HIV infection symptoms had streptococcal sore throat, acute pharyngitis, or influenza diagnosed, whereas in the emergency department unspecified viral infection, acute pharyngitis, and pyrexia of unknown origin accounted for 55% of the annual qualified visits. In the hospital clinic population, 75% of qualified visits were for unspecified upper respiratory tract infections, acute pharyngitis, and streptococcal sore throat.

Prevalence Numerator Estimates
Estimates for the annual number of patients aged 13 to 54 years with primary HIV infection seeking care in physician offices, emergency departments, and hospital clinics with complaints of fever, rash, or pharyngitis are listed in Supplemental Table 4 (available online only at http://www.annfammed.org/cgi/content/full/3/5/400/DC1). An estimated combined total of 14,394 patients newly infected with HIV who had fever, 8,568 who had pharyngitis, and 7,540 who had rash were seen in these 3 settings in 2000.

Prevalence Estimates
The prevalence estimates for patients aged 13 to 54 years with symptoms of fever, rash, and pharyngitis, as well as other symptoms and diagnoses consistent with primary HIV infection, are displayed in Table 2. Patients with fever had the highest rate of primary HIV infection, 0.66% or 6.6 cases per 1,000; the rate for those with rash was 0.56%, or 5.6 cases per 1,000; and for those with pharyngitis the rate was 0.13%, or 1.3 cases per 1,000. Prevalence estimates include a range...
Table 2. Primary Human Immunodeficiency Virus Infection Estimates

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Numerator Estimate (95% CI)</th>
<th>Denominator Estimate (95% CI)</th>
<th>Prevalence Estimate (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever</td>
<td>14,394 (13,880-15,936)</td>
<td>2,175,551 (1,723,031-2,628,071)</td>
<td>0.66 (0.53-0.92)</td>
</tr>
<tr>
<td>Rash</td>
<td>7,540 (5,998-9,082)</td>
<td>1,344,060 (966,721-1,721,399)</td>
<td>0.56 (0.35-0.94)</td>
</tr>
<tr>
<td>Pharyngitis</td>
<td>8,568 (7,026-10,436)</td>
<td>6,415,111 (5,528,853-7,301,369)</td>
<td>0.13 (0.10-0.19)</td>
</tr>
</tbody>
</table>

Note: Prevalence estimates were obtained by dividing the numerator estimate by the denominator estimate. The low range estimate was obtained by dividing the low numerator estimate by the high denominator estimate, and the high range estimate by dividing the high numerator estimate by the low denominator estimate.

Sensitivity Analysis

Sensitivity analysis involves recalculating results when assumptions about variable estimates are altered because of uncertainty. In this analysis the most uncertain estimate was the percentage of patients with primary HIV infection who sought care. The baseline analysis assumed that 50% of patients with primary HIV infection would seek medical care. Data are lacking, however, outside the urban, gay, male populations, for which the rate is 90%. If this assumption were varied from 25% to 90%, then for patients with fever the prevalence would vary from 0.33% to 1.19%, for rash from 0.28% to 1.01%, and for sore throat from 0.07% to 0.23%.

DISCUSSION

This analysis provides an initial estimate of the national prevalence of primary HIV infection in patients visiting ambulatory settings with complaints of fever, rash, or sore throat and a diagnosis consistent with an acute viral illness. A national prevalence estimate of primary HIV infection provides a useful starting point for policy decisions regarding early detection and referral programs, as emphasized in a recent CDC initiative. For instance, although current counseling, testing, and referral guidelines recommend routine HIV testing in populations with a prevalence of 1%, a more recent cost-effectiveness analysis has shown that in the era of highly active antiretroviral treatment, testing may be cost-effective when the prevalence is as low as 0.05%, ie, rates lower than those found in this study for patients with fever or rash or pharyngitis. The only previous study to develop a prevalence estimate was a prospective study of 499 patients visiting an urban urgent care center in Boston. The prevalence was 1% for patients complaining of viral symptoms who had at least 1 risk factor for HIV. The prevalence in that population would be expected to be higher than a national estimate based on patients from all settings with similar symptoms regardless of risk factors and thus corroborates these results.

This study has several limitations because of assumptions made in the analysis. The most uncertain estimate involves the percentage of patients that actually seek medical care during the symptomatic phase of primary HIV infection. Although the baseline assumption of 50% is reasonable because patients with primary HIV infection are ill for 10 to 14 days, and previous reports of care-seeking behavior are 90% for urban gay males, the true percentage is largely unknown. Poor, disenfranchised, injection drug users who acquire HIV infection through needle-sharing behavior, for instance, may be less likely to seek care because they lack insurance or transportation. This issue was addressed through a sensitivity analysis that lowered the rate of those seeking care to 25%, more one third less than that reported in the literature.

Several factors could have affected the denominator estimate. Excluding primary diagnoses that included a patient presenting with primary HIV infection symptoms could decrease the denominator and falsely increase the prevalence estimate. To minimize this potential source of error, all diagnoses possibly construed as viral (Supplemental Tables 1, 2 and 3 [available online only at http://www.annfammed.org/cgi/content/full/3/5/400/DC1]) were included in the denominator. Additionally, other diagnoses unlikely to represent viral infections but not entirely inconsistent with primary HIV infection symptoms, such as tension headache, migraine headache, unspecified abdominal pain, and gastroesophageal reflux, were included as well.

Another factor that could affect the denominator estimate is the restriction of the analysis to primary diagnoses. The analysis was concerned with only acute problems. In line with this study assumption, diagnostic coding guidelines specify that the primary diagnosis should reflect the patient’s main reason for the visit. It is possible, however, especially in private offices with familiar patients, that a chronic problem was coded first even though the patient was seeking care for an acute illness. Missing these visits would decrease the denominator estimate. To assess the impact of this source of
potential error, the data were analyzed for visits to physician’s offices for patients reporting fever. A total of 17,824 patients, or 1.75% of the 1,017,573 patients included from this data file, had secondary or tertiary diagnoses consistent with primary HIV infection. Increasing the denominator estimate by 1.75% would minimally decrease the prevalence for febrile patients from 0.66% to 0.65%.

Last, although an attempt was made to include the correct codes for the primary HIV infection symptoms reported in the literature and their likely synonyms, it is possible that some symptoms were coded for visit reasons not included in the analysis, whereas some of the inclusion symptoms, such as chills and feeling hot as possible synonyms for fever, could have overestimated the denominator estimates.

In conclusion, this study is the first to estimate the national prevalence of primary HIV infection. Prevalence estimates were developed from national ambulatory databases for patients aged 13 to 54 years regardless of risk factors who sought care for fever, rash, or sore throat and had an acute illness consistent with a viral infection. We believe these estimates will be useful in developing national testing guidelines that could ultimately aid in decreasing HIV transmission rates and improving the health of those infected through early entry into medical care.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/400.

Submitted January 30, 2005; submitted, revised, April 28, 2005; accepted May 10, 2005.

Key words: Primary HIV infection/epidemiology; prevalence studies; disease frequency surveys

Funding support: This study was supported by a Claire Weaver Scholar-ship for Research in Family Medicine.

References


‘Coming Down the Line’— Patients’ Understanding of Their Family History of Common Chronic Disease

Fiona M. Walter, MA, MSc, FRCGP
Jon Emery, DPhil, FRACGP

1 General Practice & Primary Care Research Unit, Department of Public Health & Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK
2 General Practice, School of Primary, Aboriginal and Rural Health Care (SPARHC), University of Western Australia, Claremont, Western Australia, AU

ABSTRACT

PURPOSE The family history is becoming an increasingly important feature of health promotion and early detection of common chronic diseases in primary care. Previous studies of patients from genetics clinics suggest a divergence between how persons with a family history perceive and understand their risk and the risk information provided by health professionals. This interview study aimed to explore how patients in primary care understand and come to terms with their family history of cancer, heart disease, or diabetes and how family history might affect consultations about disease risk and management.

METHODS Thirty semistructured interviews were conducted with general practice patients who had a family history of cancer, heart disease, or diabetes. The transcript data underwent a qualitative constant comparative analysis.

RESULTS What exactly constitutes having a family history of an illness varied among participants. The development of a personal sense of vulnerability to the illness in the family depended not only on the biomedical approach of counting affected relatives but also on a sophisticated interplay of other factors. The emotional impact of witnessing the illness in the family, particularly if the illness was sudden, premature, or fatal, and the nature of personal relationships within a family that determine a sense of emotional closeness and personal likeness with the affected relative, all contributed to the perception of disease risk. Different beliefs about the contributions of nature and nurture to disease can affect patients’ views on the degree of control they can exert over their risk.

CONCLUSION This study highlights potential differences between the way patients and medical professionals assess and understand familial risk of cancer, heart disease, and diabetes. Our previous systematic review findings are enhanced by showing that personal experience of disease and the emotional impact can also influence familial risk perceptions. Eliciting the patient’s perspective when discussing risk of chronic disease, particularly in the context of a family history, could inform a more patient-centered approach to risk assessment and communication and support patients to make informed decisions about the management of their disease risk.


INTRODUCTION

Taking a family history is an increasingly important component of chronic disease prevention in primary care. In the United Kingdom, the NHS National Service Frameworks (guidelines) for coronary artery disease, diabetes, and cancer1 highlight the relevance of the family history as part of risk assessment and management. Accurate assessment and effective communication of familial risk enables appropriate reassurance for those who are at population risk and the discussion of treatment options or behavior changes for those at increased risk.2 The role of the
family history in clinical practice is therefore likely to expand from providing the traditional psychosocial insights into the context for a patient’s symptoms of disease to also include risk assessment and management. Effective risk communication is a complex process that could be facilitated by adopting a patient-centered approach of incorporating patients’ ideas, feelings, and concerns.

Social science research suggests that health professionals and patients may hold different views about health and illness: while the professional perspective is informed by biomedical models, the patient’s perspective is more likely to be based on their illness experiences. Knowledge about patterns of inheritance are part of family culture in Euro-American and other cultures in which a wide range of diseases and characteristics are believed to run in the family. Clinicians should be aware of how their patients’ understand their family history, as this understanding will influence perceptions of both their risk of the disease and its management.

Disease-related risk perceptions, as well as the relationship between risk perception and health behaviors, have been studied extensively. Structured models have been developed to integrate differing health beliefs and to understand their role in predicting health-related behaviors. Among the most influential models is Leventhal’s seminal theoretical work, the Common Sense Model of Self Regulation (CSM), which arose from the observation that the medical definition of symptoms represented only one type of perceptual information needed to appraise a health-risk situation. The biopsychosocial factors that influence the representations of each of its key components were also identified: perceived health threat, potential health behavior(s), and perceived impact of engaging in those behaviors. The CSM has been widely used to examine how patients evaluate health threats by constructing their own representations or perceptions that influence their patterns of coping. If a perceived family medical history is viewed as a health threat, then the CSM may also be used to interpret patients’ understanding of their family history of common chronic disease.

We recently reported a synthesis of qualitative studies that explored how people understand a family history of cancer, coronary artery disease, or diabetes. We developed a theoretical framework to explain the processes by which a person develops and deals with personal perceptions of disease risk. It reflects several of the constructs of the CSM. Most of the studies included in our previous review were based on patients sampled from specialist care whose beliefs may have been altered by some form of genetic counseling. This qualitative study tests the theoretical model with patients from general practice. We aimed to explore understanding about family history and familial risk of common chronic disease in primary care patients.

The study was approved by the Cambridge Local Research Ethics Committee.

METHODS
Participants
Our sampling strategy aimed to gain the broadest view possible by reflecting a range of age, sex, social class, educational levels, and degree of familial risk. Participants were recruited from 2 Cambridgeshire general practices: practice 1 (list size 2,236; 30% younger than 18 years; 1% older than 75 years) was located in a developing semiurban new town, while practice 2 (list size 10,564; 21% younger than 18 years; 8% older than 75 years) was located in the city of Cambridge. Electronic searches of practice medical records looked for patients aged 18 years and older with a first-degree relative with either cancer, coronary artery disease, or diabetes. In each practice 30 patients with a family history of each disease were randomly selected by the EMIS clinical computer system, generating a total sample of 180. General practitioners (GPs) excluded patients who were unable to communicate in English or had severe mental disability or life-threatening disease. Patients were approached by a single letter from their GP.

Interviews
Semistructured qualitative interviews lasting about 1 hour were conducted mainly by the first author in the interviewees’ homes between October 2002 and March 2003. The interview guide was informed by our systematic review and pilot tested with the first 5 participants (the interview schedule can be found online only in Supplemental Table 1, available at: http://www.annfammed.org/cgi/content/full/3/5/405/DC1). Considerable flexibility during the interviews allowed participants to discuss issues that were most important to them. Participants gave details about their relatives’ health or cause of death to initiate discussions about their family history and its personal meaning.

Analysis
Audiotapes of the interviews were fully transcribed, and analyzed manually, supported by NUD*ST software. Although we were testing the theoretical model, we were concerned that using an analysis strategy, such as framework analysis, may have obscured emerging and previously unrecognized themes. We therefore applied a constant comparative technique to allow the emergence of themes and development of underlying concepts, and later we mapped the emerging concepts onto the theoretical model. This approach allowed
both the confirmation of the main constructs and the identification of further concepts leading to changes and refinement of the model. Analysis began during data collection to inform subsequent interviews, which were continued until data saturation was achieved. Analysis was conducted primarily by the first author, with half the transcripts independently read by the second author to confirm the integrity of the emerging themes and concepts. The quotations that follow were chosen to reflect a range of both consensual and dissenting views.

**RESULTS**

**Participants**

Eighty-nine patients were approached from practice 1, of whom 20 responded; 90 patients were approached from practice 2, and 24 responded, giving an overall response rate of 25%. Telephone contact with each respondent confirmed they had at least 1 first-degree relative with cancer, coronary artery disease, or diabetes. Thirty-two patients were available for interview, and 2 dropped out at the time of interview, 1 because of personal illness and 1 moved away. By the time 30 interviews were completed, no new themes were emerging. Participants’ characteristics are shown in Table 1.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td></td>
</tr>
<tr>
<td>Practice 1</td>
<td>12 (40)</td>
</tr>
<tr>
<td>Practice 2</td>
<td>18 (60)</td>
</tr>
<tr>
<td>Family history of disease</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>14 (47)</td>
</tr>
<tr>
<td>Heart disease (coronary artery disease)</td>
<td>15 (50)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
</tr>
<tr>
<td>20 – 39</td>
<td>12 (40)</td>
</tr>
<tr>
<td>40 – 59</td>
<td>14 (47)</td>
</tr>
<tr>
<td>60+</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Sex, female</td>
<td>16 (53)</td>
</tr>
<tr>
<td>Ethnic origin</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>28 (93)</td>
</tr>
<tr>
<td>Other (Japanese, Iranian)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Single, widowed, divorced</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Married, living with partner</td>
<td>26 (87)</td>
</tr>
<tr>
<td>Children</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Yes</td>
<td>23 (77)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Primary education only</td>
<td>2 (7)</td>
</tr>
<tr>
<td>2 = Some secondary education</td>
<td>7 (23)</td>
</tr>
<tr>
<td>3 = Completed O-levels (to 16 y)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>4 = Completed A-levels (to 18 y)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>5 = Further education</td>
<td>9 (30)</td>
</tr>
</tbody>
</table>

**Understanding Family History**

Our results confirm the constructs of the theoretical model derived from specialist care, and further themes emerged that have been used to refine the model (Figure 1). Once a person acknowledges that an illness runs in the family, the family history grows in meaning through many routes until there is gained a sense of vulnerability to the disease, which the person then attempts to cope with or control. Concepts about personal relationships within a family that determine a sense of risk, the effects of personal experience of familial illness on risk perception, and strategies patients apply to cope with or control their familial risk were themes that took on particular importance.

**Confirming the Familial Risk Model**

**It Runs in My Family**

The great majority of participants viewed cancer, coronary artery disease, or diabetes as “running in the family” or “coming down the line.” What exactly constituted having a family history of an illness varied among participants: whereas it meant having more than 1 affected relative to most participants, several considered an illness to run in the family when only 1 relative had suffered the illness. This finding occurred across all 3 diseases studied.

“The heart disease does seem to run in our family, because I lost my mother and aunt to it, and I’ve just lost my cousin to it, but not so much the cancer, because my dad was the only one that’s had it” (ID03; female, 42 years).

“I know with cancer in my family, my grandparent who had it was on my dad’s side, but my mum had it and my sister had it, and you sort of think, mm’hhm, 2 out of 3 of us, this is quite high, is it coincidence or something else’” (ID16; female, 52 years).

“I feel more at risk because of my father, so [diabetes] is in the family, but only my father is affected. I think genetically there is a weak link there: it could
miss me altogether, and that’d be fine, but then it could be there as well” (ID22; female; 45 years).

Explanations for why they believed a disease ran in the family were not solely based on inherited mechanisms. The differing contributions of nature and nurture were widely discussed: most participants described a multifactorial model of familial risk, and few participants viewed the development of illness as solely due to inheritance. Environmental or lifestyle causes for a relative’s illness were often reported, but when there was no obvious cause for a relative’s illness, inherited factors were more likely to be implicated.

“Well there are certain traits you know, familial traits, and I think they’re probably causing illnesses, but I certainly don’t think it’s the whole picture, I think that environment and nurture play a part as well” (ID06; female; 58 years).

“My father has been diagnosed with diabetes, and I believe that was triggered off by 2 factors actually, by diet because he was overweight at one stage and he did eat a lot of the wrong food, but also he had a very stressful time in his life that came about through a work situation, and I think that stress factor was the trigger to bring on the diabetes” (ID22; female; 45 years).

Some acknowledged that shared environmental, as opposed to genetic, factors could explain their family history, and a few discussed the contribution of chance or bad luck. For a few participants this developed into a fatalistic attitude of inevitability developing the illness in their family.

“My sister and I, we are doomed: dad’s had a quadruple bypass, and his father had strokes, and there’s high cholesterol on his side of the family” (ID07; female; 44 years).

“I suppose I’m a bit of a fatalist: I’ve smoked since I was 14, so I would have thought if there’s anything gonna happen to me, that’s something that had to happen” (ID21; male; 59 years).

About one third of participants spontaneously mentioned genes, usually in connection with transmission of traits or illnesses. When their understanding about genes was explored further, however, most participants...
were vague or even defensive. Understanding the mechanisms of inheritance was important only to a minority of concerns about the risk of familial illness for their children.

“If you inherit bad genes, then you are predisposed to have something wrong, and if you inherit strong genes, then you’ll prosper, and it’s not quite so straightforward to say just because my father’s got it, right, because that’s only half of where I come from” (ID09; male; 52 years).

“I suppose my knowledge is very limited except I just think it’s something that transmits from one generation to another. It’s like a blueprint…. I feel as if I’m doing my A levels’ (ID05; female; 60 years).

Routes to Feeling at Risk

Acknowledging that a family history of a disease does not equate with feeling personally at risk for the illness. The route to feeling at risk describes how the family history gains personal meaning, such as the emotional impact of witnessing a relative’s illness or death and the course of that illness.

Emotional impact of witnessing a relative’s illness. Experiencing a relative’s illness, recovery, or death, particularly a parent or sibling, contributed to the participants understanding of the illness itself as well as perceptions of personal risk. Participants described their relative’s illness in a chronological and detailed way, long after the event, including the emotional effects of witnessing the illness, guilt at being unable to alleviate suffering, or regret about the limited time spent with their relative before their death.

“My father died of cancer, so we were involved in it quite dramatically, and I think when it’s a family member, you never forget. I think when it’s somebody from the outside, then for a period of time, you get this immediate shock, then it sort of fades. But when it’s your father, you never forget what actually happened” (ID09; male; 52 years).

The impact of witnessing illness was not confined to cancer but was also discussed in relation to coronary artery disease and diabetes. Witnessing the illnesses of more distant relatives or friends added to their understanding of the illness but had less impact on their feeling at risk.

“There’s the thing of the genes again: if it’s the relative … I could have similar things to them. We don’t really think about that with a friend” (ID19; female; 27 years).

A few participants felt that witnessing an illness alone did not affect their sense of personal risk and that only by developing the illness themselves would they fully understand its importance.

“A heart attack is something that me mum’s talked about, and she’s told me about [her own], but it’s not something that I’ve had any experience of, you know, personally. You sort of tend to shunt that to the background a bit, you know” (ID08; male; 39 years).

Course of the relative’s illness. The onset, course, and outcome of the relative’s illness further contributed to the participants’ feeling at risk. Sudden illness, particularly sudden death, prolonged illnesses, or a silent onset of the illness, were important routes to feeling at risk:

“[My mum] was far too young at 45 to be dying from a heart attack. It wasn’t as though she had a number of warnings, it was very sudden and that’s it, just had a heart attack one day. That is much more of a threat hanging over you I suppose” (ID17; female; 38 years).

The perception that the death was premature gave it more weight and was applied even when relatives died in their 60s or 70s.

“My father was the most youthful of his generation, never a day’s illness, so to die at 65 is young, you know, he was taken too young. He still had lots of things he could do, I mean for others as well for themselves, and that’s the sad thing, because he was so energetic, squash playing, full of life, getting things out of people…. That is for us a huge part of the loss” (ID18; male; 37 years).

Symptoms or signs that appeared late in the course of the disease or delayed diagnoses, with sometimes a much stormier course of illness, also contributed to the personal meaning of the family history.

“[My mother] was 47 or so when she died, but we never knew where [the cancer] originated from, it was lots of missed diagnoses, and at the time it was diagnosed it was sort of everywhere so … no idea where it started from. And really there’s nothing you can do” (ID16; female; 52 years).

Survival and return to a nearly normal lifestyle, especially for coronary artery disease and diabetes, meant the illness and their family history were less threatening.

“I think it probably seems less serious to have heart trouble [than cancer], as the majority of my relatives have survived, and 2 of them are living relatively healthily. So there is something about outcomes as well in a way” (ID14; male; 31 years).

Patterns within the family history. Participants sometimes tried to identify patterns within their family history when considering their own risk. Examples included patterns of life events, such as age of onset of illness, or the sex of family members affected by the illness.

“[Cancer] definitely is a female thing on my side, in my mind. When I came up to the same sort of age as my Mum [when her breast cancer was diagnosed], it
was a very tense time. I think once I'd gone over 50 I felt safer almost” (ID16; female; 52 years).

**Refining the Familial Risk Model**

**Routes to Feeling at Risk**

Further themes emerged that underpin how the family history gains personal meaning. Participants described 2 related concepts about personal relationships within a family that were important influences along the route to feeling at risk. These relationships may be countered by factors that limit the personal meaning of the family history.

**Perceived Closeness.** Perceived closeness describes an emotional bond and continuity of their relationship, even when physically separated. Closeness often underpinned the emotional impact of witnessing a relative’s illness.

“Dad moved down to London, and we’d go down to see him a couple of times a month, so we were close, you know, in seeing each other and talk, but we weren’t close, you know touchy-feely close, which I do regret now that he has gone” (ID29; female; 33 years).

“I shared a lot of characteristics with [my mother], but I didn’t identify in quite the same way as with [my father], although I was close to her” (ID06; female; 58 years).

The converse of closeness limited the personal importance of a relative’s illness to the extent that the participant was unable to recall events in a relative’s life or the cause of death.

“I’m not quite sure what she died of to be honest. It’s a terrible thing to say, isn’t it? We weren’t particularly close” (ID24; male; 37 years).

**Likeness.** The concept of likeness describes similarities with family members, particularly parents, not simply in terms of physical characteristics, but also in terms of personality, mannerisms, and feelings. Many participants felt more at risk of developing their parent’s illness because of perceived likenesses with the affected parent. Likeness often extended from parents to include an overall affinity with either the maternal or paternal side of the family. This likeness to one side of the family was often linked to beliefs about illness risk.

“As I’m getting older I’m really starting to look like [my mum] now, and feel like her. I guess that it makes me think I am like her, and maybe I’m going to get the same as her” (ID09; female; 38 years).

“[The heart disease] definitely seems to be more in the male side of my family as opposed to the maternal” (ID24; male; 37 years).

Likeness was not obviously linked to the patient’s sex, despite such illnesses as breast cancer being sex specific. Likeness of personality, mannerisms, and “ways to do things” were more important than physical similarities in determining a personal sense of illness risk (as shown in Supplemental Table 2, which is available online only at http://www.annfammed.org/cgi/content/full/3/5/405/DC1). There were no participants who described feeling that a lack of similarity with an affected parent protected them from the illness in the family.

**Reducing the personal meaning of the family history.** When considering the personal meaning of their family history, some participants sought ways to minimize their sense of vulnerability through bargaining, and identifying counterexamples.

“I always had the feeling that I would get [Parkinson’s like my father]. So I suppose maybe it was a bargain, you know, I don’t want to get breast cancer [like my mother] but I’m happy to get Parkinson’s, because I’ve had a lot of friends who’ve had breast cancer and it’s really awful” (ID06; female; 58 years).

“I don’t think you can ever prevent [the heart disease in my family] because I’ve nursed people who’ve had cardiac arrests running the London marathon, and they were obviously very fit people” (ID11; female; 38 years).

The existence of a current illness and increasing age appeared to reduce the threat of their family history to the degree that their concerns shifted toward their current personal health and that of future generations.

**Vulnerability in other family members.** Despite the belief that a familial tendency had come “down the line” to themselves, only a minority of participants expressed concern that the familial illness would go on down the line to their children.

“You never think about your kids having [heart disease like my Dad], just don’t, although I must admit I fully expected for [my child] to get the asthma because my sister’s got 4 children and 3 of them had it” (ID08; male; 39 years).

“I think I worry whether I’ve got anything that I don’t want to pass on to my children. They all know that I’ve been on this genetic thing [screening for bowel cancer]. I don’t want to worry them too much, to feel that they may have to go through the same process” (ID07; female; 44 years).

While not necessarily considering the next generation, some participants identified other family members, most commonly siblings, as being more at risk, particularly those who had a family history of coronary artery disease or diabetes, perhaps because environmental or lifestyle factors were seen as more important in triggering the familial illness than for cancer.

“I think [my brother’s] a classic example of being an at-risk [for heart disease] person because to me he’s just a younger version of my Dad. He smokes, he goes out for a drink, and he’s in quite a high-up job which
is very, very stressful, and I would say his weight isn’t ideal, a bit over weight” (ID29; female; 33 years).

