The Effects of a Smoking Cessation Intervention on 14.5-Year Mortality: A Randomized Clinical Trial
Nicholas R. Anthonisen, Melissa A. Skeans, Robert A. Wise, Jure Manfreda, Richard E. Kanner, John E. Connett for the Lung Health Study Research Group*

This article provides long-term mortality results of a randomized trial of a smoking cessation program. Only 21.7% in the intervention group had stopped smoking at 5 years (compared with 5.4% of controls). Nonetheless, all-cause mortality per 1000 person-years was 8.83 deaths in the intervention group and 10.38 deaths in the control group. Smoking cessation programs substantially reduce mortality even when only a minority of patients stop smoking.

A Randomized, Controlled Trial of Combination Therapy for Chronic Hepatitis B: Comparing Pegylated Interferon-α2b and Lamivudine with Lamivudine Alone
Henry Lik-Yuen Chan, Nancy Wai-Yee Leung, Alex Yui Hui, Vincent Wai-Sun Wong, Choong-Tsek Liew, Angel Mei-Ling Chim, Francis Ka-Leung Chan, Lawrence Cheung-Tsui Hung, Yuk-Tong Lee, John Siu-Lun Tam, Christopher Wai-Kei Lam, and Joseph Jao-Yiu Sung

In patients with hepatitis B e antigen–positive chronic hepatitis B, combination treatment with pegylated interferon-α2b and lamivudine may lead to a higher rate of virologic response than lamivudine monotherapy. The rate of sustained virologic response was 36% for combination therapy and 14% for lamivudine monotherapy.

A Cost-Effectiveness Analysis of Combination Antiplatelet Therapy for High-Risk Acute Coronary Syndromes: Clopidogrel plus Aspirin versus Aspirin Alone
Mark D. Schleinitz and Paul A. Heidenreich

In patients with high-risk acute coronary syndromes, 1 year of therapy with clopidogrel plus aspirin followed by life-long aspirin results in greater life expectancy than life-long aspirin alone. The cost-effectiveness of adding clopidogrel—$15 400 per quality-adjusted life-year—is similar to that of many well-accepted interventions.
Improving Patient Care

Systematic Review: The Relationship between Clinical Experience and Quality of Health Care
Niteesh K. Choudhry, Robert H. Fletcher, and Stephen B. Soumerai

The authors systematically reviewed studies relating medical knowledge and health care quality to years in practice and physician age. Seventy-three percent of the evaluations showed decreasing performance with increasing years in practice for all or some of the outcomes assessed. Four percent of evaluations showed improving performance with increasing experience. Physicians who have been in practice longer may be at risk for providing lower-quality care.

Academia and Clinic

An Evidence-Based Guide to Writing Grant Proposals for Clinical Research
Sharon K. Inouye and David A. Fiellin

This article provides recommendations on how to write a grant for clinical research. It describes specific problems that grant reviewers frequently identify in their critiques and shows how to avoid these problems.

Updates

Update in Pulmonary Diseases
Martin J. Tobin

This year's Update in Pulmonary Diseases incorporates articles on mechanical ventilation, obstructive lung disease, and pulmonary infection.

Reviews

Narrative Review: Celiac Disease: Understanding a Complex Autoimmune Disorder
Armin Alaedini and Peter H.R. Green

This review discusses current concepts in the clinical presentation and diagnosis of celiac disease. It describes the pathogenesis of the disease; the diagnostic usefulness of serologic markers, including the sensitivity and specificity of available tests; and the association of celiac disease with other disorders.

Editorials

Smoking Kills: Experimental Proof from the Lung Health Study
Jonathan M. Samet

The new results from the Lung Health Study, reported in this issue, confirm again that smoking cessation prolongs life. In addition to their public health importance, these findings remind clinicians that interventions do increase the rate of successful quitting. The implications are obvious: Physicians should obtain a smoking history from all patients, and they should help smokers quit.
"Practice Makes Perfect" ... Or Does It?
Steven E. Weinberger, F. Daniel Duffy, and Christine K. Cassel
The medical profession cannot ignore the striking findings reported by Choudhry and colleagues and their implications: Practice does not make perfect. Physicians must make an ongoing vigorous effort to maintain their knowledge and skills, and they must work actively to sustain the quality of care in their practice.

On Being a Doctor

She Is a Beautiful Lady
Sara Sasha Battar
Dean, 87, and his wife Donna, 78, would ritualistically arrive every 3 months in my geriatrics clinic, rain or shine. They had been married for 61 years. From the beginning, Donna ably presumed the spokesperson's role and impressed me as a reliable and caring informant who knew Dean more than he knew himself.

Letters

Cardiac Resynchronization Therapy in Heart Failure
Enrique V. Carbajal, Grace W. Huang, and Billy Hu
Melanie Calvert, Nick Freemantle, and John G.F. Cleland
Finlay A. McAlister, Natasha Wiebe, and William Abraham—RESPONSE

Relative Cost-Effectiveness of Different Tests for *Chlamydia trachomatis*
Brian Jackson
Sue J. Goldie, Delphine Hu, and Edward W. Hook, III—RESPONSE

Treating Controls in Unblinded Trials
Paul S. Heckerling
Lucy Yardley—RESPONSE

Debilitating Muscle Cramps after Teriparatide Therapy
Parul J. Kakaria, David J. Nashel, and Eric S. Nylen

Statins and Nasal Polyps
Caterina Bucca, Andrea Marsico, Erika Panaro, Patrizia Bigo, and Luisa Brussino

Correction: Cardiac Resynchronization in Patients with Symptomatic Heart Failure

Medical Writings: Book Notes

Medical Malpractice: A Physician's Sourcebook
David Hsia

Contemporary Catholic Health Care Ethics
David L. Freeman

Ad Libitum

Sherbet
Jerald Winakur
The Effects of a Smoking Cessation Program on Long-Term Survival in Smokers with Mild Lung Disease

Combination Therapy for Chronic Hepatitis B

Is It Cost-Effective To Treat High-Risk Cardiac Patients with Clopidogrel plus Aspirin as Opposed to Aspirin Alone?