**Controlling the Familial Risk: ’How Can I Reduce My Risk?’**

Attempts were made to take control of the threat of a family history through a variety of means: changing behavior, medication, screening, and obtaining information about the disease. Notions of fatalism, however, sometimes countered their beliefs in the ability to control their disease risk.

“Some things we don’t have a choice about, but there are choices that we can make about keeping us in a healthy way…. I think that it basically comes back to just living a balanced normal lifestyle, and I believe if you add to it continuously by being overweight or not exercising at all or drinking too much or smoking, you are putting yourself more at risk” (ID22; female; 45 years).

Perhaps given the greater emphasis on healthy behaviors for cardiovascular disease prevention, there was a greater sense of empowerment to control disease risk in those with a family history of coronary artery disease and diabetes than in those with a family history of cancer:

“There are types of cancer that you know in the end you can’t beat. It’s almost like it seems to be more of a threat [than heart disease], it’s more worrying” (ID04; female; 56 years).

“I believe you have the power, if you like, to decide what your future direction is going to be. I am not going to die at 54 of a heart attack like my Mum did because I told myself that I have to do something about it to make sure I don’t. It’s a bit as though, yes you can inherit certain things, but then you take charge of it yourself” (ID17; female; 38 years).

Alternative strategies of control for participants with a family history of coronary artery disease and diabetes included medication or surgery, believing that this would provide them safety from their familial risk.

“I used to [feel at risk] until I had my heart by-pass, and now I’m taking medication I feel quite safe. Although anybody can have a heart attack any time, I feel a little bit safer now because I’ve had mine seen to, so it doesn’t worry me so much” (ID15; male; 62 years).

The perceived inability to alter cancer risk through lifestyle changes led to participants with a family history of cancer relying more on screening and self-examination or seeking control by keeping well informed.

“I would have wanted some sort of blood test to decide whether I was genetically more predisposed to this [breast] cancer or whether the history was just coincidental in the family. And what I worked out (which probably wasn’t accurate in retrospect) was that I couldn’t find out for sure whether it was a gene thing because I hadn’t got any living relatives with breast cancer … but I’m too scared to, actually, do breast checks on myself, I don’t actually explore to see if I’ve got any lumps” (ID05; female; 60 years).

“I think we just wanted to be as educated as we could about [our family history of breast cancer], as in having surveillance or screening or just finding out what all this was, yeah, I just wanted information to tell me what my options were. That’s just about as much you can do really, rather than maybe just being ignorant and not being informed” (ID10; female; 38 years).

**DISCUSSION**

This qualitative study of primary care patients explored how a person understands and makes sense of familial risk of common chronic disease. A theoretical model derived from specialist care was tested with general practice patients: not only were the constructs of the theoretical model confirmed, but also further themes emerged that have been used to refine the model, including concepts about personal relationships within a family that determine a sense of risk, the effects of personal experience of familial illness on risk perception, and strategies patients apply to cope with or control their familial risk (Figure 1).

Variations between patient perceptions and professional assessment of familial risk have been previously identified for coronary artery disease. This study shows that particular features occur across the spectrum of common chronic diseases. Whereas both health professionals and patients count the number of affected relatives and the age at which they were affected, patients’ understanding of familial risk can be informed by many more factors. The emotional impact of witnessing the relative’s illness, plus features of the course of the relative’s disease, were far more important to patients’ personal sense of disease risk. For instance, the notion of a premature death was widely held. Many patients minimized their perceived risk by balancing disease risk with counterexamples, such as the relative who smoked and lived to old age, or factors that could protect them, such as medication or diet.

Patients’ understanding about the causes of an illness in the family have an important effect upon a developing sense of personal risk. When considering what a family history means, most participants held a multifactorial model of familial risk, balancing the risks of nature and nurture. This study shows that understanding the mechanisms of inheritance was important only to the minority who were concerned about the risk of familial illness. Many more participants were concerned about their familial risk “coming down the
line,” referring to illnesses passing through the genera-
tions to themselves and their siblings. Fewer partici-
pants identified concerns that the risk of such diseases
might be transmitted to their children or future genera-
tions, seeing themselves as the terminus. This finding
contrasts with studies of patients from genetics clinics,
who are concerned about the risk to their offspring as
much as to themselves.18 Except for the rare subsets
of disease inherited in a Mendelian fashion (eg, familial
hypercholesterolemia, familial breast cancer BRCA1
and BRCA2), patients probably do not require in-depth
knowledge of genetic mechanisms to understand their
familial risk.19 Indeed, a multifactorial model is prob-
ably more appropriate for those with a family history
of common chronic disease.

In this study we identified 2 concepts about per-
sonal relationships within a family—closeness and
likeness—that affect understanding familial risk and
were discussed in detail by all participants. Closeness
to the affected relative underpinned the importance
of emotional bonds to developing a sense of personal
risk. The emotional impact of witnessing a relative’s
illness and the nature of the disease itself were impor-
tant determinants of risk perception,20 with a strong
negative impact leading to a higher sense of vulner-
ability. The emotional impact of a delayed diagnosis
or sudden death probably reflects similar affective
pathways to developing a personal sense of disease risk.
We therefore suggest that closeness is related to the
salience and perceived seriousness of the disease. On
the other hand, likeness was more related to perceived
susceptibility. Persons with a family history of coronary
artery disease have been shown to view familial risk as
linked to inheritance of physical characteristics.21,22 Our
findings suggest that similar views were held by those
with a family history of cancer or diabetes. Likeness or
personality and mannerisms were more important than
physical similarities in determining a sense of disease
risk among our sample in contrast to the existing lit-
erature. Patients also identified characteristics within
a particular side of the family that were not sex specific
which informed their personal sense of disease risk.

The impact of personal experience on risk percep-
tion has been described among persons who have a
relative with cancer,23,24 but the same does not hold
true for diabetes or coronary artery disease. Those
with a diabetic parent frequently underestimate their
personal risk and know little about preventive strate-
gies, such as diet and exercise.24 Persons with a family
history of coronary artery disease consider visible risk
factors such as smoking and weight to explain or pre-
dict coronary events and identify “coronary candidates”
in their family.25 There is also acceptance, however,
that these behavioral risk factors fail to explain some
“anomalous deaths” in persons with low-risk lifestyles,
and long “unwarranted” survival in those with high-risk
lifestyles.25 Our findings are the same for participants
with family histories of diabetes or coronary artery
disease: both groups used counterexamples or strategies
such as bargaining to minimize their perceived personal
risk. These tactics can undermine a person’s belief in
the value of modifying behaviors to reduce their risk.25

Fatalistic attitudes toward disease risk were fairly
common, particularly for cancer, which was perceived
to be under less personal control, with fewer modifi-
able lifestyle risk factors. Notions of fatalism, though,
were also used to account for continued risky lifestyles,
such as smoking despite a family history of coronary
artery disease. Representations of fatalism have been
described among persons with familial hypercholester-
olaemia,27 an autosomal dominant condition associated
with a high risk of coronary artery disease. Such fatal-
istic attitudes associated with beliefs about familial risk
may interfere with behavioral or lifestyle interventions
even in multifactorial disease.

Strengths and Limitations

Our study is the first to explore primary care patients’
views about their family history of 3 different common
chronic diseases. Few studies have examined patients’
understanding of familial risk of any chronic disease,
and previous research, particularly among those with a
family history of cancer, may have been influenced by
the setting of a genetics clinic. The choice of a qualita-
tive method to characterize the patients’ perceptions is
therefore not only appropriate but also novel in identi-
fying themes that are generalized across diseases.

The sampling strategy led to a broad group of
participants of varying age and educational levels.
Recruitment through general practice records meant
that we had to rely on random sampling, which led to
a more homogenous sample than we had hoped for.
Although we would have preferred a purposeful strati-
fied sampling strategy, it was not possible within the
restrictions of current UK ethics and research gover-
nance guidelines. Further strength was given by the
interviews continuing until there was saturation of data,
and analysis by 2 researchers increased validity. Given
that the low response rate and the location of the study
led to less ethnic variation than is found in the UK
today,28 these results may not necessarily be extra-
polated to other ethnic or cultural groups or countries.
Furthermore, taking part in the research process itself
may have increased or decreased participants’ concerns
about their risks.

The findings of our study have been fed back to the
participants in a pilot questionnaire for further valida-
tion, and the data generated will form part of a larger
quantitative study designed to measure the determinants of familial risk perception. This study will also address issues that emerged from these data, such as the influence of the “side of the family,” sex of affected relative, and concepts of closeness and likeness on perceived familial risk and changes in behavior.

Implications for Clinical Practice
Our findings highlight potential differences between the way patients and clinicians determine and understand familial risk of common chronic disease. Studies of women with a family history of breast cancer suggest they often overestimate their risk and have associated high levels of anxiety. Furthermore, although cancer risk counseling can improve knowledge, a systematic review of the literature failed to show improvements in risk perceptions. Family history of diabetes and coronary artery disease may not be perceived as an important risk factor by some patients or may be used to downplay the value of behavior change.

Testing the theoretical model in a primary care setting and across multiple chronic conditions served not only to confirm the original constructs but also to add new constructs and refine the model of familial risk perceptions. We have shown that personal experiences of disease and their emotional impact can have a major influence on a patient’s personal sense of vulnerability. This finding has considerable clinical value: health professionals may gain a better understanding of a patient’s risk perception by exploring beliefs and experiences of disease within the patient’s family. Such findings are consistent with the core values of the patient-centered clinical method and clinical narrative medicine.

Risk communication strategies are being developed to assist physicians communicate risks clearly and more effectively and thereby to build closer relationships with their patients. Computerized tools have also been developed to support risk assessment of certain cancers, coronary artery disease, and diabetes. Incorporating the specific beliefs and experiences that determine patients’ familial risk perceptions could improve their understanding about their risk and support informed decisions about the management of their risk through healthy behaviors and appropriate use of screening tests.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/405.

Key words: Primary health care; attitude to health; disease susceptibility; patient-centered care; risk factors; risk perception; medical history taking; patients’ understanding; family health; family history; cancer; coronary arteriosclerosis; diabetes mellitus

Funding support: Fiona Walter is funded by a NHS R& D Health Service Research Fellowship, and Jon Emery was funded by Cancer Research UK.

Acknowledgments: The authors would like to thank all the patients who kindly agreed to be interviewed, and the 2 practices for providing study patients. We also thank Professor Theresa Marteau and the 3 anonymous reviewers for their helpful comments about this paper.

References


Patient-Centered Communication and Diagnostic Testing

Ronald M. Epstein, MD1,2,3
Peter Franks, MD4
Cleveland G. Shields, PhD1,2,3
Sean C. Meldrum, MS1,3
Katherine N. Miller, BS1,3
Thomas L. Campbell, MD1,2,3
Kevin Fiscella, MD, MPH1,3,5
1Department of Family Medicine, University of Rochester Medical Center, Rochester, NY
2Department of Psychiatry, University of Rochester Medical Center, Rochester, NY
3Rochester Center to Improve Communication in Health Care, Rochester, NY
4Department of Family and Community Medicine, and Center for Health Services Research in Primary Care, University of California, Davis, Sacramento Campus, Sacramento, Calif
5Department of Community and Preventive Medicine, University of Rochester, Rochester, NY

Abstract

Purpose: Although patient-centered communication is associated with improved health and patient trust, information about the impact of patient-centered communication on health care costs is limited. We studied the relationship between patient-centered communication and diagnostic testing expenditures.

Methods: We undertook an observational cross-sectional study using covert standardized patient visits to study physician interaction style and its relationship to diagnostic testing costs. Participants were 100 primary care physicians in the Rochester, NY, area participating in a large managed care organization (MCO). Audio recordings of 2 standardized patient encounters for each physician were rated using the Measure of Patient-Centered Communication (MPCC). Standardized diagnostic testing and other expenditures, adjusted for patient demographics and case-mix, were derived from the MCO claims database. Analyses were adjusted for demographics and standardized patient detection.

Results: Compared with other physicians, those who had MPCC scores in the lowest tercile had greater standardized diagnostic testing expenditures (11.0% higher, 95% confidence interval [CI], 4.5%-17.8%) and greater total standardized expenditures (3.5% higher, 95% CI, 1.0%-6.1%). Whereas lower MPCC scores were associated with shorter visits, adjustment for visit length and standardized patient detection did not affect the relationship with expenditures. Total (testing, ambulatory and hospital care) expenditures were also greater for physicians who had lower MPCC scores, an effect primarily associated with the effect on testing expenditures.

Conclusions: Patient-centered communication is associated with fewer diagnostic testing expenditures but also with increased visit length. Because costs and visit length may affect physicians’ and health systems’ willingness to endorse and practice a patient-centered approach, these results should be confirmed in future randomized trials.


Introduction

Patient-centered communication is based on a moral philosophy that calls for physicians to expand upon the biomedical approach to care by (1) helping patients feel understood through inquiry into patients’ needs, perspectives, and expectations; (2) attending to the psychosocial context; and (3) expanding patients’ involvement in understanding their illnesses and in decisions that affect their health.1-3 Patient-centered communication is a complex construct, aspects of which have differential associations with such outcomes as patient satisfaction4 and control of chronic disease.5-8 Most physicians tend to use a biomedical rather than a patient-centered communication style,9 whereas most patients prefer a patient-centered approach.10,11

Although patient-centered communication should not be advocated on the basis of cost considerations alone, it is important to understand the cost implications of such an approach from a health policy perspec-
tive. Concerns that patient-centered communication might drive up health care costs has led to at least 2 inquiries into the relationship between communication and diagnostic testing. Even though both studies reported that elements of patient-centered communication are associated with fewer diagnostic tests, these studies have been limited by several factors. Stewart et al, in their study of 39 family physicians and 315 of their patients, found that 1 of 3 components of a newly developed patient survey of “perceived patient-centeredness” was associated with reduced diagnostic testing; 1 component of their validated observational measure was related to the survey measure but not to other outcomes. Reliance on the same patients to yield effects on both diagnostic testing and measures of patient-centered communication, however, may have introduced biases; for example, severity of illness is associated both with lower patient ratings of their physicians and with increased testing. The level of analysis was at the level of the individual encounter, precluding an overall measure of a particular physician’s style of communication and use of diagnostic tests independent of each patient’s influence on patient-centered communication.

Bertakis et al randomized 509 new patients to primary care resident physicians and measured these physicians’ interaction style and the costs generated by the patients for a 1-year period. Although Bertakis et al did not study patient-centered communication as such, a “technically oriented” visit style was associated with more diagnostic tests than was a style characterized by patient activation.

Patient-centered theory suggests that patients who think they understand their symptoms and feel understood by their physicians may be less anxious, have greater confidence in their physician’s abilities, and be more trusting of their physician. A previous report from the data presented here suggests that physicians who exhibit more observable patient-centered communication behaviors also generate higher levels of trust, not only in the observed standardized patients but overall. Patients who trust their physicians’ judgment may be less likely to demand diagnostic tests in a quest for certainty or sense of control. Similarly, mutual understanding and trust may decrease the drive to order diagnostic tests in an effort to allay the physician’s own anxiety and fear of litigation. Given that diagnostic test costs are under physician control more than are inpatient costs (which are driven by diagnostic related groups), the former would likely be more sensitive to physician communication style.

Ideally, reduced testing should occur in situations in which diagnostic testing is of unproven benefit and in which reduced testing does not reduce patient satisfaction. Among patients with back pain, for example, Deyo found that patients who reported their physicians gave “adequate explanation” of their symptoms also demanded fewer diagnostic tests and were as satisfied as those who received tests. Thus, in the original study design, we sought to compare physicians’ behavior when evaluating conditions that were medically straightforward (such as gastroesophageal reflux disease [GERD]) with their behavior when evaluating medically unexplained symptoms (atypical chest pain) likely to increase anxiety and drive a need for certainty. We report here the aggregate effect of patient-centered communication on costs; comparisons between the 2 conditions and visit-specific outcomes will be reported elsewhere.

A key challenge to studying the relationship between patient-centered communication and diagnostic test use is adequately controlling for patient factors, including type of visit, illness severity, number of concerns, stated preferences, and choice of physician based on practice style. Because it is difficult to adjust adequately for these differences among patients, observed effects may simply reflect confounding by patient factors. Furthermore, no study has examined whether patient-centeredness represents, in part, a physician’s style, observable for more than 1 patient interaction, or whether that style exhibits any relation to the use of diagnostic testing and visit length.

To address these limitations, we measured the physicians’ communication style using unannounced covert standardized patients who carried concealed audio recorders. Standardized patients have been used to generate reliable estimates of health care processes; these estimates have been found to correspond to actual physician behavior more accurately than chart audit or response to vignettes. A focus group with community physicians in preparation for this study suggested that we could achieve the required number of physicians for adequate statistical power if we used no more than 2 standardized patient visits per physician. To assess costs, we used standardized expenditures generated by those same physicians with their real patients for a 1-year period derived from a managed care claims database.

**METHODS**

The study used patient surveys and audio recordings of clinical encounters with standardized patients to arrive at descriptions of physicians’ overall clinical style. Claims data from a large managed care organization (MCO) were used to assess diagnostic testing costs. These expenditures were considered more representative of the clinician’s overall diagnostic test use patterns than were single standardized patient visits.
Physician Sample
In late 1999, we identified 594 primary care physicians in active clinical practice within 45 minutes of Rochester, NY, belonging to a large MCO serving the 8-county Rochester, NY, region (population 1.1 million). To achieve stable measures of costs, only the 506 physicians who had more than 100 patients in the MCO were eligible; thus, enrolled physicians, compared with physicians not enrolled, had larger practices. We also intentionally oversampled family physicians to allow for comparisons between family physicians and internists (reported elsewhere). A maximum of 2 physicians per practice were recruited to avoid clustering effects and to minimize physician detection of standardized patients. The remaining 297 eligible physicians were recruited by 12 physician-recruiters in random order until a total of 100 physicians were recruited. Physicians gave informed consent to participate in a study of "patient care and outcomes." They agreed to have 2 unannounced, covert, surreptitiously audio recorded standardized patient visits at any point in the subsequent 12 months during 2000-2001. Physicians were reimbursed $100 for each standardized patient visit (slightly less than usual charges for a new patient appointment for acute care) and received $100 for completing the survey instruments; $100 was provided to the office staff at each site for their help with scheduling standardized patient visits, facilitating collection of patient survey instruments, and sending standardized patient charts to study investigators. The study received institutional review board approval.

Standardized Patient Visits
To address one of the study aims (reported elsewhere)—to compare physician behavior in response to straightforward patient symptoms—we created 2 contrasting standardized patient roles, both portraying patients with chest pain. The GERD role portrayed a 48-year-old patient (male or female) with typical symptoms of GERD. The medically unexplained symptoms role portrayed a moderately distressed patient (male or female) with multiple symptoms, including atypical chest pain. Each physician saw 2 standardized patients, 1 male and 1 female, and 1 of each illness condition randomized by order, illness condition, and patient sex. Each standardized patient visit was recorded using a digital, audio disk recorder with a high-quality microphone; all equipment was hidden inside a handbag or backpack.

Two days after the visit, a fax was sent to the physician to determine whether, when prompted, the physician was able to identify the standardized patient. The physician indicated at what point in the visit detection occurred, how realistic the patient portrayal was, and whether the physician altered any practice behavior because a standardized patient was suspected.

Visit length was calculated using the audio recording, excluding waiting time and any period of more than 1 minute during which the physician left the room. Audio recordings were analyzed using the Measure of Patient-Centered Communication (MPCC), a validated instrument based on a model of patient-centered communication, which includes the 3 major communication elements of patient-centered communication, measures physician responsiveness to patient concerns, and has been positively correlated with patient trust and patient perceptions of patient-centeredness. A full discussion of the measure can be found in the online-only Supplemental Appendix, available at http://www.annfammed.org/cgi/content/full/3/5/415/DC1; theoretical considerations in measuring patient-centered communication are discussed in a recent publication.

Standardized Utilization Data
To assess standardized health care services utilization, we derived standardized expenditures in discrete categories, diagnostic testing, hospital care, and total expenditures, from the MCO 1996-9 claims data. Details about the claims data have been published elsewhere.

Analyses
Data were analyzed at the patient level using Stata (Version 8.2, StataCorp, College Station, Tex). We used ordinary least squares regression to examine factors affecting the standardized expenditures in the categories described above; we used the logarithm of expenditures to adjust for the skewing of expenditure data. All analyses were adjusted for patient age, sex, Zip code-based socioeconomic status, year of enrollment, total years of enrollment, case-mix (a dummy variable for each ambulatory diagnostic group), physician specialty, and the nesting of patient observations within primary care physician. The key independent variable of interest, the physician’s MPCC score, was expressed in terciles. Further details can be found in the Supplemental Appendix.

RESULTS
Of the 297 physicians with whom contact was attempted, 14 were later found to be ineligible (eg, planning to retire within 6 months), and 109 declined to participate (a loss of 42% of those originally eligible) in the process of recruiting the 100 study physicians. The most common reason for refusal was lack
of time. As shown in Table 1, the sociodemographic, utilization, and clinical characteristics of patients in practices of enrolled and not enrolled physicians were similar. Of the 100 physicians, 93 completed both standardized patient visits. The rest completed only 1 standardized patient visit, and later moved out of the area or withdrew from the study for personal reasons. The mean total MPCC score for both cases was .50 with similar standard deviations and ranges (GERD case: standard deviation [SD] .09, range .25-.74; medically unexplained symptoms case: SD .08, range .25-.63), there was no statistically significant difference between the scores. The correlation between the 2 MPCC scores was 0.39 (P = .0001); the reliability, calculated using the Spearman-Brown prophecy formula, for the average of the 2 cases was .56. Subsequent analyses use the mean MPCC scores.

As displayed in Table 2, physicians with MPCC scores in the lowest tercile generated greater expenditures compared with other physicians. Compared with physicians in the combined middle and highest terciles, those with MPCC scores in the lowest tercile generated greater standardized diagnostic testing expenditures (11.0% greater, 95% confidence interval [CI], 4.5%-17.8%) and total standardized expenditures (3.5% greater, 95% CI, 1.0%-6.1%). There was no significant (P > .4) relationship between MPCC score and total expenditures with diagnostic expenditures subtracted out, that is, the effect of patient-centered communication on total standardized expenditures appears to reflect its effect on standardized diagnostic testing expenditures. There was no significant relationship between MPCC scores and standardized hospital expenditures.

Mean MPCC score was correlated with mean visit length (r = 0.36, P = .0002) for both roles. Mean visit lengths by MPCC tercile were 18.8 minutes (SD 5.2 minutes), 19.6 minutes (SD 3.3 minutes), and 22.8 minutes (SD 7.1 minutes), respectively, for the lowest to highest terciles. There were no relationships between visit length and costs when adjusted for MPCC scores; however, when adjusted for visit length, the relationship between MPCC scores and costs remained significant. These findings were adjusted for physician and patient demographics and illness burden as described in the Methods section.

In 80 (40%) of the 198 visits, physicians were able to identify the standardized patient when prompted 2 days later. The most common reasons for detection were a closed physician practice (n = 19, 63%), physician notification by staff (n = 10, 33%), and poor acting by the standardized patient (n = 1, 3%). For detected visits, mean physician rating of realism was 8.1 on a scale from 1 to 10. Analyses were repeated excluding detected visits with realism scores below 7, and repeated again adjusting for visit length and prompted suspicion that the patient was an standardized patient; these factors did not significantly affect the relationships between patient-centered communication and costs.

### Table 1. Characteristics of Patients and Physicians Enrolled and Not Enrolled in the Study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Not Enrolled</th>
<th>Enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>483,094</td>
<td>121,806</td>
</tr>
<tr>
<td>Age, years (SD)</td>
<td>41.1 (11.2)</td>
<td>41.0 (11.0)</td>
</tr>
<tr>
<td>Female sex, %</td>
<td>52.7</td>
<td>53.9</td>
</tr>
<tr>
<td>Median income, $ (SD)*</td>
<td>36,874 (10,160)</td>
<td>37,830 (10,683)</td>
</tr>
<tr>
<td>High-school graduation, % (SD)*</td>
<td>63.8 (7.9)</td>
<td>64.8 (7.8)</td>
</tr>
<tr>
<td>Any visit to a physician, %</td>
<td>82.5</td>
<td>83.1</td>
</tr>
<tr>
<td>Referred, %</td>
<td>25.6</td>
<td>25.7</td>
</tr>
<tr>
<td>Years enrolled in the MCO, No. (SD)</td>
<td>3.07 (1.12)</td>
<td>3.07 (1.12)</td>
</tr>
<tr>
<td>ADGs, mean No. (SD)</td>
<td>2.99 (2.67)</td>
<td>3.02 (2.67)</td>
</tr>
<tr>
<td>Number</td>
<td>594</td>
<td>100†</td>
</tr>
<tr>
<td>Specialty, family practice, %</td>
<td>24</td>
<td>47</td>
</tr>
<tr>
<td>Patients enrolled in the MCO, No. (SD)</td>
<td>813 (776)</td>
<td>1218 (758)</td>
</tr>
</tbody>
</table>

### Table 2. The Relationship of Patient-Centered Communication to Adjusted Health Care Costs

<table>
<thead>
<tr>
<th>Standardized Expenditure Category</th>
<th>Tercile of MPCC Score</th>
<th>Lowest Tercile</th>
<th>Middle Tercile (95% CI)</th>
<th>Highest Tercile (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic testing costs</td>
<td>100</td>
<td>90.3 (84.3-96.9)</td>
<td>89.6 (83.9-95.6)</td>
<td></td>
</tr>
<tr>
<td>Inpatient costs</td>
<td>100</td>
<td>103.7 (92.2-116.7)</td>
<td>98.9 (87.5-111.8)</td>
<td></td>
</tr>
<tr>
<td>Total costs</td>
<td>100</td>
<td>96.3 (93.6-99.0)</td>
<td>96.8 (94.1-99.6)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Included are patients with at least some expenditures in each category. Adjusted percentage of standardized expenditures presented as terciles of MPCC scores. Analyses adjust (from claims data) for patient age, sex, Zip code-based socioeconomic status, ambulatory diagnostic groups, year, years of enrollment, and physician specialty.

MPCC = Measure of Patient-Centered Communication.

† As percentage of lowest tercile.
DISCUSSION
This study suggests that physicians with a more patient-centered communication style tend to have fewer diagnostic testing expenditures. Total standardized expenditures were also less for these physicians, largely reflecting the relationship between patient-centered communication and diagnostic testing.

This study adds to previous research on patient-centered communication and health care costs by bringing several specific strengths. First, the measures were not all derived solely from the patient visit; we used independent sources of data to avoid contamination of one measure by another. Second, we used actual expenditures, not just a count of tests generated only by the patients studied. Third, the relationship was based on observational measures of patient-centered communication rather than patient survey reports which are subject to unmeasured patient confounding. Fourth, despite their drawbacks, standardized patients greatly reduce the variation typically observed using real patients, avoiding some of the confounding encountered in naturalistic observation studies.

Finally, we examined standardized inpatient expenditures as a control condition. While primary care physicians have a substantial degree of direct control over outpatient diagnostic testing utilization, they have relatively little impact on the standardized expenditures of hospitalizations, which are based on diagnostic related groups and are affected mostly by patient case-mix. The finding that inpatient costs were not affected by patient-centered communication suggests that inadequate adjustment for patient characteristics (particularly case-mix) does not explain relationship between MPCC scores and expenditures.

The mechanisms for a relationship between patient-centered communication and diagnostic testing have not been explored and should be incorporated into future randomized trials of interventions to improve patient-centered communication. We have shown a relationship between trust and patient-centered communication. We have shown a relationship between trust and patient-centered communication. Patient-physician relationships characterized by a higher level of trust may result in less perceived or real pressure on physicians to order tests. Other possible mediators of the drive to test include patient factors (symptoms, anxiety, trust, requests, autonomy support, self-efficacy) and physician factors (tolerance of uncertainty, personality). Further studies should examine whether patient-centered communication is associated with a reduction in inappropriate testing, appropriate testing, or both.

We noted that visit length tended to increase with increased MPCC scores, but that MPCC, not visit length, was associated with costs. Longer visits are likely associated with improved quality of care in several domains, including prevention, prescribing, response to emotional distress, and information exchange. Further research is also needed to determine the causal relationships among patient-centered communication, visit length, and outcomes—for example, whether longer visits are a necessary but not sufficient condition for patient-centered communication, or whether shortering visits interferes with patient-centered communication and results in greater overall costs. Given current incentives to see more patients in less time, physicians using a more patient-centered consultation style may be at a financial disadvantage if their longer visits reduce the number of patients they can see, or if they take additional time without receiving additional compensation. These physicians may be at greater risk for stress, fatigue, and burnout, which in turn can negatively affect quality of care. Educational programs should develop means for training in patient-centered communication skills while emphasizing time management.

Study Limitations
The difficulties of studying the relationship between patient-centered communication and costs should not be underestimated; our results are subject to multiple interpretations for several reasons. Stable measures of physician style, extrapolating from our calculations, would require 6 standardized patient visits; however, physicians would not have participated had that been our protocol. Using real patients to measure patient-centered communication invites unmeasured confounding. The calculation of health care costs is an inexact science.

The observational and cross-sectional nature of the data precludes making causal interpretations. The mechanisms by which observed differences in physician communication might reduce diagnostic testing or increase visit duration are uncertain. Also, it is unclear whether the reduction in testing affected the quality of care.

Patient-centered communication is an aggregation of loosely related skills, such as eliciting the patient’s perspective on the problem, inquiring into the psychosocial context, and encouraging patient participation in decisions. Although MPCC components correspond to some of these skills, the case-to-case consistency of physician behavior and the interrater reliability of the components are insufficient to indicate which skills or behaviors are most directly related to health care costs. Although the MPCC shows good interrater reliability for individual standardized patient visits, the ability to detect a physician’s overall style is limited when measured with only 2 visits. Thus, some physicians may be misclassified. This bias is likely to result in underestimation of the true relationship between patient-centered communication and diagnostic testing expenditures. It is possible that unmeasured confounding explains both...
higher levels of observed patient-centered communication and fewer diagnostic testing expenditures. For example, some physicians may systematically attract patients who induce more patient-centered behaviors in their physicians while also demanding fewer tests.

Using standardized patients results in advantages and difficulties. The choice of the standardized patient roles may have selected for certain specific types of physician behavior. Additional visits would likely have increased reliability, but would have made physician recruitment and retention difficult. Even so, we exceeded the standards of nearly all other published studies using unannounced standardized patients, which have used just 1 standardized patient visit per physician. Furthermore, physicians’ behavior with a new patient might not predict their subsequent behavior as the relationship develops with time.38 To have used audio-recorded real patient visits would introduce Hawthorne effects, as physicians may have become more accommodating to patient requests.

Detection of standardized patients did not affect the results we report. There appears to be an inverse relationship between timing of the inquiry and the reported detection rate. We chose to inform physicians 2 days after the visit as a courtesy so they would not need to pursue the patient further. The proximity of the prompt to the standardized patient visit, however, allowed physicians to choose from a small pool of recent patients whom they might recall. Data from the same physician pool from a more recent study in which detection notification was requested 10 working days after the visit might not predict their subsequent behavior as the relationship develops with time.38 In our study, detection rates as low as 2% simply asked the physicians to contact study personnel if they suspected they had seen an standardized patient and were never prompted.40

Unmeasured pharmacy expenditures may have compensated for the savings realized in diagnostic testing. That is, physicians scoring higher on the MPCC may have tended to prescribe empirically rather than base prescribing on diagnostic test results. Pharmacy data in the MCO database were incomplete and inadequate for analysis.

The results apply only to those physicians selected into the study sample. While the patients of enrolled and not enrolled physicians appear to be similar, participating physicians are likely to exhibit some important differences from others in the community. Because other recruitment methods would not have provided an adequate sample size, physicians were recruited by peers; they also agreed to participate in a relatively intrusive study. Generalization to patients with different insurance plans (or no insurance), and beyond primary care physicians in the Rochester, NY, area remains unproved.

Patient-centered communication represents an attempt to balance two imperatives in the clinical encounter: the need to arrive at a diagnosis and treatment plan, and the need to understand patients and involve them in care. Our findings should reduce fears that encouraging patient-centered communication would necessarily drive up health care costs. Although we found that patient-centered communication is associated with fewer diagnostic testing expenditures, a patient-centered approach should not be implemented solely for economic reasons. Patients should also perceive their care as better, and health outcomes should be improved. Although the evidence for the former is strong, more research is needed to assess the effect of patient-centered communication training on health outcomes. Randomized trials should study the effects of patient-centered communication on health care costs to confirm our observational finding. The possibility that primary care physicians are penalized for their patient-centered communication by increasing visit length without additional compensation should also be investigated further, as it might discourage physicians from practicing in a patient-centered manner.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/first/3/5/415.

Key words: communication; health care economics; health care delivery; physician-patient relations; standardized patients; diagnostic tests, routine; health expenditures; evaluation studies

Submitted January 3, 2005; submitted, revised, May 6, 2005; accepted May 9, 2005.

Funding support: Patient Centered Care and Health Care Costs, AHRQ R01-HS09963-01A1, Peter Franks, Principal Investigator.

Acknowledgments: The authors thank Patricia Carney, RN, PhD, for her help developing standardized patient training protocols; Moira Stewart, PhD, and Judith Belle Brown, PhD, for their expertise in coding of audio recordings; and Richard Kravitz, MD, MSPH, for his careful review of the manuscript. William Bayer, MD, Marc Berliant, MD, Phillip Bonanni, MD, Paul Frame, MD, Diane Morse, MD, Timothy Quill, MD and Geoffrey Williams, MD, PhD, helped recruit physicians for the study. Judi Lardner assisted with manuscript preparation and references.

References


Antibiotic Treatment and Survival of Nursing Home Patients With Lower Respiratory Tract Infection: A Cross-National Analysis

ABSTRACT

PURPOSE Although lower respiratory tract infections are a leading cause of death in frail elderly patients, few studies have compared treatments and outcomes. We assessed the effects of different antibiotic treatment strategies on survival of elderly nursing home residents with lower respiratory tract infections in the United States and the Netherlands, where treatment approaches are quite different.

METHODS We combined data from 2 prospective cohort studies of lower respiratory tract infections conducted in 36 nursing homes in the United States and 61 in the Netherlands. We included residents whose infections were treated with antibiotics: 806 in the United States and 415 in the Netherlands. Outcome measures were 1-month and 3-month mortality. We used logistic regression to adjust for differing illness severity.

RESULTS Dutch residents had higher mortality than US residents (28.1% vs 15.1% at 1 month, respectively; P < .001). After adjusting for illness severity with logistic regression, the differences between the Dutch and US populations were not significant (odds ratio 1.34; 95% confidence interval, 0.94-1.90). Predicted mortality was overestimated for more severely ill US residents at 1 month but not at 3 months. No antibiotic regimen was consistently associated with increased or decreased mortality.

CONCLUSION Despite differences in illness severity and treatment, adjusted mortality did not differ between the 2 countries. Although we cannot exclude a short-term survival benefit from more aggressive treatment in the United States, differences in baseline health appear prognostically more important than the type of antibiotic treatment.


INTRODUCTION

Nursing home residents frequently develop lower respiratory tract infections, mostly pneumonia and bronchitis,1-5 with estimated 30-day mortality from pneumonia or lower respiratory tract infection varying between 10% and 30%.6-9 Because meaningful cultures from specimens are rarely obtained in practice,10,11 antibiotic treatment is usually empiric.10-12 Despite guidelines from authoritative groups,13-15 however, little evidence supports any specific regimen. Naughton and colleagues16 found no difference in mortality between residents treated initially with either oral or intramuscular antibiotics. In a second study, adherence to an antibiotic guideline did not affect mortality.17 In a few small trials from the late 1980s or early 1990s, researchers found effectiveness of ceftriaxone or ciprofloxacin to be equal when compared with similar antibiotics.18-21

The ideal study design to test different regimens would be a randomized controlled trial; practical and ethical difficulties make such a study unlikely. Consequently, we must rely on data from observational studies, attempting to control for differences between treatment groups. Two large,
concurrent studies of lower respiratory tract infections acquired by residents in nursing homes in the United States and the Netherlands provided the opportunity to examine how antibiotic treatment affected outcomes.20,21 In the US study, 30-day mortality was associated with 8 variables in a multivariable logistic model: blood urea nitrogen, white cell count, low absolute lymphocyte count (<0.8 × 10^9/L), lower body mass index, pulse rate, activities of daily living status, male sex, and deterioration in mood.25 One-week mortality in Dutch residents who were treated with antibiotics was related to 7 variables in multivariable logistic regression: respiratory rate, pulse rate, dependency in eating, recent increase in discomfort, decreased alertness, inadequate fluid intake, and recent occurrence of decubitus ulcers.23

Compared with treatment in the United States, hospitalization, use of intravenous antibiotics and multiple antibiotic regimens, intravenous fluids, and feeding tubes are relatively rare in the Netherlands.3,7,8,22-26 Such wide differences in care provide a natural experiment of the relationship of treatment and outcomes. We hypothesized that although there would be little overall difference in 1- and 3-month mortality among regimens, we would identify a group of residents for whom more aggressive therapy is more effective.

METHODS
Patient Recruitment and Data Collection
The Missouri LRI (lower respiratory tract infection) Study prospectively enrolled residents of 36 nursing homes in central Missouri and the St. Louis area between August 1995 and September 1998. Trained project nurses evaluated residents who had symptoms compatible with a lower respiratory tract infection. Diagnosis of lower respiratory tract infection was determined according to a standardized definition that required either (1) 3 of 6 signs or symptoms of lower respiratory tract infection (eg, cough, purulent sputum production) or general illness (eg, fever or worsening mental or functional status); or (2) 2 signs and symptoms with a chest radiograph interpreted as showing probable pneumonia. To avoid including an exacerbation of congestive heart failure or chronic obstructive pulmonary disease, we also required either probable pneumonia or a temperature of at least 38°C in association with those chronic conditions. Details of identification and evaluation are described elsewhere.22,27 About two thirds of residents had possible or probable pneumonia determined by chest radiographs. Those found to have a lower respiratory tract infection (1,409 episodes in 1,044 residents) were observed for a minimum of 3 months after evaluation. Data were collected from several sources, including nurse evaluations and medical record information. Institutional review boards at the University of Missouri-Columbia School of Medicine and Washington University School of Medicine approved the protocol.

The nationwide Dutch Pneumonia Study was conducted in psychogeriatric units in 61 nursing homes between October 1996 and July 1998.23 Enrollment criteria included psychogeriatric disease (97% had dementia), facility residence for at least 4 weeks, and pneumonia diagnosed by the attending physician. The investigators prospectively identified 706 consecutive residents with a diagnosis of pneumonia. Consistent with Dutch practice, attending physicians belonged to the permanent staff of the nursing home and cared for patients daily.24 They rarely ordered chest radiographs or blood tests. In contrast with the US study, Dutch residents were enrolled only once. The physicians caring for patients recorded their data on standardized forms at several points during 3 months of follow-up. The Medical Ethics Committee of the VU University Medical Center, Amsterdam, approved the study protocol.

Selection of Residents
For this analysis, we included US and Dutch residents aged 70 years or older whose condition explicitly met the US study definition of a lower respiratory tract infection and who received antibiotic therapy. Antibiotic therapy was defined as any antibiotic started from 2 days before diagnosis until 5 days after diagnosis. We excluded residents who did not receive antibiotics because they were dissimilar in the 2 countries. Dutch residents who did not receive antibiotics were quite ill and almost always receiving palliative care; in contrast, many US residents who did not receive antibiotics were not very ill.29,30 Among residents with dementia who did not receive antibiotics, 1-month mortality was strikingly different for Dutch and US residents, 90.2% and 17.5%, respectively.30

Only the first episode of illness was included for US residents with multiple enrollments. We excluded 193 of 1,044 US residents who did not receive antibiotics within the specified time frame and 45 who were younger than 70 years, leaving 806 for this analysis. Of the original 706 Dutch residents, we excluded 127 who did not meet the US study definition, 5 whose diagnosis was revised (to non–lower respiratory tract infection) after enrollment, 141 who did not receive antibiotics or whose antibiotic status was unknown, 16 who were younger than 70 years, and 2 with unknown age. We included the remaining 415 Dutch residents in the analysis.

Variables
We measured variables relative to the time of the treatment decision (in the Dutch study) or the time of the resident’s evaluation (in the US study). Hereafter, we
refer to this as “the time of diagnosis” for both studies. Mortality was noted for at least 3 months after diagnosis. Selection of variables and definitions have been described in detail elsewhere. We defined dehydration as either poor skin turgor or dry mucous membranes noted at the time of diagnosis. Activities of daily living (ADL) performance was measured using the 5-level self-performance items from the Minimum Data Set (MDS) in the US study and the 4-level Bedford Alzheimer Nursing Severity-Scale (BANS-S) items in the Dutch study. To create common ADL measures for self-performance of eating, walking, and dressing, we collapsed the middle categories to create 3-level items scored as 0 (independent), 1 (requires assistance), or 2 (fully dependent). The sum of these 3 scores was used to rate overall ADL performance. Our definition of severe dementia included both cognitive and ADL impairment. We defined study residents as having severe dementia if they had at least 4 of 6 possible points on the overall ADL scale and were severely cognitively impaired (a score greater than 15 on the BANS-S for Dutch residents or a score of 5 or 6 on the CPS for US residents).

Analysis
We performed all statistical analyses with SAS for Windows. To determine a relationship between antibiotic choice and mortality, we first created a statistical model to control for illness severity. After combining the data sets, we began with bivariate analyses to determine which variables were related to mortality and nationality. For the rare missing values (mostly less than 4%), we imputed the mean for continuous variables and the mode for categorical data. To avoid disproportionate influence of extreme outliers, we set pulse rates below 60 beats per minute at 60 beats per minute and pulse rates above 140 beats per minute at 140 beats per minute. Similarly, we truncated respiratory rates above 50 breaths per minute at 50 breaths per minute.

Beginning with residents’ characteristics, we developed logistic regression models for 1- and 3-month mortality. Because the purpose of the models was to control for confounding, we did not limit the number of variables in the models and retained variables of marginal statistical significance (.05 < P < .10) that we believed were clinically important. We also included variables that were not statistically significant if they were important confounders and tested several quadratic terms. After developing these initial models, we tested facility size and the physician’s age and years of experience in the models. Finally, we included a dichotomous nationality variable (Dutch = 1, US = 0). We explored whether variables performed consistently across the 2 studies by interacting variables by nationality and testing the interaction terms in the final models. Stratifying by deciles of predicted mortality, we examined actual mortality for US and Dutch residents to determine whether model performance was consistent in both populations.

We then examined initial antibiotic therapy relative to mortality. Antibiotics were divided into several classes based on the clinical relevance and prevalence of individual agents. For Dutch residents, we further consolidated the classes because some antibiotics were rarely prescribed in the Netherlands.

We found that antibiotic use differed between the 2 countries, not only in the agents used but in how their use related to illness severity (Figure 1). The data are therefore incompatible with the simple approach of adding a design variable representing classes of antibiotics to the basic logistic models that control for resident and facility characteristics. Instead, to test whether mortality differed by antibiotic choice, we stratified residents by ranges of illness severity calculated from our logistic models. Because mortality risk was unevenly distributed in the 2 populations, we used clinically meaningful cutoff values that would place a reasonable number of residents in each group. Because of the disparity in antibiotic treatment, we examined Dutch and US residents separately. After controlling for mortality risk, we used the Cochran-Mantel-Haenszel statistic of general association to test for a consistent relationship between mortality and antibiotic treatment over all strata. Fisher’s exact test was used to determine whether antibiotic treatment had different associations with mortality within different risk strata.

RESULTS
Description of Residents and Treatments
Compared with US residents, Dutch residents were more severely ill overall, more severely ADL dependent, and more often had severe dementia; however, they had fewer comorbidities (Table 1). Unadjusted mortality was substantially higher for Dutch residents than US residents at both 1 month (28.1% vs 15.1%, respectively, P < .001) and 3 months (36.5% vs 24.8%, respectively, P < .001).

Most Dutch residents (71.8%) were treated with oral amoxicillin or amoxicillin/clavulanate; only 15.3% of US residents received these regimens (Table 1). US residents most frequently received cephalosporins (37.6%). Thirty-nine different antibiotic agents were used in the United States compared with 15 in the Netherlands.

Forty-six (11.1%) Dutch residents received intramuscular ampicillin or amoxicillin, followed by oral amoxicillin, this combination was not used in the United States. Other than that combination, only 1 Dutch resident received 2 antibiotics (cefuroxime and gentamicin) com-
pared with 14.5% of US residents (for example, 9.2% received a third-generation or parenteral second-generation cephalosporin with another agent). US residents received parenteral antibiotics more often than Dutch residents (32.1% vs 12.8%, respectively, \(P < .001\)).

Patterns of antibiotic treatment differed between nations across mortality risk derived from our logistic model described below (Figure 1). As mortality risk increased, treatment of US residents with doxycycline, tetracycline, macrolides, or trimethoprim-sulfamethoxazole declined while treatment with third-generation or parenteral second-generation cephalosporins increased. In contrast to the United States, amoxicillin was commonly used across all levels of risk in the Dutch population. The combination of intramuscular ampicillin or amoxicillin followed by oral amoxicillin was given to Dutch residents with higher mortality risk.

Only 1 Dutch resident (0.2%) but almost one third (30.4%) of US residents were hospitalized. For US residents, hospitalization increased with mortality risk, from 15.6% of residents with less than 5% mortality risk, to 56.4% of residents with a mortality risk of 40% or more. More than two thirds (68.2%) of hospitalized residents received parenteral antibiotics compared with 16.4% of residents who were not hospitalized. Rehydration therapy was 10 times more common in the United States (34.5%) than in the Netherlands (3.0%) in the month following diagnosis \((P < .001)\).

**Mortality Models**

Table 2 shows logistic models for 1- and 3-month mortality. For 1-month mortality, several indicators of acute illness were associated with increased risk of mortality, as were impaired self-performance of eating and walking, male sex, increased age, and Parkinson’s disease. Age was of borderline significance \((P = .064)\), but we retained it in the model to help control for confounding. Model discrimination was good \((c\) statistic = 0.765), with good calibration over the range of mortality risk (Hosmer-Lemeshow goodness-of-fit, \(P = .895\)).

The 3-month model also included 2 variables reflecting an increasing number of chronic health problems (bladder continence and congestive heart failure) and respiratory rate, another acute illness indicator, which was of borderline significance \((P = .096)\); respiratory rate was not significantly associated with 1-month mortality \((P = .302)\). Discrimination \((c\) statistic = 0.737) and calibration (Hosmer-Lemeshow goodness-of-fit, \(P = .635\) ) were good.

Despite the increased illness severity of the Dutch residents compared with US residents, Dutch nationality was not independently significant when added to either the 1- or 3-month models (odds ratio \([OR]\) 1.34, 95% confidence interval \([CI]\) 0.94-1.90; and OR 1.20, 95% CI, 0.87-1.66, respectively). There were no significant interactions between nationality and other variables. For low to moderately high risk of 1-month
mortality (0% to 25%), predicted and actual mortality were quite similar between the 2 populations (Figure 2). At higher levels of predicted mortality, however, the model overpredicted mortality for US residents and underpredicted mortality for Dutch residents. This disparity was not evident at 3 months.

Treatment Effects
After stratifying by mortality risk, the overall association between antibiotic agent and mortality for Dutch residents was not statistically significant at 1 month ($P = .143$), but it was statistically significant at 3 months ($P = .020$), indicating that mortality is not distributed evenly between antibiotics after adjusting for mortality risk. Looking within individual risk strata (available in Supplemental Table 1 online at: http://www.annfammed.org/cgi/content/full/3/5/422/DC1), however, there was no statistically significant association between a particular antibiotic and mortality at any level of risk for Dutch residents at either 1 or 3 months. The overall association between antibiotic agent and mortality for US residents was not statistically significant at either 1 or 3 months ($P = .208$ and .174, respectively) after adjusting for stratum of mortality risk.

DISCUSSION
We examined the different approaches to care of nursing home residents with lower respiratory tract infection in the United States and the Netherlands and found little evidence that specific choice of antibiotic therapy affects mortality. Despite our hypothe-
esis, we also could not identify specific risk strata in which more aggressive antibiotic therapy appeared to result in better mortality outcomes. Based on their characteristics, the Dutch residents appeared to be much sicker and had higher unadjusted mortality. Nonetheless, the variable representing Dutch nationality was not significant in either mortality model. This finding indicates that mortality differences were mostly explained by case-mix and allowed us to substantially control for the differences between these 2 populations. We found no clear advantage for treatment with any particular antibiotic regimen after stratification by illness severity. At high-mortality risk, however, 1-month mortality for US residents was overestimated by the logistic model, and conversely, underestimated for Dutch residents (Figure 2). This difference was not evident at 3 months. These findings are consistent with some short-term increased survival associated with the more aggressive therapy observed in the United States that included more frequent use of parenteral antibiotics, multiple antibiotic regimens, and hospitalization. Even so, we cannot exclude some other difference between the populations as the explanation.

Whether short-term increased survival is of potential benefit is open to question. For residents with advanced dementia, who are facing a high 6-month mortality and the frequent necessity of restraints to prevent removal of intravenous lines, the benefit of such aggressive care has been increasingly questioned.3,26,35-39 Our findings lend further support to this position because aggressive care appeared to offer little survival benefit in the 3 months after lower respiratory tract illness. Previous analysis of Missouri LRI Study data found no difference in 1-month mortality between residents first treated in the nursing home or the hospital, after adjusting for illness severity and probability of hospitalization.40

The study’s main limitation is the universal problem with observational studies—we were unable to assure adequate control for confounding. Across levels of illness severity and between the 2 countries, residents received very different therapies, making it difficult to compare outcomes. Furthermore, the great number of regimens made comparisons difficult. Despite the absence of patterns suggesting agents with clear benefit or harm, a much larger study would be needed to provide adequate numbers in all categories of antibiotics and mortality risk.

Different study procedures and differences in the practice of medicine between the two nations could have affected results. Variables could not be used if they were not present in both studies, which forced us to omit some potent risk factors from the original

<table>
<thead>
<tr>
<th>Variable</th>
<th>1-Month Mortality</th>
<th>3-Month Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-7.020</td>
<td>-7.302</td>
</tr>
<tr>
<td>Age*</td>
<td>0.023</td>
<td>0.029</td>
</tr>
<tr>
<td>Bladder incontinence at time of diagnosis†</td>
<td>0.248</td>
<td>0.375</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.330</td>
<td>0.375</td>
</tr>
<tr>
<td>Decreased alertness</td>
<td>0.465</td>
<td>0.375</td>
</tr>
<tr>
<td>Dehydrated</td>
<td>0.488</td>
<td>0.425</td>
</tr>
<tr>
<td>Eating self-performance at time of diagnosis‡</td>
<td>0.442</td>
<td>0.314</td>
</tr>
<tr>
<td>Facility has 300 or more beds</td>
<td>0.515</td>
<td>0.515</td>
</tr>
<tr>
<td>Male</td>
<td>0.576</td>
<td>0.398</td>
</tr>
<tr>
<td>Walking self-performance at diagnosis</td>
<td>0.193</td>
<td>0.369</td>
</tr>
<tr>
<td>New or increased cough</td>
<td>-0.536</td>
<td>-0.497</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>0.596</td>
<td>0.477</td>
</tr>
<tr>
<td>Pulse*</td>
<td>0.021</td>
<td>0.015</td>
</tr>
<tr>
<td>Respiratory distress</td>
<td>0.781</td>
<td>0.469</td>
</tr>
<tr>
<td>Respiratory rate*</td>
<td>0.781</td>
<td>0.469</td>
</tr>
<tr>
<td>C statistic</td>
<td>0.765</td>
<td>0.737</td>
</tr>
<tr>
<td>Hosmer-Lemeshow goodness-of-fit statistic, P§</td>
<td>0.90</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Cl = confidence interval.
* The odds ratios shown for age, pulse, and respiratory rate are for 10-unit change.
† Coding for bladder incontinence is as follows: 1 = continent, 2 = usually continent, 3 = incontinent 2 or more times a week.
‡ Coding for eating dependency is as follows: 0 = independent, 1 = requires assistance, 2 = fully-dependent.
§ A nonsignificant P value indicates good fit over the range of mortality risk.
studies, such as body mass index\textsuperscript{22} and resident discomfort.\textsuperscript{23} In the Missouri LRI Study, project nurses actively attempted to identify potential cases, while in the Dutch Pneumonia Study, nursing home physicians included residents when they diagnosed pneumonia. The physicians may have preferentially included the more definitive cases. To make study subjects comparable, we used a common case definition to select residents, and our logistic models performed well in adjusting for differences in mortality risk.