Is an Older, More Experienced Doctor a Better Doctor?
The Effects of a Smoking Cessation Intervention on 14.5-Year Mortality
A Randomized Clinical Trial
Nicholas R. Anthonisen, MD; Melissa A. Skeans, MS; Robert A. Wise, MD; Jure Manfreda, MD; Richard E. Kanner, MD; and John E. Connett, PhD, for the Lung Health Study Research Group*

Background: Randomized clinical trials have not yet demonstrated the mortality benefit of smoking cessation.
Objective: To assess the long-term effect on mortality of a randomly applied smoking cessation program.
Design: The Lung Health Study was a randomized clinical trial of smoking cessation. Special intervention participants received the smoking intervention program and were compared with usual care participants. Vital status was followed up to 14.5 years.
Setting: 10 clinical centers in the United States and Canada.
Patients: 5887 middle-aged volunteers with asymptomatic airway obstruction.
Measurements: All-cause mortality and mortality due to cardiovascular disease, lung cancer, and other respiratory disease.
Intervention: The intervention was a 10-week smoking cessation program that included a strong physician message and 12 group sessions using behavior modification and nicotine gum, plus either ipratropium or a placebo inhaler.
Results: At 5 years, 21.7% of special intervention participants had stopped smoking since study entry compared with 5.4% of usual care participants. After up to 14.5 years of follow-up, 731 patients died: 33% of lung cancer, 22% of cardiovascular disease, 7.8% of respiratory disease other than cancer, and 2.3% of unknown causes. All-cause mortality was significantly lower in the special intervention group than in the usual care group (8.83 per 1000 person-years vs. 10.38 per 1000 person-years; P = 0.03). The hazard ratio for mortality in the usual care group compared with the special intervention group was 1.18 (95% CI, 1.02 to 1.37). Differences in death rates for both lung cancer and cardiovascular disease were greater when death rates were analyzed by smoking habit.
Limitations: Results apply only to individuals with airway obstruction.
Conclusion: Smoking cessation intervention programs can have a substantial effect on subsequent mortality, even when successful in a minority of participants.

METHODS
The design of the LHS has been described in detail elsewhere (2). The participants, all volunteers, were smokers who did not consider themselves ill but had evidence of mild to moderate airway obstruction (2). Individuals with serious disease, hypertension, obesity, or excessive alcohol intake were excluded. The primary research questions were whether a smoking cessation program and use of inhaled ipratropium would decrease the rate of decline of lung function and would affect mortality and morbidity over 5 years. These results have been reported elsewhere (3, 4). The smoking cessation program was associated with cumulative reduced decline in lung function (FEV₁) that was largest in participants who stopped smoking early in the study; inhaled ipratropium produced a small noncumulative increase in FEV₁ that disappeared when the drug was withdrawn (3). Intention-to-treat analysis after 5 years did not reveal differences in morbidity or mortality among treatment groups (4), although subgroup analysis showed that smoking cessation was associated with significant reductions in fatal or nonfatal cardiovascular disease and coronary heart disease. This paper reports the effects of the study intervention on mortality in LHS participants 14.5 years after randomization.

See also:
Print
Editors’ Notes ........................................... 234
Editorial comment ..................................... 299
Summary for Patients ................................. I-12
Web-Only
Appendix
Conversion of figures and tables into slides
The Effects of a Smoking Cessation Intervention on Mortality

Although there are many health benefits for smokers who stop smoking, we still lack evidence from randomized, controlled trials that smoking cessation programs reduce mortality.

In this randomized, controlled trial of a 10-week-long smoking cessation intervention in 5887 smokers with asymptomatic airway obstruction, 14-year mortality rates were higher in the usual care group than in the smoking cessation group (hazard ratio, 1.18 [95% CI, 1.02 to 1.37]). The mortality benefit was greatest among the 21.7% of the intervention group who actually managed to quit smoking.

Smoking cessation programs substantially reduce mortality even when only a minority of patients stop smoking.

—The Editors

a wide variety of techniques (5). In 10 clinical centers, 5887 participants were randomly assigned to 3 groups. Two special intervention groups received an intensive 10-week smoking cessation program. Briefly, the cessation intervention consisted of a strong physician message and 12 two-hour group sessions, using behavior modification and nicotine gum. Quitters entered a maintenance program that stressed coping skills. One special intervention group also received ipratropium, while the other received a placebo inhaler. A third group received usual care. About 75% of the original participants were followed continuously for the subsequent 10 years by biannual telephone contacts and 1 clinic visit at approximately 11 to 12 years after randomization (6). Telephone contacts served to check smoking status, morbidity, and mortality and were not part of the intervention.

All study participants provided written informed consent for the original LHS before beginning the study. The consent documents stated that smoking increases the risk for chronic obstructive pulmonary disease, respiratory tract cancer, and cardiovascular disease and that smoking cessation would decrease such risks. Additional written informed consent was obtained from persons who participated in the biannual telephone calls. Institutional review boards at each of the 10 clinical centers and the coordinating center approved the study design and consent documents.

When biannual phone calls revealed a participant death, staff attempted to collect death certificates, autopsy reports, relevant medical records, and interviews with attending physicians or eyewitnesses. An independent mortality and morbidity review board examined these data and classified causes of death. In addition, a National Death Index review provided date and cause of death for all U.S. study participants through the end of 2001. Vital status at 31 December 2001 or 14.5 years, whichever was earlier, was successfully determined for 98.3% of all participants; missing individuals were Canadians who had been lost to follow-up and were not accessible through the National Death Index. Mortality end points were classified in 7 categories: coronary heart disease, cardiovascular disease including coronary heart disease, lung cancer, other cancer, respiratory disease excluding lung cancer, other, and unknown. The “other” category included but was not limited to liver disease, kidney disease, sepsis, accidents, suicide, and AIDS.