Ultimately, studies involving direct experimental comparisons between antibiotic agents are needed to determine the best treatments for lower respiratory tract infection acquired by nursing home residents. This article may help to provide a basis on which such a study could be ethically justified, because differences in illness severity are likely much more strongly related to mortality than differences in antibiotic treatments. Nonetheless, in the interim, the lack of benefit in 3-month survival found for more aggressive treatment lends support for treating nursing home residents in the nursing home with simpler regimens that involve less discomfort.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/422.

Key words: Nursing homes; aged; respiratory tract infections; pneumonia; antibacterial agents; delivery of health care; health services research

Submitted December 16, 2004; submitted, revised, March 4 2005; accepted April 4, 2005.

Funding support: This study was supported by a grant from the National Institute on Aging, grant number AG18335. The original data collection in Missouri was supported by the Agency for Healthcare Research and Quality, grant number HS08551, and a Generalist Physician Faculty Scholars Award to Dr. Mehr from the Robert Wood Johnson Foundation. The original data collection in the Netherlands was supported by grants from Dutch Ministry of Health, Welfare and Sport (VWS, grant numbers 37851, 37852, 37853, and 37832) and the Society 'Het Zonnehuis.'

Figure 2. Predicted and actual 1- and 3-month mortality, by decile of predicted mortality risk for US and Dutch residents.
References


Delivery of Clinical Preventive Services in Family Medicine Offices

Benjamin F. Crabtree, PhD1,2
William L. Miller, MD, MA3
Alfred F. Tallia, MD, MPH4
Deborah J. Cohen, PhD1
Barbara DiCicco-Bloom, RN, PhD4
Helen E. McIlvain, PhD4
Virginia A. Aita, RN, PhD3
John G. Scott, MD, PhD3
Patrice B. Gregory, PhD, MPH5
Kurt C. Stange, MD, PhD6
Reuben R. McDaniel, Jr, EdD7

1Department of Family Medicine, UMDNJ-Robert Wood Johnson Medical School, New Brunswick, NJ
2Cancer Institute of New Jersey, New Brunswick, NJ
3Department of Family Medicine, Lehigh Valley Hospital & Health Network, Allentown, Pa
4Department of Family Medicine, University of Nebraska Medical Center, Omaha, Neb
5Department of Preventive and Societal Medicine, University of Nebraska Medical Center, Omaha, Neb
6Departments of Family Medicine, Epidemiology & Biostatistics, and Sociology, Case Western Reserve University, and the Case Comprehensive Cancer Center, Cleveland, Ohio
7McCombs School of Business, University of Texas at Austin, Austin, Tex

ABSTRACT

BACKGROUND This study aimed to elucidate how clinical preventive services are delivered in family practices and how this information might inform improvement efforts.

METHODS We used a comparative case study design to observe clinical preventive service delivery in 18 purposefully selected Midwestern family medicine offices from 1997 to 1999. Medical records, observation of outpatient encounters, and patient exit cards were used to calculate practice-level rates of delivery of clinical preventive services. Field notes from direct observation of clinical encounters and prolonged observation of the practice and transcripts from in-depth interviews of practice staff and physicians were systematically examined to identify approaches to delivering clinical preventive services recommended by the US Preventive Services Task Force.

RESULTS Practices developed individualized approaches for delivering clinical preventive services, with no one approach being successful across practices. Clinicians acknowledged a 3-fold mission of providing acute care, managing chronic problems, and prevention, but only some made prevention a priority. The clinical encounter was a central focus for preventive service delivery in all practices. Preventive services delivery rates often appeared to be influenced by competing demands within the clinical encounter (including between different preventive services), having a physician champion who prioritized prevention, and economic concerns.

CONCLUSIONS Practice quality improvement efforts that assume there is an optimal approach for delivering clinical preventive services fail to account for practices' propensity to optimize care processes to meet local contexts. Interventions to enhance clinical preventive service delivery should be tailored to meet the local needs of practices and their patient populations.


INTRODUCTION

The Healthy People 2010 report calls for a major effort to improve rates of preventive service delivery by primary care clinicians.1 Despite prevention being a core component of primary care practice,2 studies show that clinical preventive service delivery rates are low.34 A range of interventions to enhance the delivery of clinical preventive services have been tried with varying degrees of success, including the use of continuing medical education, audit and feedback, computerized reminder systems, the involvement of nursing staff, the use of chart-based algorithms, and the implementation of continuous quality improvement programs.5-10 Efforts to improve clinical preventive services delivery are limited by our understanding of how clinicians and practices actually incorporate and deliver preventive services within the competing demands of care.11-13 Although recent research provides insight into the emergent nature of practices as human organizations,14,15 much remains to be understood about how the specifics of patients, their families, communities, clinicians, and health systems affect the approaches used in practices to provide clinical preventive services.13,16,17
The goal of this research is to better understand the organizational features of primary care practices that help explain how and why practices deliver clinical preventive services. Specifically, we asked: What competing demands are imposed by carrying out clinical prevention and illness care? What approaches are used by practices with high vs low rates of preventive service delivery? What organizational features support preventive services? Answers may suggest ways for enhancing the delivery of clinical preventive services.

METHODS

The Prevention & Competing Demands in Primary Care study was an observational comparative case study designed to examine the organizational contexts and features that support or inhibit the delivery of clinical preventive services in family medical offices. Practices were selected as cases, and the study design involved extensive observation of clinical encounters and office systems by a field researcher who spent 4 weeks or more in each practice. This comparative case study design sought to optimize qualitative description at the practice level through an iterative, purposeful sampling strategy, and quantitative data were used selectively to enrich the qualitative data. The design is ideal for identifying possible patterns across cases, but the qualitative sampling strategy makes it inappropriate to perform quantitative comparisons across practices.

The study protocol was reviewed and approved by the appropriate institutional review board.

Sampling

Practice selection was based on results from a previous study of 91 practices from a single Midwestern state. These practices were ranked according to tobacco-related services delivery rates and then purposefully chosen in an iterative process to include a range in tobacco-related prevention rates, practice size, and geographic location (urban, suburban, and rural). Eighteen practices participated in the study. Approximately 30 patients per clinician were sampled using a strategy of seeking consent from consecutive patients in the waiting room. Details of practice and patient sampling can be found in Supplemental Appendix 1 available online only at http://www.annfammed.org/cgi/content/full/3/5/430/DC1.

Data Collection

A field researcher trained in qualitative methods collected data through direct observations of the practice and patient encounters, brief patient exit cards to obtain self-reported smoking status and other prevention activities, patient pathways in which a patient gives consent and is observed throughout the visit to the practice, individual in-depth interviews with physicians and key staff members, and chart audits. These data contained detailed descriptions of local clinic environments, patient characteristics, nursing stations, examination rooms, waiting areas, physician offices, and patient education materials. Practice personnel and their roles, duties, and relationships with other staff were characterized in practice genograms. Physical office systems including charts, flow sheets, and computer systems, as well as functional office routines and procedures, were described. Individual in-depth interviews assessed clinician and staff perceptions of prevention philosophy, knowledge, and delivery. Field researchers directly observed and dictated descriptions of approximately 30 patient encounters with each of the more than 50 clinicians from the 18 practices, and charts were abstracted on each patient. Details of these methods, including many of the data collection instruments, have been previously published.

Data Management and Analysis

Interviews and field notes were transcribed and imported into FolioViews 4.2 (Open Market, Inc, Salt Lake City, 1999), a text management software program. Chart audit and structured checklist data were entered into Microsoft Excel and imported into SPSS (Statistical Package for the Social Sciences, SPSS Inc, Chicago, Ill).

Data were analyzed and interpreted in several iterative phases, leading to summaries of each practice’s approach to preventive service delivery. Chart reviews, encounter descriptions, and patient exit card responses were combined to calculate practice rates for 3 types of preventive services: screening (combination of cholesterol screening, Papanicolaou [Pap] test, and mammogram), counseling (smoking cessation counseling), and immunization (combination of childhood series, and adult tetanus, influenza, and pneumonia). Practice summaries, calculated rates for preventive services, and descriptions of individual clinician’s approaches were combined to create a summary characterizing each practice’s strategy for delivering preventive services. These summaries were compared to identify larger patterns (see Supplemental Appendix 1 for details).

RESULTS

The 18 practices included a range of small and large practices located in different-sized communities, including 8 rural, 6 suburban, and 4 urban. The 57 clinicians were from practices that ranged from 1 solo practice to a large multispecialty practice, with most practices being small groups of 2 to 4 clinicians (n = 13). Eight practices included both physicians (doctors of medicine
or doctors of osteopathy) and other allied health clinicians (nurse practitioner or physician's assistant). In the largest practice, a multispecialty practice that included relatively independent obstetric-gynecology and pediatric sections, we focused only on the family physician unit that was housed in its own wing. With the exception of 1 general surgeon, 1 medicine-pediatrics physician, and 2 general internists, all participating physicians had training in family medicine. One half of the practices were owned by larger hospital-based health systems, whereas one half were independently-owned private practices. Not unexpectedly, most of the system-owned practices were located in suburban or urban settings. Throughout this Midwestern state fee-for-service predominated, with less than 15% of the state's population enrolled in a health maintenance organization.

The sampling strategy ensured a considerable range in clinical preventive services delivery rates. Rates were calculated from 1,637 patients whose encounters were observed and charts audited; the number of patients per practice ranged from 40 for the solo practice to 272 for the largest practice. All but 2 practices had screening rates of more than 50%, with 3 practices having rates for screening services of more than 80%. Conversely, rates for both smoking counseling and immunizations were less than 50% for most practices. Immunization rates were often strongly affected by low rates of adult tetanus vaccinations because of the large denominator for this service relative to other immunizations. Whereas 2 practices provided clinical preventive services at relatively high rates across screening (89% and 76%), counseling (50% and 63%), and immunizations (60% and 60%), most practices displayed variation across these 3 types of services. For example, a small, low-volume suburban practice that emphasized wellness had the highest rate of counseling smokers (69%) but was among the lowest in administering immunizations (16%) and modest in screening (58%).

A case-by-case summary of key practice characteristics, clinical preventive service delivery rates, and service delivery approaches can be found in tabular form in Supplemental Table 1, available online only at http://www.annfammed.org/cgi/content/full/3/5/430/DC1. In addition, there are expanded case reports of 3 of the practices to provide more detailed context in Supplemental Appendix 2, available online only at http://www.annfammed.org/cgi/content/full/3/5/430/DC1.

The qualitative analyses focused on hypothesis generation and describing and understanding differences in rates of clinical preventive services delivery. The first results section below focuses on understanding the degree of prioritization of preventive services in relation to providing illness care and other competing demands, after which there are a summary of different approaches used by practices and a section on the organizational features that appeared to influence prioritization of clinical preventive services delivery.

**Competing Demands of Care**

During in-depth interviews, every physician, nurse practitioner, and physician's assistant expressed his or her endorsement of preventive services as an integral part of a 3-fold mission that included acute care and chronic illness management. All these clinicians knew of the USPSTF guidelines and were able to recall, with some accuracy, the recommendations for the common screening procedures and immunizations used in this analysis. All believed smoking was a major risk factor and that they had a role in helping patients to stop. Nevertheless, none of these practices delivered clinical preventive services at high rates across the board, although all practices performed these services at some level. For example, the small 4-clinician rural practice with the highest rates on screening (89%) and immunization (60%) had a somewhat lower rate of counseling smokers (50%). Only 1 practice had rates below 50% for all 3 of the calculated services, and most practices tended to do relatively well in 1 or 2 areas, but rarely all 3. Multiple competing demands within each practice were observed to affect these rates. Acute illness visits accounted for most encounters in all these practices, so to be financially viable, practice systems were usually geared toward maximizing efficiency in 10- to 15-minute illness visits. The result, all too commonly, was that the acute issues crowded out prevention. How practices prioritized clinical prevention relative to acute and chronic illness care appeared to be influenced by several other factors discussed in the following paragraphs.

Physicians generally prioritized 1 to several specific services, rather than the full range of recommended clinical preventive services. For example, in a small rural practice in which 1 of the physicians championed smoking counseling, but not other services, the practice had relatively high counseling rates (50%); however, the rates for screening (44%) and immunizations (17%) were quite low. Thus, different specific preventive services compete with each other, in addition to competing with illness visits, for time on the agenda.

In all practices, regardless of approach or system innovation, the clinical encounter with the physician or allied health clinician was the primary locus for delivering preventive services. While several practices had nursing and/or front office staff involved in reminders or patient intake forms, it was up to clinicians to follow through in the examination room. Even a large rural practice with a part-time health educator depended on internal referrals from clinicians. Thus, rates of delivery...
of preventive services were ultimately tied to clinician decision making during direct care of patients.

Differing patient needs and expectations resulted in separate competing demands. For example, in a cramped, 2-clinician inner-city practice, many patients complained of multiple problems at every visit, and few scheduled health care maintenance (HCM) visits. The lead physician stated that their patients were burdened with diseases of excess and needed better diets, more exercise, and smoking counseling. The clinicians used a mental checklist to discuss regularly what they saw as these patients’ most pressing issues: sexual practices, time test for tuberculosis, diet, and caffeine use. A focus on women’s health resulted in respectable screening rates (65%); however, tobacco counseling (31%) and immunization (31%) were much lower.

Health care system expectations provided other separate types of competing demands. For example, in a busy 3-clinician practice in a highly competitive suburban community, the clinicians perceived that their hospital system owners expected high volume and referrals. Many of the women in the practice allegedly received their annual preventive care from the system-owned nearby obstetrics and gynecology practice, so comprehensive HCM visits often were not scheduled; however, the practice had no documentation that these services were provided. The net effect was low screening (47%), smoking counseling (36%), and immunization (21%) rates. It is interesting to note that another 2-clinician practice faced with the same system expectations, had similarly low screening (44%) and immunization (17%) rates, but did better in tobacco counseling (50%) because 1 physician championed tobacco counseling.

Variation in Approaches for Preventive Service Delivery

Many of the office systems and innovations introduced during the past 20 years were in evidence in 1 or more of the practices. They all utilized health care maintenance (HCM) visits, particularly annual visits for birth control, childhood wellness protocols, and mandated school physicals. Many practices had prevention flow sheets, HCM visit forms and protocols, reminder systems, patient history forms, and educational materials targeting preventive issues. In 2 practices, staff completed intake forms or entered patient data during the intake process; however, in most practices, staff members were seen as overworked, and practices hesitated to consider additional prevention-related tasks. Several practices had immunization guidelines posted in examination rooms and nursing stations, and 1 practice gave nursing staff standing orders for mammogram referrals when scheduling HCM visits. Two practices had tried chart alert stickers but found they were rarely used.

Most of the practices had introduced prevention flow sheets, but, except for 2 practices, these flow sheets were usually not filled out even when in charts.

Specific approaches did not differentiate practices with higher rates from those with lower rates. The differential adoption of approaches with time led to unique eclectic strategies that incorporated various office system approaches. Examples from different practices include meticulous charting of preventive services and reviewing charts before each encounter, completion of prevention intake forms by office staff, use of electronic medical record reminders, hiring of a part-time health educator, and review of screening services by the office nurse.

Most practices had not articulated systematic strategies for delivering preventive services; instead, they defaulted to a combination of opportunistic delivery during illness visits and periodic HCM visits (mostly for children and adult women). Clinical preventive services in these practices were often provided only in response to patient request or obvious need (eg, a smoker with a respiratory complaint). Whereas some system approaches noted above were present in these practices, their use was sporadic. Few practices had systems to help patients schedule HCM visits, making reliance on HCM visits an ineffective strategy by itself.

Organizational Features that Support Clinical Preventive Services

A common pattern among practices with higher rates of clinical preventive service delivery was having one or more physician champions who made particular preventive services a practice priority. Many times the champion was the senior physician, but in a large rural practice a new partner just out of residency brought new ideas and enthusiasm that were adopted by senior physicians. Commitment to prevention often had origins in residency, but frequently it stemmed from personal history and formative experiences. For example, several physicians recounted instances in which a patient had cancer diagnosed early as a result of screening, reinforcing the value of screening. Conversely, 1 physician vividly described the impact of failing to test a stool sample for occult blood in a woman who was subsequently had colon cancer diagnosed. Others had personal or family experiences that underscored the importance of stressing prevention (eg, a physician who suffered severe head trauma in a bicycle accident while not wearing a helmet). Having a physician champion appeared to be necessary, but it was not always sufficient for success. For example, in one 3-physician practice, 1 physician who was extremely enthusiastic about implementing strategies for enhancing prevention met resistance and apathy from partners. It should also be noted that in the 8 practices with allied health
clinicians, none of those clinicians was observed to have taken a lead in championing prevention at a practice level, even though several were specifically hired by a physician champion to focus on prevention.

Most practices with higher rates of clinical preventive service delivery either found ways to make prevention financially viable or made conscious decisions to compromise incomes to provide clinical preventive services. For example, the practice with the highest screening rates had a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory that provided an economic self-interest for cholesterol screening. On the other hand, 2 practices reduced patient volume and accepted greatly reduced incomes to have more time with patients, while a couple of practices invested in hiring a health educator or nurse practitioner to enhance clinical preventive service delivery. Economic disincentives were frequently cited as an explanation for less than optimal delivery of clinical preventive services. Most practices faced financial challenges, and clinical preventive services were often perceived as not being reimbursed proportionately to the amount of time expended, particularly when they were opportunistically added into illness visits, which were generally billed according to the major focus of the encounter.

**DISCUSSION**

This study provides in-depth examinations of how preventive services are delivered in primary care practices; however, the findings must be interpreted in the context of the study’s limitations. The data were cross-sectional. The study examines clinical prevention only from a practice perspective, it does not examine the broader frame of prevention at the community and population levels. Because the sample is high in rural practices and low in large practices, reflecting the Midwestern location, it is possible that these patterns are not present in other areas of the country. To overcome these limitations, the sampling strategy deliberately selected practices representing a wide range of practice types, geographic locations, and levels of preventive service delivery. The research team spent weeks observing and interviewing in each practice to get as complete a picture as possible of the practice and its history.

All practices embrace prevention as part of a core mission and are doing it to some degree. They are often creative, adaptive, and responsive to local needs and expectations, including those of patients, community, local health care institutions, staff, and physician past experiences—these practices learn. On the other hand, only 7 practices delivered screening services at rates of 65% or above, and even the best only had rates of 69% for smoking counseling and 60% for immunizations.

What are the competing demands related to carrying out clinical prevention? Perceived patient and health system needs and expectations were at times supportive, but often at odds with a practice’s intent to provide clinical preventive services. More importantly, the encounter-centric approach to delivering preventive services assured that prevention was ensnarled in the site of greatest competing demands, including competition with acute and chronic illness care needs, patient concerns, billing issues, and pressure on physician time and productivity. Preventive services are thus squeezed into an already overcrowded clinical encounter so that even preventive services compete with each other for attention. The structure of practices may need redesign as described in the recent Institute of Medicine report and Future of Family Medicine recommendations, so some preventive services are accomplished outside the encounter whereas others are integrated into illness visits and there is greater use of information systems.

What approaches are used by practices with high vs low rates of clinical preventive service delivery? There was no best approach. Practices tended not to use systems thinking and had not developed systematic strategies for care improvement. Instead, we saw unique, eclectic strategies. These data suggest that future interventions need to raise system-level thinking and awareness and to be individualized. This suggestion is consistent with several current intervention approaches being studied.

What organizational features support clinical preventive services? Having a physician champion and making strategic economic choices were important features shared by many practices with higher clinical preventive service delivery rates. Although this finding might suggest that every practice needs a physician champion for prevention, it might also raise concerns about the limitations of this physician-centeredness. The physician-centeredness evident in these practices could complicate efforts at creating a team approach to care, empowering staff, and enhancing practice-level reflection. This study’s findings also support those who claim that national health care finance reform is needed to obtain better alignment between prevention recommendations and practice reimbursement.

A practice systems perspective suggests that efforts at getting diverse clinical offices to adopt a standardized set of processes for implementing preventive services are likely to fail regardless of the quality of the process. Future interventions need to recognize factors leading to practice variability and use this understanding to tailor interventions to the local needs of practices, their patients, and their communities. Recent approaches to systems change using participatory learning, complexity theory, and appreciative inquiry
may be helpful.\textsuperscript{33–35} Future intervention studies should consider these strategies as part of longitudinal designs that allow evolution of change over time.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/430.

Key words: Prevention; mass screening; office visits; family medicine offices; professional practice; health care quality, access, and evaluation; quality assurance, health care; quality improvement; qualitative research; Papanicolaou smear; cholesterol testing; smoking/prevention and control; mammography; immunizations

Submitted June 26, 2004; submitted, revised, April 8, 2005; accepted May 9, 2005.

Funding support: This study was supported by a grant from the Agency for Healthcare Research and Quality (R01 HS08776) and a Research Center grant from the American Academy of Family Physicians.

Acknowledgments: The authors are grateful to the physicians, staff and patients from the 18 practices, whose participation made this study possible. The authors also wish to thank dedicated work of project staff that spent countless hours collecting, transcribing and managing an extremely large amount of data.

References


Rochester Participatory Decision-Making Scale (RPAD): Reliability and Validity

Cleveland G. Shields, PhD1,2
Peter Franks, MD3
Kevin Fiscella, MD, MPH1,4
Sean Meldrum, MA1
Ronald M. Epstein, MD1,2
1Department of Family Medicine, University of Rochester Medical Center, Rochester, NY
2Department of Psychiatry, University of Rochester Medical Center, Rochester, NY
3University of California at Davis, Calif
4Department of Community and Preventive Medicine, University of Rochester, Rochester, NY

ABSTRACT

PURPOSE We wanted to develop a reliable and valid objective measure of patient-physician collaborative decision making, the Rochester Participatory Decision-Making Scale (RPAD).

METHODS Based on an informed decision-making model, the RPAD assesses physician behavior that encourages patient participation in decision making. Data were from a study of physician-patient communication of 100 primary care physicians. Physician encounters with 2 standardized patients each were audio recorded, resulting in 193 useable recordings. Transcribed recordings were coded both with RPAD and the Measure of Patient-Centered Communication (MPCC), which includes a related construct, Finding Common Ground. Two sets of dependent variables were derived from (1) surveys of the standardized patients and (2) surveys of 50 patients of each physician, who assessed their perceptions of the physician-patient relationship.

RESULTS The RPAD was coded reliably (intraclass correlation coefficient [ICC] = 0.72). RPAD correlated with Finding Common Ground \( (r = 0.19, P < .01) \) and with the survey measures of standardized patient’s perceptions of the physician-patient relationship \( (r = 0.32 - 0.36 [P < .005]) \) but less with the patient survey measures \( (r = 0.06 to 0.07 [P < .005]) \). Multivariate, hierarchical analyses suggested that the RPAD made a more robust contribution to explaining variance in standardized patient perceptions than did the MPCC Finding Common Ground.

CONCLUSIONS The RPAD shows promise as a reliable, valid, and easy-to-code objective measure of participatory decision making.


INTRODUCTION

Participatory decision making has been reported to affect health outcomes, including control of chronic disease1 and functional outcomes.2 Based on those early results and more recent studies that show a lack of patient involvement in decisions,3 physicians have been encouraged to adopt a more participatory style. Some consider that participatory decision making is a moral imperative in medicine without regard to its impact on outcomes.4 The outcomes of efforts to improve participatory decision making have been mixed; although effects on consultation style and satisfaction have been reported,5,6 effects on control of chronic disease have not been replicated.7 These studies have often relied on patient surveys to assess participatory decision making; a validated observational instrument would provide a more objective description of behaviors and reduce the likelihood of confounding by including both measures of participatory decision making and reported outcomes on the same patient survey.

Participatory decision making emerged in the 1970s as an alternative to a more traditional paternalistic model in which physicians made decisions for their patients;8-12 initially it was influenced by consumerist and models of care, which suggest that patients have the right to information and self-determination.13,14 A contractual model elaborated on the consumerist.
model by emphasizing the importance of taking into account patients’ stated values to arrive at decisions. Participatory decision making is probably most closely related to a deliberative model in which physicians elicit and respect patients’ values, but physicians also offer expertise and recommendations, sometimes using persuasion to adopt healthier options if there is not initial consensus. Thus, participatory decision making consists of 2 processes: expert problem solving and decision making. Problem solving is the province of physicians whose expertise informs their judgment to determine treatment options. Decision making involves patients working with the physician to determine which treatment options best satisfy the patient’s preferences.

Measurement of the process of participatory decision making has been elusive. Patient surveys may not capture the level of detail to inform physician training interventions. Current interaction analysis systems, such as the Measure of Patient-Centered Communication (MPCC) and the Roter Interaction Analysis System (RIAS), offer some key behaviors that may be indicators of participatory decision making (patient question-asking), but not others. Braddock et al developed an instrument derived from a consensually derived set of behavioral criteria for “informed” decision making. Using their criteria, informed decision making occurs in only 9% of primary care office visits, raising concerns that physicians need to develop better skills in involving patients in their care. Despite its usefulness as a descriptive measure to define the conceptual domains of informed decision making, this instrument has some limitations; there is no overall scale score, and criterion validity has not been reported.

Many of the models described above focus on information sought, offered, and received. But participatory decision making also includes the responsiveness of physicians to a richer range of patient participation in decisions beyond assuring that patients have been informed. Using the Braddock et al scale as a starting point, we sought to develop a reliable and valid objective measure of physician behaviors that encourage participatory decision making. We developed new items and a simple method of scoring the scale to construct the Rochester Participatory Decision Making Scale (RPAD). While it is clear that patients also bring attitudes and behaviors that contribute to participatory decision making, our scale was developed to evaluate physician communication behavior and to be used for physician training purposes, rather than as a purely descriptive measure of conversational process. For this reason, we used unannounced and covert standardized patients to reduce patient variability so that we could observe the differences in physician participatory decision-making behavior when confronted with a nearly identical stimulus.

**METHODS**

The RPAD was developed as part of a larger study that examined the relationship between physicians’ communication behaviors and health care costs. The larger study involved audio recording and coding standardized patient visits to physicians, surveys of standardized patients (measuring their perceptions of the encounter), physician surveys (personality and demographics), patient surveys (measures of the patient-physician relationship, satisfaction, demographics, illness morbidity, physical and mental functioning), and claims data from a large managed care organization.

**Research Participants**

We had 3 sets of participants in this study: primary care physicians, standardized patients, and real patients. One hundred primary care physicians (internists and family physicians) who were members of the independent practice association of a managed care organization were recruited and enrolled in the study. Standardized patients made 2 unannounced, covert, audio-recorded visits to physicians. The first standardized patient role was constructed to mimic typical patients in primary care with straightforward symptoms of gastroesophageal reflux (GERD case). The second role was designed to simulate patients with medically unexplained symptoms so we could explore how physicians handle situations that involve potential disagreements about the meaning of symptoms, the diagnosis, and its treatment (ambiguous case). Two male and 3 female standardized patients were used. All visits were audio recorded with recorders hidden in purses and backpacks.

The order of standardized patient visits (male or female, role) was randomized for each physician. In the treatment and planning phase of the office visit, standardized patients were instructed to respond to physicians’ questions and to ask clarifying questions, but they were not to challenge directly the physician’s assessment. At one point during each visit, however, standardized patients were instructed to ask whether their symptoms could represent something serious so they could communicate to the physician a moderate level of anxiety. Thus, we sought to create typical patients in current primary care practice. Standardized patients participated in a pilot test to assure they were realistic, and we sought feedback from pilot physicians on whether the standardized patients seemed typical and ordinary.

Physicians completed questionnaires, and 50 visiting patients from each physician’s office were also recruited to complete questionnaires. We approached 4,963 eligible patients; 4,746 (95.6%) completed the questionnaire. The reasons for refusal were as follows: 185 patients stated that they disliked questionnaires, 52 felt rushed because of illness, and 52 felt rushed.
Demographic information on the physician and patient samples is contained in Tables 1 and 2.

Two days after the visit, a fax was sent to the physician to determine whether, when prompted, the physician could identify the standardized patient. The fax notified the physician that a standardized patient had visited in the past few days; the physicians were asked whether they suspected they had seen an standardized patient, and if so, to describe the patient and indicate how realistic the standardized patient portrayal was. Forty percent of physicians identified the standardized patients from this prompted recall.

**Analysis of Audio-Recorded Encounters**

Each standardized patient visit was recorded using a digital audio disk recorder with a high-quality microphone. Visit length was calculated (in minutes), excluding waiting time in the examining room before the visit and any period of more than 1 minute during which the physician left the room.

**RPAD Scale Development**

The RPAD was developed by incorporating items suggested by Braddock et al as indicative of physician behaviors that encourage patient participation in decision making. In developing the RPAD, we observed that some physician behaviors were performed fully, whereas others were completed only partially. This finding led us to create a coding scheme for each item that gave a score of 0 for no evidence of the behavior, ½ for partial presence of the behavior, and 1 for the full presence of the behavior (Table 3). We developed a coding manual with descriptions and examples for each 0, ½, and 1 score to guide raters (available from the first author).

We pilot tested the scale on 10 audio-recorded visits. We discontinued items that never received a code. We were left with 4 items; we then developed 5 more items and scoring criteria for each and pilot tested them. The final coding system is shown in Table 3. The 10 visits we used to develop the scale were recoded after all other tapes had been coded and used as data in the analysis.

We have included the discarded items in the Supplemental Appendix, available online only at http://www.annfammed.org/cgi/content/full/3/5/436/DC1.

Coders first listened to the entire audio recording and then listened again to code the instances of physician behaviors listed on the RPAD coding sheet. Each time they found an example, they stopped the tape and listened again to that section to determine whether the behavior deserved a 0, ½, or 1 full-point score.

**The MPCC**

We also coded using the MPCC, a measure of physician responsiveness to patient concerns, including participation in care. See the Supplemental Appendix for information about the MPCC.

**Patient Survey**

Patient questionnaires that were administered to 50 patients of each physician included 4 scales: the 5-item Health Care Climate Questionnaire (HCCQ), the Primary Care Assessment Survey (PCAS) knowledge and trust subscales, and a single-item satisfaction scale. Details can be found in the Supplemental Appendix.
Patient data for covariate adjustment were also collected, including demographics (age, sex, race/ethnicity, and educational level), health status medical and physical component scores of the SF-12 Health Survey (MCS-12 and PCS-12),24 SCL-90 (Symptom Checklist – 90) somatization score,25 11 patient-reported morbidities, and the length of the physician-patient relationship.

**Standardized Patient Survey**

The standardized patients also completed questionnaires after their visits with physicians. The HCCQ23 and the PCAS trust subscale were completed by both patients and standardized patients.22,23,26

**Statistical Analysis**

We examined the coding reliability of the RPAD by calculating the intraclass correlation coefficient (ICC). We also examined the case-to-case reliability of the RPAD coding of the 2 standardized patient cases as a measure of physician style using the Spearman Brown prophecy formula $\alpha = n\gamma/[1+(n-1)\gamma]$ ($n =$ number of standardized patient cases and $\gamma =$ average correlation between cases). This formula treats the 2 cases as items in a scale assessing the physician's style and calculates a coefficient of reliability. We then examined the relationship of RPAD with MPCC total score and its components. We expected the measures to be moderately related, but our primary hypothesis was that RPAD would correlate with Component 3, because MPCC measures physician-patient interaction around the delivery of the diagnosis and treatment plan. Finally, we examined the criterion validity.

<table>
<thead>
<tr>
<th>Table 3. Rochester Participatory Decision-Making Scale (RPAD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Items</strong></td>
</tr>
<tr>
<td>1 Explain the clinical issue or nature of the decision*</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Physician gives a cursory, hurried, unclear, rushed explanation, or long confusing lecture</td>
</tr>
<tr>
<td>1 Physician clearly explains his/her view of the medical/clinical problem</td>
</tr>
<tr>
<td>2 Discussion of the uncertainties associated with the situation*</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Physician acknowledges uncertainties but does not explain thoroughly or only does with active patient prompting</td>
</tr>
<tr>
<td>1 Physician thoroughly explains uncertainties in the problem or treatment</td>
</tr>
<tr>
<td>3 Clarification of agreement</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Patient expressed passive assent</td>
</tr>
<tr>
<td>1 Physician actively asks for patient agreement and tries to obtain a commitment from the patient to the treatment plan</td>
</tr>
<tr>
<td>4 Examine barriers to follow-through with treatment plan</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Patient discloses concerns or problems with following through with treatment</td>
</tr>
<tr>
<td>1 Physician actively examines patients concerns or problems with following through with treatment</td>
</tr>
<tr>
<td>5 Physician gives patient opportunity to ask questions and checks patients understanding of the treatment plan*</td>
</tr>
<tr>
<td>0 No opportunity for patient to ask questions</td>
</tr>
<tr>
<td>½ Patient has opportunity to ask questions</td>
</tr>
<tr>
<td>1 Physician asks patients for their understanding of problem or plans</td>
</tr>
<tr>
<td>6 Physician’s medical language matches patient’s level of understanding</td>
</tr>
<tr>
<td>½ Clear mismatch between the technicality of physician’s and patient’s language</td>
</tr>
<tr>
<td>½ Level of technicality or detail of the physician’s and patient’s language matches most of the time.</td>
</tr>
<tr>
<td>1 Level of technicality or detail of the physician’s and patient’s language clearly matches.</td>
</tr>
<tr>
<td>7 Physician asks, &quot;Any questions?&quot;*</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Yes, but no discussion ensues</td>
</tr>
<tr>
<td>1 Yes, and physician engages in a discussion with patient about the question</td>
</tr>
<tr>
<td>8 Physician asks open-ended questions.</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Yes, but no discussion ensues</td>
</tr>
<tr>
<td>1 Yes, and physician engages in a discussion with patient about the question</td>
</tr>
<tr>
<td>9 Physician checks his/her understanding of patient’s point of view*</td>
</tr>
<tr>
<td>0 No evidence</td>
</tr>
<tr>
<td>½ Yes, but no discussion ensues</td>
</tr>
<tr>
<td>1 Yes, and physician engages in a discussion with patient about the physician’s perceptions of patients</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
</tr>
</tbody>
</table>

**Discarded items**

| Discussion of the patient’s role in decision making* |  |
| 0 No evidence |  |
| ½ Yes, but no discussion ensues |  |
| 1 Yes, and physician engages in a discussion with patient about the patient’s role |  |
| Discussion of the alternatives* |  |
| 0 No evidence |  |
| ½ Yes, but no discussion ensues |  |
| 1 Yes, and physician engages in a discussion with patient about the alternative treatments available |  |
| Discussion of the pros (potential benefits) and cons (risks) of the alternatives* |  |
| 0 No evidence |  |
| ½ Yes, but no discussion ensues |  |
| 1 Yes, and physician engages in a discussion with patient about the pros and cons of the alternative treatments |  |

* Indicates modified Braddock items.
by examining the relationship of RPAD with patients’ and standardized patients’ perceptions of their relationships with their physicians using multivariate methods. We were particularly interested in the contribution that the RPAD variable made to patient and standardized patient perceptions independent of the other objective measure of physician-patient interaction (MPCC). The multivariate analysis methods and the results are included in an online Supplemental Appendix.

RESULTS

We analyzed 193 audio recordings from 100 physician-patient encounters. Seven recordings were not available because of equipment failure (3 encounters); 4 physicians moved their practices before completion of the study. We averaged 49.4 (SD = 6) patient questionnaires from each physician’s office. Patients reported an average of 1.25 illnesses from a list of 13 commonly treated primary care conditions. (Detailed information on patient illnesses and health status is included in the on-line Supplemental Table 1, available online-only at http://www.annfammed.org/cgi/content/full/3/5/436/DC1.)

Reliability of the RPAD

The ICC for the RPAD was 0.72. Reliability for the RPAD as a measure of physician style, using the Spearman-Brown prophecy formula based on the 2 standardized patient encounters, was 0.53. Audio-recorded encounters took approximately 50 minutes to code; 20 minutes were spent first listening to the tape, and another 30 minutes to code the 20 minutes of the recording.

RPAD Distribution and Scoring

Table 4 shows the distribution of scores on the RPAD. Each item was scored 0, ½, or 1, but when averaged over 2 cases, the scores also included ¼ and ¾. Almost 70% of the physicians gave a clear description of the clinical problem, though 53% did not discuss uncertainties in any way. Almost all the physicians attempted to clarify agreement on the diagnosis and treatment plan, 98% had at least a score of ½ or higher. Most physicians, 93%, did not discuss barriers to carrying out the treatment plan. The bulk of patients, 92%, were given some opportunity to ask questions. Most of the time, physician language matched the patients’. More than 25% of the time, physicians asked whether patients had any questions. A small percentage of physicians used open-ended questions, and a similarly small percentage checked patients’ understanding.

Table 4. Rochester Participatory Decision-Making Scale (RPAD) Descriptive Statistics

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain the clinical issue</td>
<td>0.89</td>
</tr>
<tr>
<td>Discuss uncertainties</td>
<td>0.20</td>
</tr>
<tr>
<td>Clarify agreement</td>
<td>0.57</td>
</tr>
<tr>
<td>Examine barriers</td>
<td>0.02</td>
</tr>
<tr>
<td>Patients asked questions</td>
<td>0.49</td>
</tr>
<tr>
<td>Physician’s medical language</td>
<td>0.55</td>
</tr>
<tr>
<td>Physician asks, Any questions?</td>
<td>0.25</td>
</tr>
<tr>
<td>Physician asks open-ended questions</td>
<td>0.07</td>
</tr>
<tr>
<td>Physician checks understanding</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency and Percentage*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>53</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>93</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>46</td>
</tr>
<tr>
<td>84</td>
</tr>
<tr>
<td>77</td>
</tr>
</tbody>
</table>

* In this table, the frequency is per 100 cases, so percentage is equal to frequency.
† Items for the RPAD were scored 0, ½, and 1, averaged over 2 cases.
Regression of RPAD on Patient Surveys and Standardized Patient Surveys

We conducted multilevel regression analyses examining the regression of patient survey perception measures on the RPAD and MPCC components. The optimal models for all 4 patient perception measures, based on Akaike’s and Bayes information criteria and physician variance component reduction, were the models including RPAD and MPCC Component 1 and Component 2, but not Component 3 (Supplemental Table 2, available online only at http://www.annfammed.org/cgi/content/full/3/5/436/DC1).

We conducted a similar series of regression analyses of the standardized patient survey measures on the RPAD and MPCC components. Again, the optimal models for each of the survey measures were the models including RPAD and MPCC Component 1 and Component 2, RPAD but not Component 3 (Supplemental Table 3, available online only at http://www.annfammed.org/cgi/content/full/3/5/436//DC1).

Consistent with the univariate Pearson correlations, the parameter estimates for the standardized patient survey measures were much larger than those for the patient measures in terms of standard deviation units on the scales examined. For the standardized patient measures, a 1 SD difference in participatory decision making was associated with a 30.3% SD difference in HCCQ and a 25.6% SD difference in satisfaction, whereas for the patient perception measures, a 1 SD difference in RPAD was associated with only a 4.8%-6.1% SD difference in measures of patient perceptions of autonomy support, physician knowledge of patient, trust, and satisfaction.

DISCUSSION

We report exploratory data on a new quantitative objective measure of participatory decision making. The RPAD can be coded reliably, correlates with standardized-patient and real-patient measures of constructs related to participatory decision making, and takes only 50 minutes to code 20-minute office visits. Based on the Braddock et al scale and other literature on participatory decision making, the scale items have face validity.28,29 The scale items address behaviors that physicians use to encourage patient participation in decision making. A difference between our scale and the Braddock et al scale is that we set out to capture physician behaviors that might encourage patient participation, whereas the Braddock et al scale focuses on behaviors that should have occurred during informed decision making. Although we developed the measure in conjunction with our use of the MPCC, we think that the RPAD could be used independently of the MPCC.

The use of standardized patients is both a strength and a weakness of the study. We do not know how the RPAD might work with real patients; however, by using standardized patients, we focused on the physician as an agent encouraging participatory decision making rather than on measuring patient participation in decision making. Future studies should examine using RPAD with real patients.

Because there are no reliable measures of participatory decision making, it was challenging to establish construct validity of the scale. The closest we came to evidence of construct validity was the correlation of MPCC Finding Common Ground with the RPAD. It is difficult to determine whether the modest correlation reflects poor reliability of the MPCC Finding Common Ground subscale or that the 2 scales share variance but measure somewhat different constructs.

Interestingly, RPAD correlated with the MPCC Exploring the Disease and Illness Experience subscale. This finding suggests that the RPAD scale is tapping into other communication processes that are important to patient centered care, or that exploring disease and illness experience is a necessary precursor to participatory decision making. The RPAD includes items that measure physicians’ use of active encouragement for patients to express their ideas and thoughts about the treatment plan. Thus, it includes domains that may not be captured using the MPCC Finding Common

Table 5. Correlation of RPAD Score With Self-Report Measures

<table>
<thead>
<tr>
<th>Patient Self-Report</th>
<th>RPAD Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding of audiotapes (n = 193)</td>
<td>0.24*</td>
</tr>
<tr>
<td>Total MPCC score</td>
<td>0.18*</td>
</tr>
<tr>
<td>C1 - Exploring the Disease and Illness</td>
<td>0.06</td>
</tr>
<tr>
<td>C2 - Whole Person</td>
<td>0.08</td>
</tr>
<tr>
<td>C3 - Diagnosis and Treatment</td>
<td>0.14</td>
</tr>
<tr>
<td>Physician characteristics (n = 193)</td>
<td>0.36*</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
</tr>
<tr>
<td>Female</td>
<td>0.08</td>
</tr>
<tr>
<td>Years in practice</td>
<td>0.02</td>
</tr>
<tr>
<td>Solo practice</td>
<td>0.14</td>
</tr>
<tr>
<td>Number of partners</td>
<td>0.07*</td>
</tr>
<tr>
<td>SP survey (n = 193)</td>
<td>0.07*</td>
</tr>
<tr>
<td>Health care climate</td>
<td>0.07*</td>
</tr>
<tr>
<td>Trust in physician</td>
<td>0.14</td>
</tr>
<tr>
<td>Patient survey (n = 4,746)</td>
<td>0.07*</td>
</tr>
<tr>
<td>Health care climate</td>
<td>0.07*</td>
</tr>
<tr>
<td>Knowledge of patient</td>
<td>0.07*</td>
</tr>
<tr>
<td>Trust in physician</td>
<td>0.36*</td>
</tr>
<tr>
<td>Patient satisfaction</td>
<td>0.06*</td>
</tr>
</tbody>
</table>

RPAD = Rochester Participatory Decision-Making Scale; MD-SP = physician-standardized patient; MPCC = Measure of Patient-Centered Communication; C1, C2, C3 = Components 1, 2, 3.

* P < .005.
† P < .01.
Ground subscale, which focuses more on patient question asking, but does not address whether the physician actively encouraged the patients’ participation.

RPAD significantly contributed to the model explaining variance in the degree to which the standardized patients believed that their autonomy was supported by physicians, lending convergent validity. Because no similar relationship was found for MPCC Finding Common Ground subscale, the RPAD may capture the construct of patient-perceived participatory decision making at least as well as other available objective instruments. Not surprisingly, RPAD did not account for as much variance in patient surveys as it did with standardized patient surveys. Patients’ tendency to accommodate to their physician's communication style may have caused them to judge their physicians' less critically than standardized patients did, thus muting the association between communication style and patient perceptions of their physicians. In addition, the standardized patients were reporting their perception of the same encounter that was coded using the RPAD, whereas the patients were reporting their perceptions about their ongoing relationship with the physician. Finally, patients’ perceptions were correlated with a measure of physician style assessed from physicians. It is possible that correlations with real patients’ perceptions of their physicians would be stronger had the interactions been with the real patients. These preliminary findings suggest that the RPAD offers promise as a reliable, valid, and easy-to-code objective measure of participatory decision making.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/436.

Key words: Physician-patient relations; medical decision making; informatics

Submitted September 7, 2004; submitted, revised, February 21, 2005; accepted February 28, 2005.

Funding support: This project was supported by grant No. R01HS10610 from the Agency for Healthcare Research and Quality (Dr. Epstein).

References

Social Network Analysis as an Analytic Tool for Interaction Patterns in Primary Care Practices

John Scott, MD, PhD1,2
Alfred Tallia, MD, MPH1,2
Jesse C. Crosson, PhD1,2,4
A. John Orzano, MD, MPH1,2
Christine Stroebel, MPH1,2
Barbara DiCicco-Bloom, RN, PhD1,2
Dena O’Malley, BA1,2
Eric Shaw, PhD1,2
Benjamin Crabtree, PhD1,2,3
1Department of Family Medicine, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, New Brunswick, NJ
2Center for Research in Family Practice and Primary Care, Cleveland, Ohio
3Cancer Institute of New Jersey, New Brunswick, NJ
4University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, NJ

ABSTRACT

PURPOSE Social network analysis (SNA) provides a way of quantitatively analyzing relationships among people or other information-processing agents. Using 2 practices as illustrations, we describe how SNA can be used to characterize and compare communication patterns in primary care practices.

METHODS Based on data from ethnographic field notes, we constructed matrices identifying how practice members interact when practice-level decisions are made. SNA software (UCINet and KrackPlot) calculates quantitative measures of network structure including density, centralization, hierarchy and clustering coefficient. The software also generates a visual representation of networks through network diagrams.

RESULTS The 2 examples show clear distinctions between practices for all the SNA measures. Potential uses of these measures for analysis of primary care practices are described.

CONCLUSIONS SNA can be useful for quantitative analysis of interaction patterns that can distinguish differences among primary care practices.


INTRODUCTION

Primary care practices are complex systems that are characterized by dynamic patterns of interactivity among practice members and their environment.1-3 One feature of complex systems is the property of emergence, which is the tendency of organized patterns to emerge that cannot be predicted from the properties of individual parts of the system.4 Thus, to understand how primary care practices function, it is necessary to study not only the individuals within the practice or individual practice components but also the relationships among individuals.5 Study of such patterns and how they change with time or in response to interventions requires an ability to look at the entire complex web of relationships and interactions within a primary care practice. Although qualitative description6-10 and practice genograms11 have demonstrated utility for understanding the complex interactions in practices, a tool that captures quantitative aspects of the patterns of relationships within practices would be a useful aid in studies of primary care practices. Social network analysis (SNA) is such a tool.

SNA combines the concept of the sociogram (a visual representation of relationships in a social group) with elements of graph theory to analyze patterns of interaction among people in various kinds of networks, allowing quantitative comparisons between different network structures.12 There is a large body of scholarly literature describing the use of SNA. Much of this work addresses the basic science of SNA, that is the development of
theoretical models of network organization and the mathematical derivation of quantitative measures of network characteristics. More recent work examines the association of these quantitative measures with organizational performance outcomes. Cummings and Cross, for example, found that degree of hierarchy, core-periphery structure, and structural holes of leaders correlated negatively with performance in 182 work groups in a large telecommunications company, and Aydin et al found that increased network communication density was associated with higher use of an electronic medical record system by nurse practitioners and physician’s assistants. There have also been studies showing how network parameters change with time. Shah, for example showed that network centrality decreased after downsizing in a consumer electronics firm, whereas Burkhardt and Brass documented increased network centrality after introduction of a new computer system in a federal agency.

In this article, using data from 2 primary care practices as examples, we detail SNA measures that can be used to quantify patterns of decision making and discuss how these measures could be used to facilitate the design and measure the outcomes of interventions to change organizational behavior in primary care practices.

METHODS

The data for the example practices were collected as part of a large, group-randomized clinical trial of primary care practices designed to evaluate a practice improvement intervention funded by the National Heart, Lung, and Blood Institute. Trained observers collected data for 2 weeks in each practice. They directly observed interactions among the office staff and clinicians, conducted key informant interviews of clinicians and staff, tape-recorded in-depth interviews with a sample of practice participants, and collected samples of various practice documents. Observers recorded brief jottings throughout the day and expanded these each evening into extensive field notes. A practice environment template, which detailed data domains for observation, guided data collection. These baseline data were available on 30 primary care practices in the intervention group. We chose the 2 example practices for this article because they illustrated maximal variation in methods of decision making based on our initial analysis of the qualitative data.

Construction of the SNA Matrix

SNA analyzes patterns of connections (ties) among information-processing agents (nodes). Data are recorded in the form of an adjacency matrix, where each node is assigned both a column and a row in the matrix. A matrix constructed in this way will have two cells representing the intersection of any 2 nodes, 1 above and 1 below the diagonal. If a connection or tie exists between 2 nodes, then a 1 (or another positive number representing the strength of the tie) is entered in the matrix cell representing the intersection of these 2 nodes. If no tie exists, then a 0 is entered. It is not necessary that the 2 cells for each pair of notes have the same value. For example, in Figure 1A, A consults B when making a decision, but B does not consult A. In this case, the cell at the intersection of row A and column B would contain a 1, but the intersection of row B and column A would contain a 0. A matrix that has this property is called a directed matrix. If the tie being studied is not directional (ie, if A talks to B, then B must also talk to A), then the matrix cells below the diagonal are identical to those above the diagonal and are ignored in the SNA calculations (Figure 1B).

SNA data are usually collected using interviews or surveys. Each member of the network under study is asked to identify every other member with whom he or she has the interaction. For example, a survey item for an SNA study designed to identify decision-making patterns in a network might be constructed as follows: “When you need to solve a problem related to your work, whose input (by e-mail or telephone, or in person) do you regularly seek?” Survey or interview responses are then tabulated and entered in an adjacency matrix as described above.

Because our data consisted of ethnographic field notes rather than survey responses, our approach to constructing the SNA matrices was somewhat more
complex than the method described above. A team consisting of the observers who had recorded the data for the practices and a family physician analyzed the field notes and the transcribed interviews. Each team member individually constructed a social network adjacency matrix for both practices using a spreadsheet (Microsoft Excel) with each practice member assigned both a column and a row. For every person represented on a row of the spreadsheet, team members used qualitative data to answer the question, “Whom does this person consult when significant decisions need to be made in the practice?” For each column in the row, the entry of a 1 or 0 indicated the presence or absence of consultation of that practice member by the person represented on the row. This resulted in a directed as opposed to an undirected matrix, meaning in this case that connections between members were not necessarily reciprocal. Reciprocal connections would occur only if 2 members consulted each other when making decisions.

Unlike structured interviews or surveys, these ethnographic field notes may not provide specific information about each pair of practice members, however, they provide important contextual data for interpreting the SNA results. Statements or observations often describe the interaction of an individual with a group of practice members or interactions between groups. Thus we were able to assign values in the matrix to all the members in each group. In practice 1, for example, we assigned 0 to all the columns for the practice leader, whom we will call “Dr. Smith,” indicating that he consulted no one in the practice when making decisions. The following illustrates how data elements from observation, in-depth interviews, and key informant interviews were used to assign those values:

Observations—Practice leader: “Dr. Smith seems to have a very hands-on approach to management of his office. Regarding office organization in decision making, I noticed that most of the decision-making seems to come from Dr. Smith and that there is little in the way of staff empowerment. Decisions are communicated throughout the practice via memo and not very systematically.”

In-depth interviews—Office manager: “Long-term goals are set by Dr. Smith, who has a clear goal and change is managed principally by Dr. Smith. Decision-making tends to be top-down and all financial management decisions are made by Dr. Smith.” Front desk supervisor: “When significant changes need to be made, Dr. Smith makes them. Dr. Smith often sets policies and doesn’t follow them himself.” Dr. Smith (referring to a new electronic medical record system): “I came up with those plans and I tried to get (the office manager) and the staff to enact the goals I recommended.”

Key informant interviews—The office manager: “Reports that Dr. Smith generally doesn’t appreciate some of the expertise around. She describes a sort of micromanaging that he does.” The nursing supervisor: “Reports that one thing that she does want to mention and make sure it gets into the notes is that Dr. Smith often does not hear when he is spoken to.”

Matrix values for the remaining practice members in practice 1 and practice 2 were assigned in similar fashion using the 3 kinds of qualitative data noted above. The team met together to share the individual matrices. There were only minimal differences among the observers, which were resolved by consensus. The resulting matrices were then imported into UCInet and Krack-Plot software to compute several quantitative measures of network structure for each of the 2 practices.