Analyses were performed on an intention-to-treat basis, comparing the special intervention group with the usual care group. The special intervention group was a combination of the groups originally assigned to receive inhaled ipratropium or placebo therapy. Both of these groups, which were very similar at baseline, received the smoking cessation program and exhibited similar rates of smoking cessation (3). Participants were also divided into 3 groups according to smoking history during the initial 5 years of the trial. Sustained quitters were participants who stopped smoking in the first year after randomization and maintained biochemically validated abstinence (3) throughout follow-up. Continuing smokers were participants who reported smoking at all follow-up visits. Intermittent quitters were participants who reported smoking at some but not all of their follow-up visits or during the time between visits.

Statistical Analysis

Baseline differences between the special intervention and usual care groups were tested by using t-tests for continuous variables and chi-square statistics for categorical variables. Cause-specific death rates and times to events were analyzed by using the Kaplan–Meier product-limit method (7). Survival was compared among groups by using the log-rank test. Hazard ratios and adjusted analyses were obtained by using the Cox proportional hazards model. Interactions were assessed by comparing hierarchically related proportional hazards models. All P values result from 2-sided tests; no adjustments were made for multiple comparisons.

Role of the Funding Source

This study was funded by a contract and grants from the National Heart, Lung, and Blood Institute of the National Institutes of Health. The funding source had a role in the design of the study and approved the manuscript before it was submitted for publication.

RESULTS

Baseline characteristics of LHS participants are shown in Table 1. Most were middle-aged; smoked heavily; and had substantial smoking histories, airway obstruction (FEV1/FVC ratio ≤ 70%), and borderline low FEV1 val-
On average, participants were normotensive and had normal body mass indices. Most participants were of white ethnicity; 37% were women. The average participant had some postsecondary education and did not drink heavily.

There were 731 known deaths among LHS participants, as shown in Table 2. Lung cancer was the most common cause of death (n = 240 [33%]). Coronary heart disease accounted for 77 deaths (10.5%), and cardiovascular disease including coronary heart disease accounted for 163 deaths (22%). One hundred fifty-four participants (21%) died of cancer of organs other than the lung. Deaths due to respiratory disease other than cancer were relatively uncommon (n = 57 [7.8%]). The cause of death was unknown in only 17 participants (2.3%). Mortality did not significantly differ between the special intervention groups originally assigned to ipratropium or placebo (Table 2).

Figure 1 shows all-cause survival rates in the 2 treatment groups. Death rates were significantly higher in the usual care group than in the special intervention group (10.38 per 1000 person-years vs. 8.83 per 1000 person-years; P = 0.03). The hazard ratio for mortality in the usual care group was 1.18 (95% CI, 1.02 to 1.37) compared with the special intervention group. Figure 2 shows categorical causes of death in the 2 treatment groups. In all categories except “other,” death rates were higher in the usual care group than in the special intervention group, but the difference was significant only for deaths from respiratory diseases not related to lung cancer (1.08 per 1000 person-years vs. 0.56 per 1000 person-years; P = 0.01). When survival was analyzed according to smoking habit, it differed significantly between groups (P < 0.001), even after adjustment for baseline differences (data not shown). Mortality was 6.04 per 1000 person-years in sustained quitters, 7.77 per 1000 person-years in intermittent quitters, and 11.09 per 1000 person-years in continuing smokers. No significant differences in death rates were seen between special intervention and usual care participants in any of the 3 smoking categories. Figure 3 shows categorical causes of death among the 3 smoking groups. Death rates were significantly related to smoking habit for coronary heart disease (P = 0.02), cardiovascular disease (P ≤ 0.001), lung cancer (P = 0.001), and other causes (P = 0.03). Death rates were not significantly related to smoking habit for cancer other than lung cancer and for respiratory deaths not related to lung cancer. Baseline FEV\textsubscript{1}, expressed as a percentage of the predicted normal value, was inversely related to all-cause mortality (P ≤ 0.001) and to deaths from coronary heart disease (P = 0.003), cardiovascular disease (P = 0.002), lung cancer (P = 0.02), other cancer (P = 0.03), and respiratory disease other than cancer (P ≤ 0.001).

Differences between the special intervention group and the usual care group in all-cause mortality were examined in relation to subgroups identified at baseline (Table 3). There was a significant mortality difference between the special intervention and usual care groups in the youngest tertile of participants, those younger than 45 years of age.
The striking feature of our findings is the statistically significant difference in all-cause mortality in the intention-to-treat analysis. Mortality was higher in the usual care group than in the special intervention group even though the intervention—smoking cessation—was successful in only a minority of special intervention participants. Since death rates between special intervention and usual care participants with similar smoking habits did not differ, the differences observed in the groups as a whole were almost certainly due to differential cessation rates. It must be emphasized that this finding applied to a special group of heavy smokers who had preexisting airway obstruction. From its inception, the LHS was characterized by very high follow-up rates. Of the original cohort, only 75 participants (1.27%), all of whom were Canadian, were censored because of loss to follow-up at less than 12.5 years. Cause of death was adjudicated by a mortality and morbidity review board, which had access to data in 653 of the 731 deaths. In the remaining cases, cause of death was derived from the National Death Index. The 17 deaths due to unknown causes showed trends similar to the remainder of deaths in terms of treatment group and smoking status (Figures 2 and 3) and therefore were probably not a source of bias. Smoking status was ascertained at the fifth year following entry into the LHS, that is, 5 years after randomization. We have shown that smoking status established at 5 years changed relatively little in the next 6 years, especially among sustained quitters (6).