**RESULTS**

**SNA Network Diagrams**

Using NetDraw, a program within the UCInet suite, we constructed a visual representation of the networks, illustrating the web of decision-making consultations in each entire network. The network diagrams from our example practices are shown in Figure 2. An arrow leading from 1 member to another indicates that the first member consults the second member when making decisions. A double-headed arrow between 2 members indicates that both members consult each other in decision making.

It is clear from inspection of these 2 diagrams that decision-making consultation patterns are different in these 2 practices. Practice 1 has a hierarchical structure. There are no 2-way arrows. The practice leader, MD1, holds all final decision-making authority, yet he seldom consults the office staff when he makes decisions. Practice 2 has a much more collaborative decision-making process, with members empowered to make some decisions on their own. The network diagram is much denser, and there are many 2-way arrows. The visually apparent differences in the 2 practices can be quantified, as described below.

**SNA Quantitative Measures**

Some of the quantitative measures developed for use in SNA include density (a measure of the relative number of connections), clustering coefficient (the tendency of the network to aggregate in subgroups), centralization, (the degree to which a network approaches a perfectly symmetric or “star” network), and hierarchy (the extent to which network relations are ordered). We used these measures in our examples and constructed these measures as follows.

**Network Density**

The first measure constructed was network density. The density of a network is the number of actual connections between members divided by the number of pos-
Figure 2. Network diagrams.

Practice 1

 MD1  Practice Owner - MD
 NS  Clinical Supervisor
 CB  Chief Biller
 RS  Referral Specialist
 OM  Office Manager
 FOS  Front Office Supervisor
 NS  Nursing Supervisor

 MD2  Partner - MD
 PA1  Physician’s Assistant
 PA2  Physician’s Assistant
 FO1  Front Office
 FO2  Front Office
 FO3  Front Office
 FO4  Front Office

 FO5  Front Office
 MA2  Medical Assistant
 MA3  Medical Assistant
 MA4  Medical Assistant
 MA5  Medical Assistant
 MA6  Medical Assistant

 Practice 2

 MG  Management Group
 PS2  Practice Site 2
 OM  Office Manager
 FO1  Front Office
 FO2  Front Office
 FO3  Front Office
 FO4  Front Office
 FO5  Front Office

 FO6  Front Office
 F1  Filing
 F2  Filing
 F3  Filing
 R  Referrals
 MD1  Lead Physician
 MD2  Physician
 MD3  Physician

 MD4  Physician
 MD5  Physician
 NP  Nurse Practitioner
 RN1  Nurse – RN
 RN2  Nurse – RN
 RN3  Nurse – RN
 RN4  Nurse – RN
 RN5  Nurse – RN
 RN6  Nurse – RN
 RN7  Nurse – RN
 RN8  Nurse – RN
 RN9  Nurse – RN
 RN10  Nurse – RN
 RN11  Nurse – RN
 RN12  Nurse – RN

 Circles - Females
 Squares - Males
 Triangles - Outside Organizations
possible connections. Density values range from 0 to 1. In these decision-making networks, higher density indicates a greater degree of interaction among the members in the process of making decisions. In our example practices, the practice 2 network is more than twice as dense as the practice 1 network, indicating that there are many more decision-making interactions in practice 2 than in practice 1 (Table 1).

Clustering Coefficient
The clustering coefficient is the likelihood that any 2 nodes that are connected to the same node are connected themselves. For these networks, the clustering coefficient measures the degree to which decision making is done in collaborative groups. In our example practices, the clustering coefficient in practice 2 is more than 2.5 times greater than in practice 1, indicating that more collaborative groups exist in practice 2 than practice 1 (Table 1).

Hierarchy
Hierarchy is the degree to which the network approaches a perfect hierarchy, that is, the degree to which all relations are unidirectional. An organization chart would be an example of a perfect hierarchy. Hierarchy values were calculated using KrackPlot software. Practice 1 has a very high hierarchy value, whereas the hierarchy value of practice 2 is 0 (no unidirectional relations) (Table 1).

Centralization
Centralization is the degree to which a network approaches the configuration of a “star” network. A star network has 1 node in the center that connects to all other nodes. No nodes have connections to any other node except the central node. The centralization score is expressed as a percentage and can vary from 0 (every member is connected to every other member) to 100 (all members are connected to only 1 member). The centralization percentage thus indicates the degree of asymmetry in the distribution of connections in the network. A high centralization score indicates that some members have many more connections than others. For these directed decision-making networks, there are 2 components to centralization. High indegree centralization would indicate that a small number of members are consulted by the rest of the members. High outdegree centralization would indicate that a small number of members do most of the consulting of others. Table 1 shows the indegree and outdegree centralization percentages for practices 1 and 2. Both show a relatively high degree of centralization, but practice 1 has much more marked asymmetry in the indegree and outdegree percentages than practice 2.

**DISCUSSION**

It is easy to see from the network diagrams that decision-making patterns differ widely in our 2 example practices. The SNA measures, however, allow us to compare several aspects of those pattern differences quantitatively. Nevertheless, there are, some limitations of both the method and our sample data that should be kept in mind. SNA differentiates members of a network only by the pattern of interactions of each member. It is not a good tool, therefore, for measuring how individual differences in members affect the function of an organization. The quantitative measures provided here are for illustrative purposes to demonstrate the kinds of SNA measures that can be used for structural analysis of organizations. We recognize that because we calculated these measures from qualitative data, which required imputing data for some ties between individual nodes, they may be different than had they been calculated from survey data.

The measures we calculated have been associated with performance outcomes in other organizations and have been used to measure organizational change with time. Although we cannot make any statistical inferences from the differences in only 2 practices, adaptations of regression and correlation statistics have been developed for use with network data that provide a practical way to examine association of network parameters with performance outcomes in large numbers of practices. We might hypothesize based on the outcome studies cited above, for example, that network density would correlate positively with practice adherence to treatment guidelines for hypertension, and that degree of hierarchy would correlate negatively. These hypotheses could be tested using network data. Significance testing using SNA measures could also be used in randomized controlled trials to evaluate the impact on decision making or other communication patterns of interventions aimed at organizational change.

Another possible use of SNA quantitative measures would be to help tailor interventions in practices based on their network parameters. An intriguing article by McGrath and Krackhardt explores how different network configurations might predict what sorts of

<table>
<thead>
<tr>
<th>Measure</th>
<th>Practice 1</th>
<th>Practice 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network density (SD)</td>
<td>0.1570 (0.3638)</td>
<td>0.3398 (0.4136)</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.223</td>
<td>0.590</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>0.93</td>
<td>0.00</td>
</tr>
<tr>
<td>Centralization – indegree, %</td>
<td>73.440</td>
<td>57.889</td>
</tr>
<tr>
<td>Centralization – outdegree, %</td>
<td>6.880</td>
<td>64.778</td>
</tr>
</tbody>
</table>

Table 1. Social Network Analysis Quantitative Measures
interventions would be most effective in producing sustainable organizational change.

We used decision-making patterns in this analysis, but the technique would lend itself to the study of any other observed interaction among agents in the complex social systems of primary care practices, such as office communication patterns, clinical information flow, or referral patterns. SNA could also be used to complement and triangulate other quantitative measures of organizational performance that use survey methodology.\(^\text{28}\)

SNA is a useful tool for quantitative analysis of the complex systems represented by primary care practices. A broad range of potential applications of this tool is possible, including using it to help design interventions to promote practice organizational change.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/443.

**Key words:** Evaluation studies; health care delivery; health services research; social networks

Submitted October 3, 2004; submitted, revised, April 21, 2005; accepted April 25, 2005.

**Funding support:** This study was funded by a National Heart Lung Blood Institute Grant RO1 HL070800-01. Analysis of these data was also supported in part by a Research Center grant from the American Academy of Family Physicians (Center for Research in Family Practice and Primary Care).

**Acknowledgments:** The authors wish to thank the clinicians and staff of the practices who were willing to open their practices up for us.

**References**


Efficacy and Tolerability of Tricyclic Antidepressants and SSRIs Compared With Placebo for Treatment of Depression in Primary Care: A Meta-Analysis

Bruce Arroll, MBChB, PhD
Steve Macgillivray, MA
Simon Ogston, PhD, MA, MSc
Ian Reid, MB, PhD, MRCPsych
Frank Sullivan, PhD, FRCP, FRCGP
Brian Williams, PhD
Iain Crombie, PhD, C-stat, FFPHM

1Department of General Practice and Primary Health Care, University of Auckland, NZ
2Department Community Health Services, University of Dundee, Dundee, UK
3Department of Psychiatry, University of Dundee, Dundee, UK
4Department of Social Work, University of Dundee, Dundee, UK

ABSTRACT

PURPOSE Depression is common in primary care. There are no systematic reviews of depression treatment comparing antidepressants with placebo; hence, we do not know whether these medications are effective in primary care.

METHODS We searched the Cochrane Collaboration Depression, Anxiety and Neurosis Group register of controlled trials, MEDLINE, International Pharmaceutical abstracts, PsycINFO, and EMBASE. Abstracts of potential studies were reviewed independently by 2 authors. Studies needed to include randomized controlled trials of either a tricyclic antidepressant (TCA) or selective serotonin reuptake inhibitor (SSRI), or both, and placebo in a primary care setting. The data and quality of the studies were extracted and assessed by 2 authors blind to the other’s choice. Disagreements were resolved by discussion. The main outcome measures were the standardized mean difference and weighted mean difference of the final mean depression scores, the relative risk of improvement, and the number withdrawing because of side effects. Pooling of results was done using Review Manager 4.2.2.

RESULTS There were 10 studies in which TCAs were compared with placebo, 3 in which SSRIs were compared with placebo, and 2 with both compared with placebo. One half of the studies were of low methodological quality, and nearly all studies were of short duration, typically 6 to 8 weeks. Pooled estimates of efficacy data showed a relative risk of 1.26 (95% CI, 1.12-1.42) for improvement with TCAs compared with placebo; for SSRIs, relative risk was 1.37 (95% CI, 1.21-1.55). Most patients, 56% to 60%, responded well to active treatment compared with 42% to 47% for placebo. The number needed to treat for TCAs was about 4, and for SSRIs it was 6. The numbers needed to harm (for withdrawal caused by side effects) ranged from 5 to 11 for TCAs and 21 to 94 for SSRIs. Low-dose (100 mg or 75 mg) as well as high-dose TCAs were effective.

CONCLUSION This systematic review is the first comparing antidepressants with placebo for treatment of depression in primary care. Both TCAs and SSRIs are effective. This review is also the first to show that low-dose TCAs are effective in primary care. Prescribing antidepressants in primary care is a more effective clinical activity than prescribing placebo.

INTRODUCTION

It is a paradox that whereas the great majority of patients with clinical depression are cared for by primary care physicians, most research findings upon which decisions are made have involved secondary care patients. This discrepancy is important because research suggests that patients with depressive disorders in primary care have different causes, abnormalities, and natural history than those of psychiatric inpatients or
management of depression. Up to 40% of depressed patients fail to show a response to first-line antidepressant drug treatment, and of those that do respond, only a proportion will achieve full recovery. One cohort study of primary care patients found that 60% of depressed patients treated with medication and 50% with milder depression still met the criteria for caseness at 1 year.

Recent reports have indicated an urgent need to review the evidence of only those studies of antidepressant efficacy on patient samples based in primary care. Several completed reviews and protocols are currently registered on the Cochrane database that consider the efficacy of antidepressants for the treatment of depression. None has specified a focus on patients treated in primary care. Systematic reviews of antidepressant medication often include patients seen in outpatient facilities rather than in primary care or at least recruited from primary care. Concern has been expressed about the relevance of secondary care studies to primary care patients. We are aware of only 2 published systematic reviews based on patients either seen or recruited in primary care. Both compared newer antidepressants with older antidepressants. The Mulrow et al review had a small section on antidepressant drugs vs placebo but reviewed only 4 studies. The MacGillivray et al review compared selective serotonin reuptake inhibitors (SSRIs) with tricyclic antidepressants (TCAs) and therefore comment only on relative efficacy. Comparison with placebo is needed to obtain absolute efficacy. These considerations indicate a persistent need to review the evidence of only those studies that have been conducted comparing antidepressant efficacy with placebo using patient samples based in primary care. The aim of this review was to review systematically the efficacy of antidepressant compared with placebo studies in treating depression in primary care.

**METHODS**

**Study Selection**

Studies were considered for this review if they were randomized, placebo-controlled trials using TCAs and/ or SSRIs and included primary care patients. Primary care patient samples were defined as an undifferentiated group of patients who are able to access medical care from a primary health care clinician. Participants had to be adults who had a diagnosis of depression (studies with predominately children or the elderly were excluded), because we wished to assess the dosage levels in adults, and lower doses of antidepressants are often used in the elderly. Studies could be in any language.

Our primary outcomes were the efficacy of TCAs and SSRIs in comparison with placebo. Secondary outcomes were the number reporting drug-related adverse events and the number withdrawing because of drug-related adverse events. We also explored the effect of study quality on outcome, the effect of having primary care physicians as at least one half of the clinicians prescribing medications and assessing patients, and the effectiveness of low-dose of medication.

**Data Sources**

We electronically searched the Cochrane Collaboration Depression, Anxiety and Neurosis Group (CCDAN) register, MEDLINE, International Pharmaceutical abstracts, PsycINFO, and EMBASE up to February 2003 for any trials in which antidepressants were used in primary care. A follow-up search was done in September 2003, and a final search was done in December 2004. The search terms used were the names of all known antidepressant medications. There was no language restriction. Two reviewers independently assessed abstracts of all studies possible for inclusion for relevant study inclusion criteria. We searched for additional trials in the reference lists of initial studies identified and by scrutinizing other relevant review articles. We also contacted selected authors and experts.

**Data Extraction**

The methodological quality of the selected studies was assessed according to the recommendations of the Cochrane Collaboration Handbook. The components of quality were adequacy of sample size, allocation concealment, clear description of treatment, representative source of subjects, use of diagnostic criteria or clear specification of inclusion criteria, and outcome measures described clearly or use of validated instruments. A score of 0 on any component caused the study to be rated as poor quality. Two reviewers independently extracted data using data extraction forms, and disagreements were resolved by discussion. A similar process was used for the validity assessment.

**Data Synthesis**

All data were analyzed using Review Manager (RevMan) 4.2.2, which is the Cochrane Collaboration software used for preparing systematic reviews. For continuous outcomes we calculated the standardized
mean difference. Where the same outcome scale was used, we calculated the weighted mean difference. For dichotomous outcomes we calculated relative risk (RR) and the range of numbers needed to treat for statistically significant studies. We assessed heterogeneity using the Q statistic.\textsuperscript{15} Where data were available in graphic format only, an approximation of the mean was made to assess the outcomes. For data reported without standard deviations, the highest standard deviations in the outcome scores from the other studies were used. We also performed a funnel plot analysis to check for publication bias. There were a number of definitions for outcomes we described as “improvement”: 4 definitions used ≤50% reduction in the Montgomery-Asberg Depression Rating Scale (MADRS),\textsuperscript{16-19} ≤50% reduction in Hamilton Depression Rating Scale (HAMD),\textsuperscript{20} ≤7 on the HAMD scale,\textsuperscript{21} and ≤4 points on HAMD\textsuperscript{22}; and 3 definitions used global evaluation of improvement.\textsuperscript{23-25}

**RESULTS**

Of the 284 articles identified from the initial search strategy, only 12 met the study criteria (Figure 1). Three additional studies were found in a search undertaken in September 2003. No further studies were found in December 2004. There were 890 participants in SSRI studies, 596 in TCA studies, and 1,267 patients on placebo (Table 1).\textsuperscript{16-30} Of the 5 possible SSRIs available, 2 studied sertraline, 3 studied escitalopram (a precursor of citalopram), and 1 studied citalopram. Of the TCAs available, 2 studied dothiepin, 4 studied amitriptyline, 2 studied mianserin, and 3 studied imipramine. Ten of the 15 studies were identified as having a competing interest.

Our results confirm that both TCA and SSRI are significantly effective compared with placebo (Figures 2 and 3). For depression scores the standardized mean difference for TCA vs placebo was -0.42 (95% confidence interval [CI], -0.55 to -0.3). The relative risk for improvement using TCA medications was 1.26 (95% CI, 1.12 to 1.42). For SSRI medications the relative risk for improvement was 1.37 (95% CI, 1.21 to 1.55). The number needed to treat for 1 improved patient ranged from 3 to 4 for the TCA studies that were statistically significant. Likewise, the number needed to treat was 6 for SSRIs. We performed an analysis with 5 studies (not shown) that had treatment group scores of <8 on the HAMD. The weighted mean difference was -3.68 (95% CI, -5.89 to -1.47). There was no significant heterogeneity for any analyses, so a fixed effects analysis was used. No significant differences were found for those studies in which means were approximated from graphs or standard errors were assumed from other studies compared with studies that had published data. A funnel plot of the TCA studies suggested that small studies with a small effect size might be missing (the funnel plot is not shown). The funnel plot methodology gives a qualitative view of publication bias but not a quantitative perspective and is therefore difficult to interpret.

The relative risk for adverse effects leading to study withdrawal for TCAs was 2.35 (95% CI, 1.59 to 3.46) (Figure 4) and for SSRIs the relative risk was 2.01 (95% CI, 1.1 to 3.7) (Figure 5). The number needed to harm in terms of study withdrawal resulting from adverse effects for 2 statistically significant TCA studies was 5 and 10. None of the 4 SSRI studies had statistically significant findings for adverse effects leading to withdrawal, but using the pooled figure and the range of baseline risks, the number needed to harm ranged from 21 to 94.

Seven studies did not meet the minimum quality criteria on at least 1 of the key components of methodological quality.\textsuperscript{18,22,35,27-29} A score of 0 on any component caused the study to be rated as being poor quality.
Only 4 studies used an intention-to-treat analysis, and these studies were the most recent. When studies of low methodological quality for the TCAs (n = 6) were removed from analysis, the pooled standardized mean difference or depression score for TCA vs placebo was -0.50 (95% CI, -0.65 to -0.35). For improvement for the TCAs the relative risk was 1.34 (95% CI, 1.16 to 1.55). When studies in which at least one half of its assessors were family practitioners were pooled, the standardized mean difference was -0.43 (95% CI, -0.58 to -0.28) and the relative risk was 1.2 (95% CI, 1.03 to 1.4). There were sufficient data to assess continuous outcomes for TCAs at 1 week, 2 weeks, and 4 weeks. The standardized mean difference at 1 week was -0.02 (95% CI, -0.17 to 0.13), at 2 weeks it was -0.2 (95% CI, -0.36 to -0.04), and at 4 weeks it was -0.34 (95% CI, -0.5 to -0.18). For studies that used a HAMD <8 as an outcome (considered to be a remission) the weighted mean difference was -3.68 (95% CI, -5.89 to -1.47). For the 3 studies that reported no conflict of interest, the weighted mean difference was -4.59 (95% CI, -6.82 to 2.36).

Ten studies included an arm with 100 mg or more of a tricyclic antidepressant or more than 60 mg of mianserin. For the 10 studies in which a high dose was given, the standardized mean difference was -0.42 (95% CI, -0.56 to -0.29). The relative risk for these studies was 1.32 (95% CI, 1.15 to 1.5). For the 4 studies of tricyclic antidepressants using a dose of 100 mg/d or less, the weighted mean difference (all used the HAMD) was -3.15 (95% CI, -5.05 to -1.24). For the 2 studies of tricyclic antidepressants using a dose of 75 mg/d, the weighted mean difference was -3.93 (95% CI, -7.65 to -0.21).

Most studies had heterogeneous diagnoses in their participants. Only 2 TCA studies had major depressive disorder as the single diagnosis, and the weighted mean difference for that study was -1.37 (95% CI, -2.52 to -0.22). For the SSRI studies there were 4 studies...

Table 1. Features of 15 Randomized Trials Comparing Either a Selective Serotonin Reuptake Inhibitor (SSRI) or Tricyclic Antidepressant (TCA) With Placebo

<table>
<thead>
<tr>
<th>Study</th>
<th>Quality*</th>
<th>Diagnosis</th>
<th>TCA Dose†</th>
<th>Responsible for Treatment</th>
<th>Competing Interest</th>
<th>Study Period</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCA vs placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blashki et al26</td>
<td>High</td>
<td>Heterogeneous</td>
<td>High and low</td>
<td>GP and psychiatrist</td>
<td>No</td>
<td>4 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>Brink et al27</td>
<td>Low</td>
<td>Heterogeneous</td>
<td>High</td>
<td>GP</td>
<td>Yes</td>
<td>6 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>Doogan &amp; Langdon28</td>
<td>High</td>
<td>MDD</td>
<td>High</td>
<td>GP</td>
<td>Yes</td>
<td>6 wk</td>
<td>MADRS 50%M</td>
</tr>
<tr>
<td>Feighner et al29</td>
<td>High</td>
<td>Heterogeneous</td>
<td>High</td>
<td>Psychiatrist</td>
<td>No</td>
<td>4 wk</td>
<td>HAMD 50%H</td>
</tr>
<tr>
<td>Hollyman et al30</td>
<td>High</td>
<td>Heterogeneous</td>
<td>High</td>
<td>Psychiatrist</td>
<td>No</td>
<td>6 wk</td>
<td>HAMD 50%H</td>
</tr>
<tr>
<td>Lerubier et al31</td>
<td>Low</td>
<td>Heterogeneous</td>
<td>High</td>
<td>Psychiatrist</td>
<td>Yes</td>
<td>12 wk</td>
<td>HAMD 50%H</td>
</tr>
<tr>
<td>Malt et al32</td>
<td>High</td>
<td>Heterogeneous</td>
<td>High</td>
<td>GP</td>
<td>Yes</td>
<td>24 wk</td>
<td>MADRS 50%M</td>
</tr>
<tr>
<td>Mynors-Wallis et al33</td>
<td>High</td>
<td>Heterogeneous</td>
<td>High</td>
<td>GP</td>
<td>No</td>
<td>12 wk</td>
<td>MADRS</td>
</tr>
<tr>
<td>Philipp et al34</td>
<td>Low</td>
<td>ModDD</td>
<td>Low</td>
<td>GP</td>
<td>Yes</td>
<td>8 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>Thompson &amp; Thompson35</td>
<td>Low</td>
<td>Heterogeneous</td>
<td>Low</td>
<td>GP</td>
<td>No</td>
<td>4 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>Thomson et al36</td>
<td>Low</td>
<td>Heterogeneous</td>
<td>High</td>
<td>GP</td>
<td>Yes</td>
<td>12 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>Barge-Schaapveld &amp; Nicholson37</td>
<td>Low</td>
<td>MDD</td>
<td>High</td>
<td>Unclear</td>
<td>Yes</td>
<td>6 wk</td>
<td>HAMD</td>
</tr>
<tr>
<td>SSRI vs placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doogan &amp; Langdon38</td>
<td>High</td>
<td>MDD</td>
<td>100 mg sertraline</td>
<td>GP</td>
<td>Yes</td>
<td>6 wk</td>
<td>MADRS 50%M</td>
</tr>
<tr>
<td>Lepola et al39</td>
<td>Low</td>
<td>MDD</td>
<td>10 mg escitalopram</td>
<td>Unclear</td>
<td>Yes</td>
<td>8 wk</td>
<td>MADRS 50%M</td>
</tr>
<tr>
<td>Malt et al32</td>
<td>High</td>
<td>Heterogeneous</td>
<td>100 mg sertraline</td>
<td>GP</td>
<td>Yes</td>
<td>24 wk</td>
<td>MADRS</td>
</tr>
<tr>
<td>Montgomery et al40</td>
<td>High</td>
<td>MDD</td>
<td>10 mg escitalopram or 20 mg citalopram</td>
<td>Unclear</td>
<td>Yes</td>
<td>4 wk</td>
<td>MADRS</td>
</tr>
<tr>
<td>Wade et al41</td>
<td>High</td>
<td>MDD</td>
<td>10 mg escitalopram</td>
<td>Unclear</td>
<td>Yes</td>
<td>8 wk</td>
<td>MADRS</td>
</tr>
</tbody>
</table>

HAMD = Hamilton depression scale; 50%H = discrete outcomes where improvement is a greater than 50% reduction in HAMD; MADRS = Montgomery-Asberg Depression Rating Scale; 50%M = discrete outcomes where improvement is a greater than 50% reduction in the MADRS; heterogeneous = patients thought by their general practitioner to be depressed, which may or may not include patients with major depression as opposed to the studies with only patients with major depression; MDD = major depressive disorder; ModDD = moderate depressive disorder.

* Quality high if adequate sample size, concealment, description of treatment, representative sample, specified inclusion, details of withdrawals, valid outcomes.
† High dose defined as majority of TCA treated patients receiving at least equivalent of 100 mg/d amitriptyline (60 mg mianserin).
‡ Study has 3 arms (SSRI vs TCA vs placebo).
in which all participants had major depressive disor-
der.16,18,19,30 Only 3 of the 4 studies had data suitable for pooling, and the relative risk was 1.39 (95% CI, 1.21 to 1.61).

**DISCUSSION**

We believe our systematic review is the first to compare the efficacy of antidepressants with placebo in the primary care setting. Our results confirm that both TCAs and SSRIs are significantly more effective than placebo for discrete and continuous outcomes. The results seem to apply to major depressive disorder and heterogeneous depression (commonly seen in primary care) and suggest that treating depression with antidepressants is an appropriate activity in primary care. We can only speculate as to the composition of the heterogeneous group, as they had a wide range of definitions but did consider patients with levels of depression less than major depression. Only 1 study analyzed major and minor depression and found no effect for minor depression.23

We found only 15 studies based in primary care that met inclusion criteria and provided evidence for the comparative efficacy of TCAs and SSRIs vs placebo. We similarly found relatively few studies in a previous review of trials comparing SSRIs with TCAs in primary care,13 as opposed to the considerably larger number of studies conducted with patients from all settings. Williams et al31 found 206 studies comparing a newer with an older antidepressant (123 of which involved an SSRI). They found a benefit for the newer antidepressants (RR = 1.6, 95% CI, 1.3 to 2.3). Steffens et al, in a US-only based review, found 36 trials comparing a tricyclic with an SSRI.32 Most studies included in our
current review were small, phase 3 studies supported by commercial funding (Table 1). In fact, all the SSRI vs placebo studies had some commercial involvement. Many studies reviewed were of low methodological quality, and nearly all studies were of short duration, typically 6 to 8 weeks. Our findings are in keeping with a review of 108 studies of newer antidepressants that found both TCAs and SSRIs to be effective in treating depression.33 Previous reviews have tended to show that SSRIs are generally more tolerable than TCAs, but evidence is conflicting. Meta-analyses using dropout rates as an index of tolerability have varied findings. While one review14 found no difference in dropout rates between SSRIs (32.3%) and TCAs (33.2%), another13 found a small but statistically significant lower dropout rate for SSRIs (30.8%) relative to TCAs (33.4%). In our review focusing only on primary care samples, we found dropout rates for SSRIs of 5.4% and TCAs of 12%. The numbers needed to harm for the withdrawals from the statistically significant TCA studies ranged from 5 to 11. In another review of antidepressants in primary care, the relative risk of withdrawal of patients resulting from side effects from SSRIs compared with TCAs was 0.6 (95% CI, 0.6 to 0.88).13 The National Institute of Clinical Evidence (UK NICE) review group of antidepressants considered a weighted or a standardized mean difference of 3 or more to be clinically significant (D Goldberg, personal communication, April 30, 2004).
We have reported significant numbers needed to treat, which are another way of assessing clinical significance. Primary care clinicians may be more likely than hospital colleagues to alter therapy when side-effects are experienced, even during clinical trials.66

Most systematic reviews concerning efficacy of antidepressant medications fail to report a detailed examination of methodological quality and therefore fail to include such criteria when examining treatment effects. Bias in primary studies that is due to poor methodological quality (e.g., selection bias, ascertainment bias, inappropriate handling of withdrawals, protocol violations) can lead to exaggeration of treatment effects. A study of trial quality in systematic reviews showed that if low-quality studies were included in pooled estimates of treatment effect, there was a 30% to 50% exaggeration of treatment effectiveness.37 We did not, however, find any appreciable differences between effects for the high-quality studies compared with the lower quality studies. Another form of bias with meta-analysis is that of publication bias. Our funnel plot suggests that small studies with small effect sizes may be missing. This finding is consistent with a review of all applications to the US Food and Drug Administration, which, on examination of all submitted trials of newer antidepressant medications, found that the benefit of antidepressant medications was much smaller when all studies were considered than when only the published studies were considered.38 The use of standardized mean difference was necessary, as the studies with continuous outcomes used a number of different scales—HAM-D in the older studies and MADRS in more recent years. It would be helpful for future meta-analyses for both the HAM-D and MADRS to be used to facilitate pooling.

Our finding of a significant benefit when pooling the results of 2 studies is consistent with a recent meta-analysis of studies in all settings, which found a benefit from treatment with low-dose tricyclic antidepressants.39 Neither of the findings from the 2 studies in our review were statistically significant, which suggests that larger trials are needed in the primary care setting to clarify such issues as dose of antidepressant medication.36,28 The review of studies of low-dose medications found that no evidence of increased benefit, but there was an increase in side effects with higher dosages. Our results were similar to those of that review, but we did not find a significant increase in adverse effects.

Most of the trials reviewed here studied patients with a range of depression severity. Only 2 TCA studies included patients with major depressive disorder.19,29 Three studies of SSRIs included only patients with major depressive disorder.66,18,19 As patients in primary care settings have a range of depression severity, the generalizability of the results of these studies to primary care is reasonable.40 Advice on using TCAs has stressed that patients will not obtain a benefit from medication until 2 weeks of treatment has passed. Our findings are in agreement. Also at issue is that primary care populations may benefit from antidepressant medication only when it is given by a psychiatrist. Our significant findings for continuous and discrete outcomes contradict this concern.

We found evidence that both TCAs and SSRIs are more effective than placebo. This finding needs to be tempered with the knowledge that some publication bias may have occurred and that many studies in the review were small and of variable quality. Gaps in the literature include a lack of attention to the treatment of specific diagnostic groups, in particular patients with minor depression. Further research is needed on these groups of patients in addition to longer and larger trials of low-dose TCAs. In terms of practice, many guidelines are recommending SSRIs rather than TCAs because of safety. Both are effective, and if safety is not an issue, then individual tolerability to side effects will determine types of medications used.

Funding support: Funding Body Chief Scientist Office, Scotland.

References


Metformin as Treatment for Overweight and Obese Adults: A Systematic Review

Kara M. Levri, MD, MPH
Elizabeth Slaymaker, MD
Allen Last, MD
Julie Yeh, MD
Jonathan Ference, PharmD
Frank D’Amico, PhD
Stephen A. Wilson, MD
1University of Pittsburgh Medical Center
St. Margaret, Pittsburgh, Pa
2Grace Medical Associates, Pittsburgh, Pa
3University of Oklahoma-Tulsa, Tulsa, Okla

ABSTRACT

PURPOSE We wanted to determine whether metformin is an effective medication for treatment of overweight or obese adults who do not have diabetes mellitus or polycystic ovary syndrome (PCOS).

METHODS We searched MEDLINE (1966–2003), EMBASE (1986–2003), Allied and Complementary Medicine Database (1985-2003), International Pharmacetical Abstracts (1970-2003), the Cochrane Library, American College of Physicians Journal Club, Database of Abstracts of Reviews of Effects, Cochrane Controlled Trials Register, MEDLINE In-Process & Other Non-Indexed Citations, reference lists of retrieved articles, and articles by selected authors and pharmaceutical manufacturers. Inclusion criteria were being obese or overweight determined by a BMI of 25 kg/m2 or greater or waist-to-hip ratio (WHR) of more than 0.8, metformin use, and aged 18 years or older. Exclusion criteria were a diagnosis of diabetes mellitus, polycystic ovarian syndrome or descriptors of polycystic ovarian syndrome, human immunodeficiency virus infection, and concomitant antipsychotic medications. Trials were graded on an 11-point Jadad scale. Only randomized controlled and blinded trials were accepted. Two reviewers independently extracted data from each trial. Primary outcomes measured were changes in BMI, WHR, and weight.

RESULTS Fifty-seven potentially relevant studies were initially identified; 48 were excluded because of lack of randomization, lack of blinding, failure to meet inclusion or exclusion criteria, inaccessible outcomes, or improper study design. Nine clinical trials met criteria for validity assessment. Four studies used the parameter of waist-to-hip ratio, 3 studies included BMI, and 8 used weight. Two of the 9 studies showed a small reduction in WHR.

CONCLUSION Insufficient evidence exists for the use of metformin as treatment of overweight or obese adults who do not have diabetes mellitus or polycystic ovary syndrome. Further studies are needed to answer this clinical question.


INTRODUCTION

Obesity has become a major public health concern in the United States. More than 60% of adults are overweight or obese,1 with a body mass index (BMI) greater than 25 kg/m2.2 Obesity affects persons regardless of age, race, ethnicity, and sex.2 It is associated annually with 400,000 deaths, and associated health care costs exceed $177 billion.3

A combination of genetic, metabolic, behavioral, environmental, cultural, and socioeconomic factors contribute to a person’s body weight.2 An elevated BMI increases the prevalence, morbidity, and mortality of type 2 diabetes mellitus, hypertension, heart disease, stroke, osteoarthritis, respiratory tract disorders, gallstones, certain types of cancer, and psychological disorders.2,3

Many treatment options are available for overweight and obese adults: behavioral strategies, medications approved by the US Food and Drug Administration (FDA), and bariatric surgery for those at greatest risk. Several FDA-approved drugs for conditions other than obesity have been...
investigated as treatment of excess body weight. Metformin is one such drug.

Metformin, the biguanide most widely used for the treatment of type 2 diabetes mellitus, may be useful in aiding weight loss. In diabetic patients, it suppresses endogenous glucose production and may also act as an insulin sensitizer. It also helps diabetic patients lose weight or at least keep their weight stable. In addition to its use in treatment of diabetes, metformin has also become commonly prescribed for patients with polycystic ovary syndrome (PCOS), and its use has resulted in weight reduction in those patients as well.

Given the problem of increasing rates of obesity, we conducted a systematic review of the literature regarding metformin and weight loss in patients who did not have diabetes or PCOS. To our knowledge, no other systematic review has been performed on the use of metformin for weight loss in this patient population.

METHODS

MEDLINE (1966–2003), EMBASE (1986–2003), Allied and Complementary Medicine Database (AMED) (1985-2003), International Pharmaceutical Abstracts (1970-2003), the Cochrane Library, American College of Physicians (ACP) Journal Club, Database of Abstracts of Reviews of Effects (DARE), Cochrane Controlled Trials Register (CCTR), and MEDLINE In-Process & Other Non-Indexed Citations were searched electronically. The search was limited to publications in English. Search terms were “metformin” or “Glucophage” or “biguanides” or “diguanide” and “obese/obese therapy” or “weight/weight loss” or “body fat” or “diet/diet therapy” or “overweight” or “fitness,” using appropriate suffixes and derivatives. Reference lists of all retrieved articles were examined. Attempts were made to contact authors who had expertise regarding the clinical question (ie, multiple publications on the issue), as well as the pharmaceutical manufacturer of Glucophage, Bristol-Myers Squibb, asking for additional data, unpublished studies, or missed references.

Inclusion criteria were predefined as follows: obese or overweight determined by a BMI of 25 kg/m² or greater or waist-to-hip ratio (WHR) of more than 0.8, metformin use, and aged 18 years or older. Exclusion criteria were predefined as a diagnosis of diabetes mellitus, PCOS, or descriptors of PCOS, such as hirsutism or oligomenorrhea; infection with the human immunodeficiency virus (HIV); and concomitant antipsychotic medication use.

Articles were screened for inclusion and exclusion criteria by multiple reviewers. Authors of studies with the potential to meet inclusion or exclusion criteria by additional unpublished information were contacted. Each trial was scored independently by 2 researchers. Trials were graded on an 11-point Jadad scale. Only randomized controlled and blinded trials were accepted; no studies with a Jadad score of less than 7 were included in analysis. Studies lacking pretreatment and posttreatment weight, BMI or WHR were excluded. Disagreements were settled by group consensus.

Two reviewers independently extracted data from each trial. Primary outcomes measured were changes in BMI, WHR, and weight. All outcomes were assessed at the maximum follow-up time or the end of treatment.

No formal statistical procedures or tests were performed because there was not sufficient similar information from each article to achieve a meaningful meta-analysis.

RESULTS

Fifty-seven potentially relevant studies were initially identified. Of these, 49 were identified through a MEDLINE search, 1 from MEDLINE In-Process & Other Non-Indexed Citations; 7 additional studies were found through a reference search of those studies. Of the studies with the potential to meet inclusion-exclusion criteria by additional unpublished information, response was achieved from 2 of the 7 authors contacted. One author was unable to provide this information because of the ongoing nature of the study. The second author was “unable to provide the time to review the sought information.” The remaining 5 authors were unable to be contacted after 2 attempts. These 7 potential studies for review were among the 10 that were excluded after a more detailed evaluation and were incorporated into the category of “inaccessible outcomes” (Figure 1).

A total of 48 studies were excluded because of lack of randomization, lack of blinding, failure to meet inclusion or exclusion criteria, inaccessible outcomes, and improper study design. As a result, 9 clinical trials, published between 1970 and 2002, met criteria for validity assessment (Figure 1).

Clinical features included study durations ranging from 15 days to 1 year and metformin dosages ranged from 750 to 1700 mg/d. Six of the 9 studies included both male and female participants. Four studies used the parameter of waist-to-hip ratio, 3 studies included BMI, and 8 used weight (Table 1).

Table 2 displays quality assessment features of the 9 clinical trials. One study specifically reported allocation concealment (Kantola et al¹⁶). Three studies reported intention-to-treat analysis, 3 did not utilize intention-to-treat, and 3 were not applicable for this analysis because of a lack of study dropouts.

Table 3 summarizes the quantitative results of each
METFORMIN FOR OVERWEIGHT AND OBESITY

of the 9 clinical trials. Two studies showed statistically significant improvements. Fontbonne et al found significant reduction in weight and waist-to-hip ratio; Paolisso et al showed a significant decrease in weight.

DISCUSSION

Based on this review, the evidence for using metformin as treatment of overweight or obesity in adults who had no diabetes mellitus or polycystic ovary syndrome was insufficient. We reviewed articles relating to our clinical question in the absence of any other systematic review on this topic. We found 9 high-quality articles in which patients were randomized into 2 or more groups, baseline measures were taken, and subsequent measures taken at follow-up time points. Our study has both strengths and weaknesses, and further study is needed for definitive answers.

This systematic review is methodologically sound. Reviewers evaluated double-blind randomized controlled trials using doses of metformin that resulted in effects on glucose metabolism. Beyond electronic databases, we sought information from expert authors, manufacturers, unpublished studies, and published trials with unpublished data. Two reviewers independently assessed each article for inclusion criteria and validity. Our failure to blind reviewers to author and journal of publication is an area that could have introduced bias; however, the reviewers did not know the authors of articles published in the journals.

The primary limitation of this systematic review is that 5 of the 9 studies were not specifically designed to study weight loss, change in BMI, or change in WHR as their primary outcome measure, which might have affected the ability of studies to detect a change in weight, BMI, or WHR, because the studies were powered for a different outcome (affecting sample size and treatment dose). Another concern is the short duration of treatment (15 days to 3 months) in 6 of the 9 studies. Furthermore, the BMI measurements of study participants were limited to overweight (BMI 25-29.9 kg/m²) or grade I obesity (BMI 30-34.9 kg/m²). Indeed, metformin might enhance weight loss more effectively in obese patients whose BMI is greater than 35 kg/m², because that is when insulin resistance, a potential mediator of weight gain and inhibitor of weight loss, becomes more prevalent. A randomized controlled trial designed with adequate power for the outcomes of weight loss, BMI, or WHR that compared metformin with placebo in patients with and without diabetes or polycystic ovary syndrome, stratifying for BMI measurements of 25 to 35, 35 to 40, and greater than 40 kg/m², and observed patients for at least 1 year would provide a more definitive answer to what role, if any, metformin has in assisting weight loss.

For reasons cited in the methods, a meta-analysis of the individual studies could not be performed. Various studies gave the results in terms of differences from baseline, some reported the raw means and standard deviations at each time point, others reported the percentage of change, and a few gave their results in graphical form. Three or 4 of the studies had data that could be used for meta-analysis, but the authors collectively agreed that because only 9 studies were included in the review, a meta-analysis on such a small subset of the studies was not suitable. We do not consider this lack of meta-analysis to be a limitation of the review. All of these studies had similar designs, yet the reporting of the findings is so dissimilar, they do not allow a final quantitative answer to be obtained. Researchers
### Table 1. Clinical Features of Studies Included for Review

<table>
<thead>
<tr>
<th>Source</th>
<th>Subjects</th>
<th>Dosage (mg)</th>
<th>Study Duration</th>
<th>Main Outcome Measure</th>
<th>Treatment Arms</th>
<th>Baseline Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles et al,12 2000</td>
<td>168 men</td>
<td>850 bid</td>
<td>3 mo</td>
<td>Lipids and blood pressure</td>
<td>2</td>
<td>WHR ≥ 0.95</td>
</tr>
<tr>
<td>Charles et al,13 1998</td>
<td>457 persons</td>
<td>850 bid</td>
<td>1 y</td>
<td>Fibrinolysis</td>
<td>2</td>
<td>WHR ≥ 0.95 men</td>
</tr>
<tr>
<td>Fontbonne et al,14 1996</td>
<td>108 men</td>
<td>850 bid</td>
<td>1 y</td>
<td>Metabolic parameters and fat distribution</td>
<td>2</td>
<td>WHR ≥ 0.95 men</td>
</tr>
<tr>
<td>Guigliano et al,15 1993</td>
<td>12 women</td>
<td>850 bid</td>
<td>12 wk x 2</td>
<td>Blood pressure and left ventricular mass</td>
<td>2</td>
<td>BMI 34 ± 0.9</td>
</tr>
<tr>
<td>Kantola et al,16 2002</td>
<td>15 men</td>
<td>1,500 bid</td>
<td>8 wk</td>
<td>Blood pressure and insulin sensitivity</td>
<td>2</td>
<td>Weight 86.4-87.7 kg</td>
</tr>
<tr>
<td>Lawson et al,17 1970</td>
<td>34 women</td>
<td>250 tid, 500 tid</td>
<td>8 wk x 4</td>
<td>Weight loss</td>
<td>4</td>
<td>Weight 74.5-125.5 kg</td>
</tr>
<tr>
<td>Lehtovorta et al,18 2001</td>
<td>20 men</td>
<td>500 bid</td>
<td>6 mo</td>
<td>Metabolic parameters</td>
<td>2</td>
<td>Weight 87.8-91.0 kg</td>
</tr>
<tr>
<td>More et al,19 1999</td>
<td>11 women</td>
<td>850 bid</td>
<td>6 wk x 2</td>
<td>Glucose disposal</td>
<td>2</td>
<td>BMI 30-48.5 kg</td>
</tr>
<tr>
<td>Paolisso et al,20 1998</td>
<td>30 persons</td>
<td>500 bid, 850 bid</td>
<td>15 d</td>
<td>Food intake</td>
<td>2</td>
<td>Weight 79.0-159.1 kg</td>
</tr>
</tbody>
</table>

**bid = twice a day; WHR = waist-to-hip ratio; BMI = body mass index; tid = 3 times a day.**

### Table 2. Quality Features of Articles Included for Review

<table>
<thead>
<tr>
<th>Source</th>
<th>Randomized</th>
<th>Double-Blinded</th>
<th>Description of Dropouts</th>
<th>Design</th>
<th>Description of Allocation Concealment</th>
<th>Power Analysis</th>
<th>Intention to Treat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles et al,12 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Indeterminant</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Charles et al,13 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Indeterminant</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fontbonne et al,14 1996</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Indeterminant</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Guigliano et al,15 1993</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Crossover</td>
<td>Indeterminant</td>
<td>Yes</td>
<td>No dropouts</td>
</tr>
<tr>
<td>Kantola et al,16 2002</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lawson et al,17 1970</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Crossover</td>
<td>Indeterminant</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Lehtovorta et al,18 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Indeterminant</td>
<td>Yes</td>
<td>No dropouts</td>
</tr>
<tr>
<td>More et al,19 1999</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Crossover</td>
<td>Indeterminant</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Paolisso et al,20 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>RCT</td>
<td>Indeterminant</td>
<td>No</td>
<td>No dropouts</td>
</tr>
</tbody>
</table>

**RCT = randomized controlled trial.**

### Table 3. Results of Clinical Trials Included for Review

<table>
<thead>
<tr>
<th>Source</th>
<th>Weight (kg)</th>
<th>BMI (kg/m²)</th>
<th>Waist-Hip Ratio†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Metformin</td>
<td>Placebo</td>
<td>Metformin</td>
</tr>
<tr>
<td>Charles et al,12 2000</td>
<td>-0.50</td>
<td>0.10</td>
<td>—</td>
</tr>
<tr>
<td>Charles et al,13 1998</td>
<td>-2.0</td>
<td>-0.8</td>
<td>—</td>
</tr>
<tr>
<td>Fontbonne et al,14 1996</td>
<td>-2.0</td>
<td>-0.8</td>
<td>—</td>
</tr>
<tr>
<td>Guigliano et al,15 1993</td>
<td>—</td>
<td>0.2</td>
<td>—</td>
</tr>
<tr>
<td>Kantola et al,16 2002</td>
<td>-0.2</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>Lawson et al,17 1970</td>
<td>-3.6</td>
<td>-2.22</td>
<td>—</td>
</tr>
<tr>
<td>Lehtovorta et al,18 2001</td>
<td>-2.6</td>
<td>-1.2</td>
<td>—</td>
</tr>
<tr>
<td>More et al,19 1999</td>
<td>-0.6</td>
<td>0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>Paolisso et al,20 1998</td>
<td>-2.8*</td>
<td>-0.25</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

* Indicates statistically significant (P < .05) difference from baseline.
† A change in ratios cannot be expressed in terms of units.
can improve the quality of randomized controlled trials by following guidelines for reporting data.\textsuperscript{20,23} Evidence is insufficient to conclude that metformin can serve as a treatment of overweight or grade I obesity in adults who do not have diabetes mellitus or polycystic ovary syndrome. At this time, the use of metformin is not recommended as treatment of these conditions. If clinicians and patients are considering pharmacologic assistance for weight loss, the National Heart, Lung, and Blood Institute and Cochrane Reviews\textsuperscript{24,25} recommend sibutramine or orlistat as therapeutic options for patients who have a BMI of greater than 30 and who have no concomitant risk factors and for patients who have a BMI greater than 27 and who have concomitant risk factors; they recommend bariatric surgery as a therapeutic option for patients who have a BMI greater than 40.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/457.

Key words: Metformin; obesity; anti-obesity agents; body mass index

Submitted October 4, 2004; submitted, revised, April 19, 2005; accepted April 28, 2005.

A version of this report has been presented as "Metformin as Treatment for Overweight or Obese Adults: a Systematic Review," as a research in progress poster presentation at the 37th Society of Teachers of Family Medicine Annual Spring Conference, Toronto, Ontario, May 2004; as "Metformin as Treatment for Overweight or Obese Adults: a Systematic Review," at the Pennsylvania Academy of Family Physicians Research Day, Philadelphia, Pa, April 2004; as "Metformin as Treatment for Overweight or Obese Adults: a Systematic Review," as a UPMC St. Margaret Fellowship Research Day presentation, UPMC St. Margaret, Pittsburgh, Pa, June 2004; and as "Metformin as Treatment for Overweight or Obese Adults: a Systematic Review," as a research in progress poster presentation at the 38th Society of Teachers of Family Medicine Annual Spring Conference, New Orleans, La, May 2005.

Funding support: University of Pittsburgh Medical Center St. Margaret, Pittsburgh, Pa.

References


The Decade Dance

Jon O. Neher, MD
Valley Family Medicine, Renton, Wash

ABSTRACT
This essay portrays the moment an adult survivor of childhood sexual abuse tells his physician about his traumatic past. Because of denial by both parties, the diagnosis had remained buried for years, and opportunities for appropriate intervention were repeatedly missed. When the patient is finally able to face his past, it unleashes a torrent of potentially lethal anger that challenges everyone involved.


Mr. Kelley (not his real name) sits on the examination table, a collection of tics and spasms, refusing to meet my gaze. His obvious anxiety suggests that he is building up to some important disclosure. “You know,” he finally says, looking out the window, “I’ve never told anybody about this.…”. He clears his throat. “I was sexually abused as a kid.”

I stare at him blankly a moment, his statement not registering. Then slowly, the enormity of what he just said starts to sink in. “I’m sorry to hear that,” I tell him, knowing the words are inadequate. His revelation, however, instantly sheds a chilling light on his many and chronic peculiarities.

Ten years ago, I had welcomed him into my practice and took my first history and performed my first physical examination on him. At the time, he was in his early forties, with thinning red hair but a boyish, freckled face. Throughout his initial visit (and ever since), he avoided making eye contact and spoke only in short, nervous sentences. I noted in the chart that he had never married, and he reported no close friends. Although I was concerned about his social isolation, he seemed so acutely ill at ease that I thought it might actually be unkind to press for more details. I promised myself I would ask about it later, but never did.

Eight years ago, I found a small melanoma on his back. Mr. Kelley had accepted the news as if it was no consequence, as if the possibility of disfigurement or death was a mere annoyance. Fortunately, he did well medically. Mr. Kelley, I now understood, had already faced far greater challenges than skin cancer.

Four years ago, Mr. Kelley developed a rectal prolapse that required surgical repair. I asked him if he had any thoughts about why the condition might have developed. Mr. Kelley said only that he had a problem with chronic constipation. That explanation seemed to be enough for the surgeon. I don’t think the surgeon or I even considered the possibility of perineal injury.

Back in the present, Mr. Kelley tugs at his collar and coughs. “I was fired from my job, too.”

“When was that?” I ask.

“Last week.”

Two years ago, Mr. Kelley had been injured at his job in a warehouse when he slipped on a wet floor and hurt his back. Although lumbar imaging showed only some arthritis and minor disk disease, his pain incapacitated him. The episode also unleashed from him a torrent of anger. As I filled out innumerable work capacity forms, white-hot rage poured out of...
him at my office. It poured out of him at the job site too, until everyone there became just a little afraid of him, marking him as the type of man who might just "go postal."

Part of his anger was directed at his injury and his work situation. But I knew the man well enough by now to suspect that, at its root, this rage went far deeper. I suggested that he see a counselor for anger management. Initially reluctant, he eventually followed through on the suggestion. Those anger management sessions must have been effective. He now knew with absolute certainty where his anger was coming from.

"The abuse went on for years, you know," he adds. A sardonic grin flashes across his face. "I tracked him down on the Internet—found out that he's still alive. He's down in Rock Ridge … across the street from a grade school."

His personal darkness seems to close in around us. Mr. Kelley stiffens. "All I want to do now … is go down there and kill that bastard!" he spits.

I suddenly feel lost, not expecting and certainly not prepared for this. "Is that something you are planning to do, Mr. Kelley?" I ask, guessing that having a plan or purchasing a weapon might increase the likelihood of violence. Mr. Kelley shifts his gaze out the window again and does not answer. In a moment of near panic, I believe he might actually carry out his threat.

"You know it's not a good idea," I blurt out, not knowing if confrontation is the right thing to do. "In the long run, it won't make you any happier."

He sits silently for a long moment, his eyes focused somewhere far away. My pulse pounds in my ears as I watch his face intently. Finally, he takes a deep breath, looks down at the floor, and says, "No, I suppose not."

I relax a little, but I am not ready to trust that the situation has been diffused so quickly. "Is it all right with you if I call your anger management therapist and tell her about this?" I ask, extremely thankful that I have solid backup already in place.

Mr. Kelley squirms where he sits on the examination table, as if the thought of anyone else knowing makes him very uncomfortable. Finally, the internal struggle ends. "Sure, I guess. I guess that's okay."

"Good. Let me do that right now."

I step out of the room and a wave of fatigue causes me to sag against the wall. Intense emotions vie for recognition. I feel anger at the abuser, sorrow for Mr. Kelley, apprehension about his dangerous rage, and disbelief that we have been running circles around the central issue of this man's life for as long as I have known him.

I take a few slow, deep breaths and try to regain my composure. Was there any way of avoiding this crisis by getting to the truth sooner? Mr. Kelley's defenses had obviously been strong, but sadly, I realize I never asked any probing questions that might have helped him tell his story.