To our knowledge, no directly comparable studies have examined the long-term effects of a randomly applied smoking intervention. Many intervention trials aimed at cardiovascular disease have used smoking cessation along with other interventions. Of these, the most directly comparable to the LHS was the Multiple Risk Factor Intervention Trial (MRFIT), which was conducted in North Americ-

Table 2. Causes of Death by Treatment Group*

<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>Special Intervention with Ipratropium (n = 1961)</th>
<th>Special Intervention with Placebo Inhaler (n = 1962)</th>
<th>Usual Care (n = 1964)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
<td>24 (10.6)</td>
<td>22 (9.4)</td>
<td>31 (11.5)</td>
</tr>
<tr>
<td>Cardiovascular disease including CHD</td>
<td>54 (23.9)</td>
<td>46 (19.6)</td>
<td>63 (23.3)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>74 (32.7)</td>
<td>77 (32.8)</td>
<td>89 (33.0)</td>
</tr>
<tr>
<td>Other cancer</td>
<td>50 (22.1)</td>
<td>52 (22.1)</td>
<td>52 (19.3)</td>
</tr>
<tr>
<td>Respiratory disease other than cancer</td>
<td>15 (6.6)</td>
<td>14 (6.0)</td>
<td>28 (10.4)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (11.5)</td>
<td>42 (17.9)</td>
<td>32 (11.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>7 (3.1)</td>
<td>4 (1.7)</td>
<td>6 (2.2)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>226</strong></td>
<td><strong>235</strong></td>
<td><strong>270</strong></td>
</tr>
</tbody>
</table>

* CHD = coronary heart disease. Numbers of deaths in each column do not sum to the totals because deaths in the CHD category are listed separately and are also included in the cardiovascular disease category.

(hazard ratio, 1.88; \( P = 0.001 \)), but not in the middle tertile (45 to 52 years of age) or oldest tertile (53 to 60 years of age). Interaction between treatment group and age was significant \( (P = 0.001) \). Mortality did not differ significantly between groups by sex, and no significant interaction between treatment group and sex was observed. There was a significant mortality difference between the usual care and special intervention groups among participants smoking at least 40 cigarettes per day (hazard ratio, 1.30; \( P = 0.03 \)), but not among those smoking 25 to 39 cigarettes per day or fewer than 25 cigarettes per day. In addition, no significant interactive effect on mortality was observed between smoking intensity and treatment group.

There was a significant difference in mortality between the special intervention and usual care groups for participants in the middle tertile of baseline FEV\(_1\) (75% to 83% predicted) (hazard ratio, 1.39; \( P = 0.01 \)), but not in tertiles with higher or lower values of FEV\(_1\). No significant interactive effect on mortality was observed between treatment group and FEV\(_1\).

**DISCUSSION**

The only significant difference was in respiratory disease other than lung cancer (log-rank test). CHD = coronary heart disease; CVD = cardiovascular disease.

![Figure 2. Mortality rates at 14.5 years by cause.](image-url)
ica, enrolled participants of similar age, and had similar follow-up periods (8). At 16 years after randomization, MRFIT had slightly lower all-cause mortality than the LHS at 14.5 years (10.5% vs. 12.4%). More than 50% of MRFIT deaths were attributed to cardiovascular disease, reflecting the fact that MRFIT participants were selected for cardiovascular risk factors while the LHS attempted to avoid them. However, LHS participants had substantially higher death rates for lung cancer and respiratory disease than did MRFIT participants, reflecting their heavier tobacco use and abnormal lung function. All-cause mortality did not differ significantly between treatment groups in MRFIT at 10.5 years (9) or at 16 years, perhaps in part because smoking habits did not differ greatly between the intervention and control groups after the initial 6-year follow-up (10). Similar convergence in smoking habits was observed in long-term follow-up of 2 European cardiovascular trials (11, 12), both of which initially reported significant decreases in cardiac events in the intervention groups but did not observe significant differences in all-cause mortality at 8.5 and 10 years, respectively. All-cause mortality probably differed between the special intervention and usual care groups in our study because smoking cessation has a powerful effect on mortality in heavy smokers with airway obstruction and because more than 90% of LHS participants who quit smoking during the first 5 years of the study were able to maintain cessation thereafter (6, 13).

We did not measure the cost of the LHS smoking cessation program, and researchers who worked with the intervention group had other roles in the study, such as obtaining follow-up data. However, a unit price of $2000 would probably cover the LHS smoking intervention, including intensive initial counseling, nicotine replacement therapy, and the long-term maintenance program. This seems a modest price for a life-saving intervention. An inexpensive intervention with a relatively low success rate can make an important difference if it has great potential and is applied early in the course of the diseases of interest. Indeed, the most prominent difference between the special intervention and usual care groups was observed in the youngest participants. It could be argued, therefore, that smoking cessation was most effective in preventing truly premature death.

The leading causes of death in the LHS were lung cancer and coronary heart disease, and smoking cessation was of benefit in both (Figure 3). These results are not unprecedented. In MRFIT, smoking cessation in conjunction with other risk modification strategies was shown to decrease morbidity and mortality from coronary heart disease (14), and we observed such an effect within the first 5 years of LHS follow-up (4). These results are compatible with those of many cohort and case–control studies that have shown a decline in death from coronary heart disease within 2 years of smoking cessation (15). In MRFIT, risk for myocardial infarction in participants who still smoked was roughly 3 times that in participants who had stopped smoking more than 5 years previously (15); this finding was similar to our data on death from coronary heart disease (Figure 3).

The mechanisms by which smoking induces coronary events are apparently reversible to some extent in the short term. To our knowledge, our data are the first to show an effect of smoking cessation on the rate of death from lung cancer in the context of a clinical trial. Our data are consistent with those of previous cohort and case–control studi-

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Hazard Ratio (95% CI)</th>
<th>P Value for Special Intervention vs. Usual Care</th>
<th>P Value for Subgroup–Treatment Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35–44 y</td>
<td>1.88 (1.28–2.77)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>45–52 y</td>
<td>1.07 (0.82–1.41)</td>
<td>&gt;0.2</td>
<td></td>
</tr>
<tr>
<td>53–60 y</td>
<td>1.09 (0.89–1.34)</td>
<td>&gt;0.2</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.17 (0.97–1.40)</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.19 (0.92–1.56)</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>FEV1 at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;75%</td>
<td>1.11 (0.88–1.40)</td>
<td>&gt;0.2</td>
<td></td>
</tr>
<tr>
<td>75%–83%</td>
<td>1.39 (1.08–1.81)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>&gt;83%</td>
<td>1.04 (0.76–1.40)</td>
<td>&gt;0.2</td>
<td></td>
</tr>
<tr>
<td>Cigarettes per day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>1.14 (0.86–1.52)</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>25–39</td>
<td>1.07 (0.82–1.40)</td>
<td>&gt;0.2</td>
<td></td>
</tr>
<tr>
<td>≥40</td>
<td>1.30 (1.03–1.65)</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Mortality rates at 14.5 years by cause and smoking status.
ies showing that measurable effects of cessation on lung cancer are usually not evident in the first 5 years and that lung cancer risk is probably still elevated 15 years after smoking cessation (16). In our study, death from lung cancer was roughly 2.2 times more common in current smokers than in sustained quitters (Figure 3), a finding similar to data from cohorts observed for similar lengths of time (16). Smoking is thought to cause potentially irreversible genetic changes in epithelial cells. Therefore, the effects of cessation are probably due to the absence of further insult rather than to reversal of existing disease.

To some extent, the LHS was a study about the FEV1, and our results again demonstrate the prognostic value of this test. It is obvious and axiomatic that death from lung disease other than cancer should be related to FEV1. However, it is not yet clear why FEV1 independent of smoking habits, predicts death from cardiovascular disease (17) and lung cancer (18, 19). The mechanisms involved are likely to be different because FEV1 predicts coronary artery disease in both smokers and nonsmokers (20) but apparently predicts lung cancer only in smokers and former smokers (21). Of interest, our data showed that death from other types of cancer was related to FEV1 but not to smoking habits. These results differ from those of the larger Renfrew and Paisley population study (22), which found that death due to nonrespiratory cancer was not related to FEV1 after smoking had been considered. In addition, good data link smoking to many types of nonpulmonary cancer.

The LHS was one of the few studies that examined a substantial cohort of smoking women (Table 1). Of interest, lung cancer mortality was very similar between sexes: 3.02 per 1000 person-years in men and 3.14 per 1000 person-years in women. This is in agreement with most of the other studies that have examined this issue (22, 23) but is at variance with the case-control study suggesting that women are more likely to develop lung cancer than men given the same smoking exposure (24). In the LHS, female continuing smokers smoked an average of approximately 3 fewer cigarettes per day than did male continuing smokers (5). However, it is difficult to argue that our results support the hypothesis that women are more sensitive to cigarette smoke than men.

In summary, we demonstrated that an intensive smoking cessation program followed by 5 years of reinforcement leads to a substantial and significant reduction in all-cause mortality in people with mild to moderate airway obstruction.

From University of Manitoba, Winnipeg, Manitoba, Canada; University of Minnesota, Minneapolis, Minnesota; Johns Hopkins University, Baltimore, Maryland; and University of Utah, Salt Lake City, Utah.

Grant Support: Lung Health Study III was supported by a cooperative agreement with the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI-1U10-HL59275).

Potential Financial Conflicts of Interest: Honoraria: J.E. Connett (National Institutes of Health/National Heart, Lung, and Blood Institute); Grants received: J.E. Connett (National Institutes of Health/National Heart, Lung, and Blood Institute).

Requests for Single Reprints: John E. Connett, PhD, Coordinating Centers for Biometric Research, University of Minnesota, 2221 University Avenue SE, Room 200, Minneapolis, MN 55414-3080; e-mail, john-c@ccbr.umn.edu.

Current author addresses and author contributions are available at www.annals.org.

References


APPENDIX

The principal investigators and senior staff of the clinical and coordinating centers, the National Heart, Lung, and Blood Institute, members of the Safety and Data Monitoring Board, and members of the Mortality and Morbidity Review Board are as follows.

Case Western Reserve University, Cleveland, Ohio: M.D. Altose, MD, Principal Investigator; C.D. Deitz, PhD, Project Coordinator.

Henry Ford Hospital, Detroit, Michigan: M.S. Eichenhorn, MD, Principal Investigator; K.J. Braden, AAS, Project Coordinator; P.A. Fantuz, RN, BSN; R.L. Jentons, MALLP, Project Coordinator.

Johns Hopkins University School of Medicine, Baltimore, Maryland: R.A. Wise, MD, Principal Investigator; C.S. Rand, PhD, Co-Principal Investigator; K.A. Schiller, Project Coordinator.

Mayo Clinic, Rochester, Minnesota: P.D. Scanlon, MD, Principal Investigator; G.M. Caron, Project Coordinator; K.S. Mieras, L.C. Walters.

Oregon Health Sciences University, Portland, Oregon: A.S. Buist, MD, Principal Investigator; L.R. Johnson, PhD, LHS Pulmonary Function Coordinator; V.J. Bortz, Project Coordinator.

University of Alabama at Birmingham: W.C. Bailey, MD, Principal Investigator; L.B. Gerald, PhD, MSPH, Project Coordinator.

University California, Los Angeles, California: D.P. Tashkin, MD, Principal Investigator; I.P. Zuniga, Project Coordinator.

University of Minnesota Coordinating Center, Minneapolis, Minnesota: J.E. Connett, PhD, Principal Investigator; P.G. Lindgren, MS; M.A. Skeans, MS; H.T. Voelker.

University of Pittsburgh, Pittsburgh, Pennsylvania: R.M. Rogers, MD, Principal Investigator; M.E. Pusateri, Project Coordinator.