I push up from the wall, vowing to take more diligent and courageous social and sexual histories from now on. While I may not get a full account right away, I owe it to Mr. Kelley and every other hidden abuse survivor to at least start the conversation.

I need to make a phone call. With the right diagnosis, we can finally start the process of healing.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/3/5/462.

Submitted January 11, 2005; submitted, revised April 26, 2005; accepted May 2, 2005.

Key words: Child abuse, sexual; physician-patient relations; essay
ON TRACK

Strength from Vulnerability

Kurt C. Stange, MD, PhD, Editor


The Annals online discussion since the last issue challenges recent guidelines for hypertension management, informs the use and implementation of electronic health records, draws inspiration from a practice-based research network health behavior change research initiative, and thoughtfully considers the power of sharing vulnerability in clinical practice.

CHALLENGING THE VALUE OF PREHYPERTENSION

The study by Liska and colleagues in the last issue of Annals led many to challenge the JNC 7 guidelines, which the study's data were interpreted to support. Noting that the absolute cardiovascular disease risk for prehypertensive patients is on average quite low, a physician and epidemiologist from the Framingham Heart Study suggest, "Only with global vascular risk assessment is it possible to avoid needlessly alarming or falsely reassuring these prehypertensive patients and subjecting them to therapy they do not need." These writers report the gradual, continuous nature of risk from higher blood pressure, and further note that 80% to 90% of prehypertensive patients in the Framingham Study had at least 1 other cardiovascular risk factor.

A reader from the Center for Medical Consumers raises similar concerns about medicalization of a risk factor that conveys a low absolute risk. She supports the Framingham writers' argument for considering the individual's entire risk profile. She further comments: "Whenever a committee of experts expands the boundaries for who has a disease or condition, I always look for the pharmaceutical industry's influence. It's not hard to find. Strong financial ties to drug companies were found in 9 of the 11 committee members who created prehypertension."

THE ELECTRONIC HEALTH RECORD

Studies from the last issue of Annals showed the challenges of implementing an electronic medical record and found a positive effect of an electronic medical record on the process of diabetes care, but not patient outcomes. These studies stimulated a thoughtful sharing of experience and related research on the inevitability and potential of electronic health records, and the need for supportive implementation strategies that are focused on practice transformation. Furthermore, several authors call for electronic records that move beyond the physician-centric focus which provokes a cacophony of prompts and reminders. They point to the potential for electronic health records that engage other health care team members and patients in community- and population-oriented systems. A further challenge is to provide a "computer simulation-based evidence integrator to calculate the most valuable clinical actions for each individual patient." Together, these commentaries call for more sophisticated systems that provide added value to current care and for supportive and realistic implementation approaches.

PRESCRIPTION FOR HEALTH

Prescription for Health is a collaboration of practice-based research networks to generate new knowledge about how primary care practices can promote health behavior change. In discussion of the early findings published in an Annals supplement, both systems and frontline perspectives are reflected.
From the front lines, a patient reminds us that a little support at the right time from a primary care practice can make a difference. The lay health coach for one of the projects discovered, when working with more than 100 primary care patients, “a sense of longing, a lack of fulfillment.” She notes that an emphasis on counting fruits and vegetables, or carbs misses the broader psychological issues. Self-love and its link to self-efficacy may well be underlying attitudes that lead some patients to take action to alter their habits in favor of life and longevity. Having something to move toward and live for may unconsciously impact eating and exercising choices.

Proposed systems solutions from the online discussion include a uniform approach to managing health behaviors, specific diseases and chronic care patients, and the courage, belief, and vision to move beyond a litany of barriers to make a new model of practice happen.

**STRENGTH IN VULNERABILITY**

A qualitative research study of doctors exposing their vulnerability and an essay in which a physician shares a patient’s story and her own vulnerability unleashed eloquent further reflections. About the qualitative study, Candib notes that “Malterud and Hollnagel teach us to find health in sick people, find strength in doctors’ vulnerability, and find objective ways to study doctors’ personal experiences.” Regarding Shield’s essay, she notes that “revealing ourselves to our patients at critical moments can be an act of strength that promotes healing.”

Brody, quoting Reich, reminds us that compassion means “to suffer with,” and notes that “there can be no compassion without vulnerability.” He describes the phases of silent, then expressive compassion, followed by the formation of a new identity in compassion.

Perhaps an increased opportunity for compassion, or at least the ability to abide with patients through both suffering and joy, is the reason that an international study of physicians’ valuation of personal continuity in care yielded calls for developing systems to support these ongoing relationships. A strong dissenting voice identifies family physicians’ advocacy for continuity as a stumbling block to critical evaluation of whether alternatives to personal continuity might have more beneficial effects on patient outcomes.

Swanson was stimulated by Shield’s essay to relate her theory of caring and 5 ways of relating that reveal the process of caring: knowing, being with, doing for, enabling, and maintaining belief.

Other writers shared their insights and modeled how sharing vulnerability can be healing. Discussants spoke of the connectedness of all mothers, the power of birth, and the trauma of the loss of human possibility.

A discussant of the qualitative study of frequent attenders in the last issue points out that many frequent attenders (“heartsink patients”) present with somatic complaints to the GPs, returning again and again because they underlying issues are not addressed.

Foreshadowing the message of the essay by Tarn in this issue, she notes how important the handling of these underlying issues is in helping patients to listen to the messages carried by their bodies and in fostering healing.

Interestingly, Candib’s essay “Making Time to Write” elicited calls to use writing as a way to witness, reflect, integrate, consolidate relationships, and exchange ideas. Others took inspiration for their own writing and called for a forum for “giving voice to our personal experiences as physicians.”

These reflections on writing are relevant here because even though only “the most talented writers among us will get some of those writings published… all of us will benefit from reflecting on what we do, and from the healing that comes in the creation of stories”—a parallel process for what shared stories can do for our patients.

Please join this community of knowledge at http://www.annfammed.org. Click on “Discussion of articles” or follow the links for the article on which you wish to comment.

**References**


STFM Launches New Partners Initiative


First, STFM will work to realize the vision of the Future of Family Medicine (FFM) report and the New Model of practice. The STFM Board has commissioned a special task force to focus the Society’s efforts and to prioritize how we can bring the New Model into our residency and departmental practices. Our efforts need to be consistent with our strategic plan, take advantage of our unique capacities, and make a difference. Watch for future developments and announcements.

Second, STFM will continue its focus on providing valuable opportunities for growth and renewal. The STFM Board will work closely with its standing committees to ensure that STFM meetings and scholarly forums provide the right mixture of presentation opportunities, scholarship, and careful examination of the issues we face as a discipline.

Third, STFM is launching the New Partners Initiative (NPI) to assist departments, residency programs, and individual faculty members to develop relationships with new partners, to develop new attitudes and fund-raising skills, and to develop new sources of support.

The impetus for NPI lies in the following trends:
- Title VII support hangs on precariously; funding has been reduced and, if it survives, it will likely take a very different form in the next several years
- Medicare/Medicaid indirect and direct medical education payment rates have dropped considerably in the past 10 years, and current pressures on the federal budget suggest that they will continue to drop
- Insurance reimbursement and managed care payment rates, such as IME and DME, continue to decline

Over the years, academic family medicine organizations and faculty have become dependent on these declining funding streams. In recent years we've spent enormous amounts of time and effort on sustaining them, with increasingly diminishing returns. The FFM report noted that, "Faculties are ... consumed by the demands of clinical care and teaching, with little time, energy, motivation, or resources for scholarly inquiry."1

The good news from FFM is that the American people want a good relationship with a personal physician. There are foundations, corporations, individuals, and community groups poised to become our partners in health care and education. We need to develop new attitudes, new partnerships, and new funding approaches that project excellence, competence, and the value of our contributions.

During the coming months NPI will roll out in 3 phases:
- The NPI Think Tank will convene on September 28, 2005 at the AAFP Scientific Assembly in San Francisco to identify appealing big ideas and “funding magnets,” to suggest key strategic relationships, and to provide overall direction for the project. The Think Tank will be chaired by Michael Rosenthal, MD, Thomas Jefferson University. Other Think Tank members include William Mygdal, EdD, STFM president; Macaran Baird, MD, MS, University of Minnesota; Roland Goertz, MD, Heart of Texas Community Health Center, Waco, Tex; Evelyn Lewis & Clark, MD, MA, Pfizer Pharmaceuticals and Uniformed Services University, William Hueston, MD, Medical University of South Carolina, Stephen Bogdewic, PhD, Indiana University; Laurence Bauer, MSW, MEd, Family Medicine Education Consortium; Susan Kaye, MD, Atlantic Health System, Summit, NJ; David Lanier, MD, AHRQ; and Roger Sherwood, STFM executive director.
- The Academic Fundraising Fellowship is accepting enrollees now, and each year it will offer 2 workshops and an optional year-long one-on-one coaching program. Participants will learn how to clarify their academic vision, develop new relationships, and improve fund-raising skills. Faculty will include Mr Bauer, Dr Bogdewic, James Gillespie, PhD, CFRE, and Alan Douglass, MD, Middlesex Hospital, Middletown, Conn. Contact Mr Bauer at 937-428-7866 or lrbauer@infinet.com for more information.
- The NPI Task Force will take the advice and ideas of the Think Tank and translate them into usable and accessible fund-raising training materials. Task Force members will be appointed by the STFM Board and
will include liaison members from other family medicine organizations.

NPI will be taking STFM in some new directions as it develops promising approaches to the needs of the discipline. We believe this initiative is an important step by the Society as we move the Future of Family Medicine forward.

William Mygdal, EdD
STFM President

Call For Nominations for Research Awards
Submit Your Nomination for 2006 Curtis Hames Research Award
The Society of Teachers of Family Medicine is accepting nominations for the 2006 Curtis G. Hames Research Award in Family Medicine to be presented at the 2006 Annual Spring Conference, April 26-30, in San Francisco, Calif. The award, supported by the Hames Endowment of the Medical College of Georgia, is intended to honor those individuals whose careers exemplify dedication to research in family medicine.

The award recipient is selected by a committee representing STFM, the American Academy of Family Physicians, and the North American Primary Care Research Group. Previous Hames Award recipients are on the STFM Web site listed at http://www.stfm.org/awards/awardhub.html.

Nomination letters and CVs must be postmarked by November 11, 2005, and should be addressed to STFM, 11400 Tomahawk Creek Parkway, Leawood, KS 66211. Contact Kay Frank, STFM, with questions at 800-274-2237, ext. 5402, kfrank@stfm.org.

Could Your Last Study Win the STFM Best Research Paper Award?
The Research Committee of the Society of Teachers of Family Medicine is now accepting nominations for the 2006 STFM Research Paper Award, to be presented at the 2006 Annual Spring Conference, April 26-30, in San Francisco, Calif.

The award is intended to recognize the best research paper published by an STFM member in a peer-reviewed journal between July 1, 2004, and June 30, 2005. The STFM Research Committee bases the award selection on the quality of the research and its potential impact. Previous STFM Best Research Paper Award recipients are listed on the STFM Web site at http://www.stfm.org/awards/awardhub.html.

Ten copies of the paper should accompany each nomination letter that documents the potential effect of the paper and its importance to patients’ health and well-being.

November 11, 2005, is the postmark deadline for nominations. Send nominations to STFM, 11400 Tomahawk Creek Parkway, Leawood, KS 66211. Contact Kay Frank, STFM, with questions at 800-274-2237, ext. 5402, kfrank@stfm.org.

Traci Nolte
STFM Communications Director

References

FAMILY MEDICINE LEGISLATIVE ADVOCACY: OUR POWERFUL MESSAGE
The discipline of family medicine stands at a critical juncture. We face the pressures of dwindling medical student interest and a shift in workforce policy toward a greater emphasis on market-driven forces. Population-based studies show that care provided by family physicians results in lower health care costs and improvements in quality and health outcomes. The regulatory influence of government agencies upon health policy has never been more important than it is now.

The Academic Family Medicine Advocacy Alliance (AFMAA) organizes legislative activities for ADFM, STFM, AFMRD and NAPCRG. At the annual Congressional Conference in April 2005, our members met with legislators and government officials. Senator Barak Obama expressed a clear understanding of our character when he pronounced: “Family physicians are the doctors who always put the interest of their patients ahead of their own.” This sentiment was shared by other legislators who, if properly informed, could be champions for policies that will benefit the health of the people of the United States. Unfortunately, they are not yet properly informed.

The Power of the Proper Composition of the Physician Workforce
Legislators were attentive to information from recent studies of health outcomes, which indicate that higher quality care can be achieved at a lower cost when the physician workforce is composed of the appropriate proportion of generalist physicians. These data suggest a potential cure for a health care system that Sena-
tor Obama described as “in the throes of a meltdown.” They were captivated by this information because the changes necessary for improvement in quality and reduction in spending could be described in simple terms. Even though this information is the compelling foundational argument for all legislation that supports the practice of family medicine, it is neither understood nor well articulated by most family physician clinicians or educators.

The studies from Johns Hopkins1-4 and Dartmouth4-6 are powerful population-based investigations that examined health outcomes and quality indicators in industrialized nations, states, and counties. In composite, the data suggest that optimal health outcomes occur when 40% to 50% of the physician workforce is made up of family physicians, general internists, and general pediatricians.

The Dartmouth studies examined entire Medicare data sets for several years, and compared the spending by each state with 24 quality indicators.3-7 As annual spending per Medicare beneficiary increased, quality of care declined significantly. As the number of generalist physicians increased, the quality of care improved and the costs declined. Conversely, as the number of specialist physicians in the population increased, the quality indicators declined and the costs rose.

States at the 75th percentile of quality spent about $1,600 less per beneficiary per year than states at the 25th percentile, and states at the 75th percentile in spending had about 40% fewer generalist physicians per capita than states at the 25th percentile (2.4 vs 3.9 per 10,000 people). An appropriate increase in the proportion of generalist physicians will lead to improved quality and savings of perhaps $60 billion or more per year for care of the nation’s 41,000,000 Medicare beneficiaries.

Radical changes in the US health care system must occur to support this balanced workforce, including examination of medical school admissions, increased reimbursement for generalist physicians who provide personal medical homes for patients, and incentives for systems that demonstrate high quality. The balance of spending for health care must shift toward preventive medicine and public health policies that provide access to health care for all.

Two things must be done to properly inform those who make laws and implement policy. First, we must develop enduring relationships with our legislators. We must also become conversant in the studies that show the positive effect of our discipline on the nation’s health outcomes. Our legislators already know that we are passionate about the health of our patients and our nation. Now we must become their trusted advisors who can demonstrate that our passion improves outcomes and lowers costs. They will be eager to listen to this story.

Jerry Kruse, MD, MSPH, Association of Departments of Family Medicine

References
economic health care disparities are increasing. Clearly, now is the time for innovation.\(^3\text{,}^4\)

The AFMRD is making a concerted effort to assist residency programs that are considering an electronic health record (EHR) system. At our Program Director’s Workshop (PDW) in June 2005, vendors were invited to demonstrate their EHR systems. They presented business plans to program directors, describing how their EHR systems would affect not only the residency programs’ financial bottom line but also quality and patient safety. The AFMRD has also been working with family medicine residency programs and the Residency Review Committee for Family Medicine to promote innovative ideas in family medicine education, with several residency programs suggesting alternative curricula and educational tracks while still meeting all program requirements.

To encourage high-quality leadership, the AFMRD created 3 levels of awards: bronze, silver, and gold, which were presented for the first time in June. To promote ongoing faculty development, key to providing high-quality education for our trainees, the PDW joined with the National Institute for Program Director Development (NIPDD) to offer introductory and advanced fellowships.

Mastering complex chronic cases will require a skilled workforce. Residency training will equip our trainees with the tools needed to ensure lifelong learning, to use technology and evidence-based medicine, and to optimize the health of the populations they serve. The AFMRD, which is also committed to recruiting and mentoring students, is publishing a “cookbook” for mentoring medical students as well as to help students understand the benefit of a career in family medicine. The goal is to recruit only the best and brightest, both intellectually and emotionally.\(^2\text{,}^4\)

The osteoporosis and diabetes preceptorships continue to be well received as evidence-based educational training programs for residents. Besides their educational value, the programs are designed as a resource for research and training in the core competencies of “systems-based practice and practice-based learning.” The AFMRD board is researching the programs’ impact on clinical care and core competencies. Outcomes- and clinical-based research are vital components of the future of family medicine.

Collaborations and effective medical teams are central to health care environment improvements and quality improvement. Our preceptorships represent a successful collaboration of family medicine, other primary care disciplines, specialists, allied health professionals, educators, and researchers. Within our own family of organizations, we have ongoing dialogue and project planning with STFM, ADFM, ABFM and AAFP. Discussions with STFM leadership have focused on joint curriculum development. Conversations with ADFM have focused on implementation of electronic medical records and their role in the future of family medicine. AFMRD leaders visited the ABFM at their home office in Lexington, Ky. Our 2 organizations have maintained an active dialogue as both plan strategies and innovations that will benefit both organizations.

Webster’s dictionary defines quality as a distinguishing attribute, an inherent feature, a degree of excellence, and an acquired skill or accomplishment.\(^5\)

Perhaps, the most important qualities we can instill in our graduating residents are the ability to ask questions that lead to research, and open minds that will lead to innovation. We must also instill the desire to change and evolve, to confront complacency, and challenge the specialty to a higher standard.

\(\text{Penny Tenzer, MD, President Association of Family Medicine Residency Directors}\)

**REFERENCES**


**RESEARCH INVOLVING LATINO POPULATIONS**

The US Latino population, at more than 31 million, is the largest minority group in this country, and their numbers are expected to reach more than 96 million in the next 50 years.\(^1\) This reality, along with the health disparities faced by Latinos,\(^2\) highlights the need to promote research involving this population. As stated by the Office on Minority Health: “Hispanics/Latinos are disproportionately underrepresented in research activities. Without adequate and targeted research, Hispanics/Latinos are disadvantaged in policy making.
resource allocation, program planning, and program implementation activities. A number of challenges must be addressed, however, when planning research involving Latinos.

First, much variability within the Latino population is based on country of origin and acculturation status. Two of 5 Latinos are foreign-born, and many born in the United States adhere to customs from their country of origin. The cultural variability among Latinos from different countries can affect health behaviors. Thus, we need to routinely assess Latinos’ country of origin and understand its importance. Acculturation, the process of assimilating to the majority culture, has also been shown to influence health-related behaviors and is affected by individual factors, such as a person’s age upon arrival to the United States, level of education, number of years in the United States, and support systems. The environment into which one is assimilating can also affect acculturation. In the past, Latinos were likely to settle in metropolitan areas with already large Latino populations, such as New York and Miami. This strategy allowed Latinos to live in areas that already accommodated their culture and language. More recently Latinos have settled in smaller cities or more rural areas that do not have a large Latino community. Latinos in these areas may face difficulties in overcoming language and cultural barriers not encountered by those in more urbanized areas. Because of these differences, research from both urban and outlying communities is needed, and care must be taken during sample selection to make a project’s results meaningful to different Latino communities.

Recruiting Latinos for research projects involves unique issues. Standard recruitment practices, such as recruiting from clinics or through mainstream media, may not reach a desired population. It is essential to understand the influence of language barriers, immigration status, and distrust toward the mainstream culture on recruitment. Misconceptions, poor education, and distrust regarding research must also be addressed. Minority communities might not want to participate in research because of past experience in which they have felt used, rather than engaged as full partners. Involving knowledgeable community members early in the research process can help overcome this barrier. The principles for community-based research articulated by NAPCRG are particularly relevant for research with and for Latino communities.

A final key ingredient is including qualified bilingual/bicultural researchers on the team. NAPCRG and academic health science centers should focus on how such investigators can be developed and retained. Some examples of developing centers of excellence for Latino health research include the University of Texas Health Science Center, which includes sites in San Antonio and Houston, and the University of California, which includes sites in San Diego and Los Angeles. It is noteworthy that NAPCRG is planning its 2008 meeting for Puerto Rico in a specific effort to make the meeting more accessible to Spanish-speaking countries in the Western Hemisphere. Look for a growing visibility of research for and with Latinos at NAPCRG meetings.

References


AAFP’S POLITICAL ACTION COMMITTEE ADDS TO FAMILY PHYSICIANS’ VOICE IN WASHINGTON

Family physicians have a new tool for voicing their positions on everything from Medicare physician payment to medical liability reform to incentives for providing health care to the underserved.

That tool: FamMedPAC, the AAFP’s new federal political action committee. Since its June launch, FamMedPAC has retained Mark Cribben, JD, as its director and garnered more than $70,000 in donations. The PAC board has set a goal of raising and contributing $1 million in this election cycle. The donations will be allocated to political candidates who meet the PAC’s criteria for family medicine’s support.

Approved by the 2004 Congress of Delegates, FamMedPAC will offer several benefits, according to PAC Board Chair, Michael Fleming, MD, of Shreveport, La.

“The chief benefit is ensuring that your specialty is adequately represented when Congress addresses long-term issues such as skyrocketing medical liability costs, unfair reimbursement rates, reduced physician training or the growing number of uninsured,” he said in an introductory letter about FamMedPAC to Academy members. “Without a consistent, concerted voice that compels politicians to listen, your views may not be heard.”

FamMedPAC donors also will receive insider information on political dealings influencing family medicine, potential candidates, and indications of which candidates deserve FamMedPAC support.

“Other benefits to you are intangible, but quite real,” said Fleming in his letter. “By joining FamMedPAC, you will have the ability to become more actively involved in the legislative and political action process. Through a greater understanding of the issues, you can become more comfortable initiating person-to-person contact with your elected officials and you strengthen your relationship with your representative and senators.”

Moreover, according to Cribben, family physicians can make their political contributions stretch farther via the PAC. Federal law limits individual donations to a political candidate to $2,100 per election cycle. But AAFP members can contribute up to $5,000 per year to FamMedPAC.

And the PAC can combine the donations from AAFP members to make even larger contributions to candidates who deserve the support of family medicine, said Cribben.

As a result, family medicine will reap several benefits, including:

- Opportunities to help lawmakers understand the implications of their decisions on issues, such as physician payment, family medicine training, and access to health care coverage
- The ability to disseminate political information to AAFP membership without jeopardizing the Academy’s tax-exempt status
- A tangible method of using family physicians’ collective clout to hold federal lawmakers accountable for their legislative actions

“And it will encourage members to get involved,” said Cribben. “We encourage members to get in touch with me or PAC board members and express their opinions about candidates.”

Board members will use several criteria—including whether the candidates are family physicians, their leadership position in Congress or membership on important legislative committees, and their voting records and positions on issues related to priorities of family physicians and their patients—for identifying federal candidates to receive FamMedPAC support.

“Another consideration may be the difficulty of the candidate’s race,” said Cribben. For example, the board may provide more support to friendly candidates who face difficult races.

Like all groups that provide information and feedback to lawmakers, the AAFP and its sister organizations have a solid grassroots program through action alerts and the key contact program. Most recently, both effectively communicated family medicine’s position on federal funding for primary care education through Section 747 of Title VII of the Public Health Service Act. As a result of those efforts, the Senate voted to appropriate $90 million for Title VII, despite House and Bush administration efforts to zero out federal support.

FamMedPAC will augment those efforts with direct political involvement, according to Cribben.
The PAC began accepting donations online in August. The FamMedPAC Web site directs members to a password-protected page that will provide information about the committee’s philosophy, goals, and current activities. Members then can click to a contribution page.

The PAC will offer contributors several options, said Cribben.

“The site will allow online credit card contributions, and we will accept contributions by cash, check and credit card at meetings throughout the year,” he said. “We also give members who contribute by credit card the option of giving through periodic, automatic payments.”

For more information about FamMedPAC, go to http://www.aafp.org/x34131.xml.

Leslie Champlin
AAFP News Department

LISTENING TO THE DIPLOMATES:
PHYSICIANS’ FEEDBACK ON SELF-ASSESSMENT MODULES

The Self-Assessment Module (SAM) is an integral part of the ABFM’s Maintenance of Certification for Family Physicians (MC-FP) process as well as an important learning opportunity. Contrary to what some may think, the SAMs are not tests, but rather self-evaluations and self-teaching experiences. A SAM consists of 2 parts: a knowledge assessment and a clinical simulation. Once a Diplomate completes the knowledge assessment portion for the first time, the missed questions will appear on second and subsequent tries with both references and critiques, making the SAM much more of a learning process than any type of examination. Whenever a physician completes a SAM, we solicit feedback regarding their experience to continually improve these modules. The results of this feedback for our first 4 modules reflect a strong positive response. In the areas of (1) relevance of information to clinical practice, (2) currency of information, (3) usefulness of information, and (4) overall value of the module, the average ratings fall at 5.4 on a 6 point scale (Figures 1-4).

Perhaps the most important feedback indicator is whether the information relayed in these educational modules effects positive outcomes by improving practice. In the case of the first 2 health topics offered in 2004, for which we have at least 1 year’s record of physician evaluation, 54% of those taking the diabetes module indicated that they would change their practice as a result of participating. Ninety percent of those physicians volunteered written comments describing what changes they would make. For the hypertension module these values were within 1% of the diabetes module (nearly identical), providing a strong suggestion that both are high-quality educational materials.

These results bear out the goals established in the development of the SAMs as the primary facet of Part II of MC-FP, Self-Assessment and Lifelong Learning.
namely, to enhance family physicians' knowledge and skills in areas that are of greatest importance to them and to provide continuous opportunity for improvement in the quality of care they deliver to the American public.

Any Comments?
Physicians are also offered the opportunity to make an open response comment at the end of the SAM evaluation. The ABFM has used this feedback—both positive and negative—to modify and update the SAM during the first 18 months to improve both user comfort and the functionality of the SAM itself. Examples of physicians' comments regarding the SAM's relevance as an educational tool are as follows:

“I think that this is an excellent tool for assessing and attaining knowledge in a non-biased format. Excellent update on clinical guidelines, as well as a review on the important basic science aspects of diabetes. I was very impressed with the whole set.”

“The knowledge assessment portion was quite decent and appropriate. Clearly an educational activity.”

“I think improvements have already been made since I took the SAM on diabetes last year. I really liked the way the reference came up with the question in the review instead of having to shift from page to page. That reinforces learning in a friendly way.”

“It was excellent ... the immediate critique of missed answers was most helpful.”

Michele Mason
Jason Rinaldo, PhD
American Board of Family Medicine