University of Utah, Salt Lake City, Utah: R.E. Kanner, MD, Principal Investigator; G.M. Villegas, Project Coordinator.

Safety and Data Monitoring Board: C. Furberg, MD, PhD; J.R. Landis, PhD; E. Mauger, PhD; J.R. Maurer, MD; Y. Phillips, MD; J.K. Stoller, MD; I. Tager, MD; A. Thomas Jr., MD.

Mortality and Morbidity Review Board: R.S. Crow, MD; T.E. Cuddy, MD; R.S. Fontana, MD; R.E. Hyatt, MD; C.T. Lambrew, MD; M. Lertzman, MD; B.A. Mason, MD; D.M. Mintzer, MD; R.B. Wray, MD.

National Heart, Lung, and Blood Institute, Bethesda, Maryland: T. Croxton, MD, PhD (Project Officer); J.P. Kiley, PhD (Division of Lung Diseases), G. Weinmann, MD (Airway Biology and Disease Program, Division of Lung Diseases), M.C. Wu, PhD (Division of Epidemiology and Clinical Applications).

Current Author Addresses: Dr. Anthonisen: University of Manitoba, Respiratory Hospital, 810 Sherbrook Street, Room 319, Winnipeg, Manitoba R3A 1R8, Canada.
Ms. Skeans and Dr. Connett: University of Minnesota, Coordinating Centers for Biometric Research, 2221 University Avenue SE, Room 200, Minneapolis, MN 55414-3080.
Dr. Wise: Johns Hopkins Asthma and Allergy Center, Division of Pulmonary and Critical Care Medicine, 5501 Hopkins Bayview Circle, Baltimore, MD 21224.
Dr. Manfreda: University of Manitoba, Respiratory Hospital RS 115, 810 Sherbrook Street, Winnipeg, Manitoba R3A 1R8, Canada.
Dr. Kanner: University of Utah Medical Center, Pulmonary Division, 701 Wintrobe Building, 50 North Medical Drive, Salt Lake City, UT 84132.

Author Contributions: Conception and design: N.R. Anthonisen, R.A. Wise, R.E. Kanner, J.E. Connett.
Analysis and interpretation of the data: N.R. Anthonisen, M.A. Skeans, R.A. Wise, J.E. Connett.
Drafting of the article: N.R. Anthonisen, M.A. Skeans, J. Manfreda, R.E. Kanner.
Critical revision of the article for important intellectual content: N.R. Anthonisen, M.A. Skeans, J. Manfreda, R.E. Kanner.
Final approval of the article: N.R. Anthonisen, M.A. Skeans, R.A. Wise, J. Manfreda, R.E. Kanner.
Provision of study materials or patients: R.A. Wise, R.E. Kanner.
Statistical expertise: M.A. Skeans, J.E. Connett.
Obtaining of funding: R.A. Wise, R.E. Kanner, J.E. Connett.
Collection and assembly of data: R.A. Wise, J.E. Connett.
A Randomized, Controlled Trial of Combination Therapy for Chronic Hepatitis B: Comparing Pegylated Interferon-α2b and Lamivudine with Lamivudine Alone

Henry Lik-Yuen Chan, MD; Nancy Wai-Yee Leung, MD; Alex Yui Hui, MB, BCH; Vincent Wai-Sun Wong, MBChB; Choong-Tsek Liew, MD; Angel Mei-Ling Chih, BS; Francis Ka-Leung Chan, MD; Lawrence Cheung-Tsui Hung, MB, BC; Yuk-Tong Lee, MD; John Siu-Lun Tam, PhD; Christopher Wai-Kei Lam, PhD; and Joseph Jao-Yiu Sung, MD, PhD

Background: Conventional interferon and lamivudine monotherapy are unsatisfactory in treating hepatitis B virus (HBV) infection.

Objective: To evaluate the efficacy and safety of pegylated interferon-α2b and lamivudine combination therapy for chronic hepatitis B.

Design: Randomized, controlled, open-label trial.

Setting: Outpatient clinic in a referral center.

Participants: 100 treatment-naive patients with hepatitis B e antigen (HBeAg)-positive chronic hepatitis B and moderately elevated alanine aminotransferase levels.

Measurement: The primary end point was sustained virologic response (HBeAg seroconversion and HBV DNA level < 500 000 copies/mL at 24 weeks after cessation of treatment.

Intervention: A staggered regimen of combination therapy with pegylated interferon-α2b (1.5 μg/kg of body weight per week; maximum, 100 μg) given for 32 weeks plus lamivudine (100 mg daily) given for 52 weeks versus lamivudine (100 mg daily) monotherapy given for 52 weeks. Of the 100 participants, 96% completed treatment and 80% completed post-treatment follow-up.

Results: The rate of sustained virologic response was 36% for the combination treatment group and 14% for the lamivudine monotherapy group (absolute difference, 22 percentage points [95% CI, 6 to 38 percentage points]). End-of-treatment outcomes showed that, compared with monotherapy, patients receiving combination therapy more often had virologic response (60% vs. 28%; absolute difference, 32 percentage points [CI, 14 to 50 percentage points]); had more substantial reductions of HBV DNA (3.91 log₁₀ copies/mL vs. 2.83 log₁₀ copies/mL); and less often had lamivudine-resistant mutants (21% vs. 40%). The percentages of patients with normalization of alanine aminotransferase levels and histologic improvement did not differ. Adverse effects, such as transient influenza-like symptoms, alopecia, and local erythematous reactions, were more common with combination therapy.

Limitations: This study lacked a double-blind design and was conducted at 1 institution. Because of the staggered pegylated interferon–lamivudine regimen, patients assigned to combination therapy received treatment for 8 weeks longer than those assigned to monotherapy.

Conclusions: In patients with HBeAg-positive chronic hepatitis B, staggered combination treatment with pegylated interferon-α2b and lamivudine may lead to a higher rate of virologic response than lamivudine monotherapy.


See also:

Print
Editors’ Notes .............................................. 241
Summary for Patients........................................ 1-30
Web-Only
Conversion of figures and tables into slides

Chronic hepatitis B virus (HBV) infection affects more than 300 million people globally (1). Patients who have HBV infection with positivity for hepatitis B e antigen (HBeAg) and persistently active disease have increased risks for progressive disease leading to liver cirrhosis and hepatocellular carcinoma (2). Conventional interferon treatment with injections given up to 3 times per week for 12 to 24 weeks may lead to HBeAg seroconversion in 33% of treated patients compared with 12% of untreated controls (3). The treatment response to conventional interferon treatment among Asian patients seems less satisfactory (4, 5). Pegylated interferon-α2b is synthesized by bonding recombinant interferon-α2b to polyethylene glycol and is given once weekly rather than 3 times weekly (6). The antiviral efficacy of pegylated interferon-α2b for treating chronic hepatitis C is superior to that of conventional interferon (7, 8), but few studies have evaluated pegylated interferon in patients with chronic hepatitis B.

Lamivudine is an oral nucleoside analogue that effectively suppresses HBV replication (9, 10). Studies suggest that only 16% to 18% of patients treated with lamivudine for 1 year develop HBeAg seroconversion (9–11). Extending lamivudine treatment for up to 4 years is associated with development of drug-resistant viral mutants in about 70% of patients (12). The durability of HBeAg seroconversion is estimated to be 46% to 64% up to 3 years after the cessation of lamivudine treatment (13–15).

Successful elimination of HBV depends on a durable immune clearance of the existing pool of intrahepatic HBV, particularly the closed covalent circular DNA (16). Combining an immunomodulator (such as interferon) and an antiviral agent (such as lamivudine) is an appealing ap-
The approach for treating chronic hepatitis B. However, past studies in patients with HBeAg-positive and HBeAg-negative chronic hepatitis B showed conflicting results about the superiority of combination therapy over lamivudine monotherapy (11, 17–20). We evaluated whether the combination of pegylated interferon-α2b and lamivudine improves antiviral efficacy and increases HBeAg seroconversion rates more than lamivudine monotherapy in patients with HBeAg-positive chronic hepatitis B and moderately elevated alanine aminotransferase (ALT) levels. Since extending interferon treatment from 16 to 32 weeks is associated with higher rates of HBeAg seroconversion (21), combination therapy includes pegylated interferon-α2b given for 32 weeks.

**Methods**

**Patients**

We recruited consecutive patients 18 to 65 years of age with chronic hepatitis B from the Hepatitis Clinic of the Prince of Wales Hospital, Hong Kong, China, a secondary referral center serving around 1 million people. All patients had been positive for hepatitis B surface antigen (HBsAg) for at least 6 months, were HBeAg-positive, and had a serum HBV DNA level of at least 500,000 copies/mL and an ALT level that was 1.3 to 5 times the upper limit of normal. We excluded patients who had decompensated liver disease or a history of interferon or antiviral agent use. Other exclusion criteria were co-infection with hepatitis C virus, hepatitis D virus, or HIV; history of hepatic cellular carcinoma; other causes of liver disease, including autoimmune hepatitis; Wilson disease; hemochromatosis and α1-antitrypsin deficiency; serious medical or psychiatric illness; concurrent use of corticosteroid or immunosuppressive agents; and pregnancy. We conducted the study in accordance with the guidelines of the Declaration of Helsinki. The ethics committee of The Chinese University of Hong Kong approved the protocol, and all patients gave witnessed, written informed consent.

**Study Design**

The study was a phase III, open-label, randomized trial. Within 4 weeks of screening for eligibility criteria, patients were randomly assigned to either combination therapy with pegylated interferon-α2b (PegIntron, Shering-Plough Corp., Kenilworth, New Jersey) and lamivudine (Zeffix, GlaxoSmithKline, Middlesex, United Kingdom) or lamivudine monotherapy in a ratio of 1:1. We based study group assignment on a computer-generated list, and research staff who were not involved in patient management placed the random numbers in opaque envelopes. A research nurse prescribed study drugs after receiving the information about treatment allocation at the baseline visit.

**Figure 1** shows the study design. Pegylated interferon-α2b was given as a subcutaneous injection at a dosage of 1.5 μg/kg of body weight per week for patients who weighed less than 65 kg or 100 μg per week for patients who weighed more than 65 kg for 32 weeks (6). Lamivudine was administered as 100 mg orally daily for 52 weeks in both groups of patients. In patients receiving combination therapy, pegylated interferon-α2b was administered 8 weeks before lamivudine was administered. Then both treatments were given in combination for 24 weeks, followed by lamivudine monotherapy for a further 28 weeks. Patients in the combination group were asked to return at weeks 1, 2, 4, 8, 12, 16, 20, 24, 32, 40, 52, and 60 (end of treatment). Patients in the lamivudine monotherapy group received lamivudine for 52 weeks and were asked to return for follow-up at weeks 4, 8, 12, 16, 24, 32, 40, 48, and 52 (end of treatment). We followed patients in both groups every 8 weeks in the post-treatment period for 24 more weeks. We gave open-label lamivudine treatment to patients who experienced severe post-treatment relapse of chronic hepatitis B (defined as an ALT level > 10 times the upper limit of normal and HBV DNA level > 500,000 copies/mL).

**Safety**

The investigators interviewed patients for symptomatic adverse effects and closely monitored laboratory tests at each follow-up visit. They recorded symptoms and events that patients reported spontaneously, symptoms and events elicited in response to open-ended questions, and adverse effects observed at the follow-up visits. They assessed all adverse events on the likelihood of causality by the study drug or drugs. They assessed the severity of adverse events according to a preset table and classified the event as mild (grade 1), moderate (grade 2), severe (grade 3), or life-threatening (grade 4).
The dosage of pegylated interferon was reduced, as necessary, according to the severity of the adverse events. The dosage was reduced from 100 μg per week (or 1.5 μg/kg per week if body weight < 65 kg) to 50 μg per week (or 1.0 μg/kg per week if body weight < 65 kg) for grade 3 adverse events (or sometimes grade 2 adverse events at the discretion of the investigator). The dosage could be further reduced to 25 μg per week (or 0.5 μg/kg per week if body weight < 65 kg) if the adverse effect persisted despite initial dosage reduction. Pegylated interferon therapy was stopped in case of grade 4 adverse events. Patients receiving combination treatment were allowed to continue lamivudine if investigators thought that the adverse effect was related to pegylated interferon use. Lamivudine therapy was stopped if the adverse event persisted despite cessation of pegylated interferon therapy.

We tested serum for HBV DNA levels, HBeAg, and antibody to HBeAg (at baseline, then 8 weekly until the end of treatment, and weeks 8, 16, and 24 after treatment) and HBsAg and antibody to HBsAg (at baseline, end of treatment, and 24 weeks after treatment). We determined the presence of lamivudine-resistant mutations in the serum sample at the end of treatment. Liver biopsy was performed within 4 weeks before treatment began and at the end of treatment.

Laboratory Assays

Virologic Assays

We based our sample size calculations and initial screening for eligibility on the DNA cross-linking assay (NAXCOR XLnt, NAXCOR, Menlo Park, California), which has a lower limit of detection of 500,000 copies/mL for quantification of HBV DNA (22). Since the TaqMan real-time polymerase chain reaction assay (Applied Biosystem, Foster City, California) became available in our laboratory, we used this assay to measure HBV DNA levels at baseline and in all follow-up visits (23, 24). The range of HBV DNA detection was from $10^2$ to $10^9$ copies/mL; the correlation coefficient of the standard curve was routinely greater than 0.990. We performed HBV genotyping by restriction fragment length polymorphism in a residual serum sample taken from each patient at their initial visit (25, 26). We determined the presence of lamivudine-resistant mutants by using the INNO-LiPA HBV DR line probe assay (Innogenetics N.V., Ghent, Belgium) according to the instruction of the manufacturer (27).

End Points

We defined virologic response as HBeAg seroconversion (that is, loss of HBeAg), detection of antibody to
Gave patients would be required to provide a power of 80% at an estimated response rates, we calculated that 94 patients in the lamivudine monotherapy group, 45% vs. 15%). Given sustained virologic response in combination groups vs. 30% higher than that of lamivudine (that is, 17), we anticipated that the use of pegylated interferon conventional interferon and lamivudine was up to 33% sustained virologic response for combination of treatments. We scored histologic necroinflammation by using the Knodell scoring system (score range, 0 to 10) and liver fibrosis by using the Ishak scoring system (score range, 0 to 6). The necroinflammation score was the sum of 3 histologic components: severity of periportal necrosis (score range, 0 to 10), intralobular necrosis (score range, 0 to 4), and portal inflammation (score range, 0 to 4). We considered a reduction of at least 2 points to be a clinically meaningful indicator of histologic changes. We regarded liver biopsy results with at least 3 portal tracts as sufficient for histologic scoring. One histopathologist, who was blinded to treatment assignments or the times at which the specimens were obtained, assessed all histologic specimens. The proportion of patients developing lamivudine-resistant mutant in the 2 treatment groups was also compared.

Statistical Analysis

On the basis of the literature, the proportion of patients receiving lamivudine monotherapy who achieve sustained virologic response was around 15%. Since the reported sustained virologic response for combination of conventional interferon and lamivudine was up to 33% (17), we anticipated that the use of pegylated interferon and lamivudine combination treatment had superior efficacy and that the rate of sustained virologic response was 30% higher than that of lamivudine monotherapy (that is, sustained virologic response in combination groups vs. lamivudine monotherapy group, 45% vs. 15%). Given these estimated response rates, we calculated that 94 patients would be required to provide a power of 80% at an α level of 0.05, allowing for a dropout rate of 10%

We used a modified intention-to-treat analysis to assess treatment responses. In the analysis, we included all randomly assigned patients who received at least 1 dose of the study medication. For the assessment of virologic or biochemical response, we considered patients who withdrew before the end of treatment (week 52 for monotherapy and week 60 for combination therapy) or for whom data were missing at the end of treatment to have failed response. For analyses of sustained response, we regarded patients who were classified as having a failed response at the end of treatment due to early withdrawal or missing data as having treatment failure. Among patients assigned to combination therapy who stopped pegylated interferon treatment prematurely, we assessed virologic and biochemical responses if lamivudine treatment was continued to the end of treatment. For the analysis of adverse events, we included all patients who were randomly assigned to either treatment group and who received at least 1 dose of study medication.

We performed all statistical tests by using SPSS, version 11.0 (SPSS, Inc., Chicago, Illinois). We expressed continuous variables as the median (range). Hepatitis B virus DNA was logarithmically transformed for analysis. We compared continuous variables, including patient age, liver biochemistry, log_{10} HBV DNA levels, and histologic scores, by using the Mann–Whitney U test. We compared categorical variables and the proportions of patients with virologic and biochemical responses, histologic improvements, lamivudine-resistant mutants, and adverse events by using the Pearson chi-square test or Fisher exact test, as appropriate. We compared the timing of HBeAg seroconversion by using Kaplan–Meier survival analysis. We used a logistic regression model to compare virologic response, with adjustment for baseline imbalance in ALT levels. A P value less than 0.05 was considered statistically significant. All statistical tests were 2-sided.

Role of Funding Source

Schering-Plough Corp. supplied pegylated interferon-α2b, and GlaxoSmithKline supplied lamivudine. The authors developed the study protocol and were responsible for data collection and progress of the study. The authors had full access to all of the study’s data files and were responsible for statistical analysis, reporting of data, and manuscript submission. Representatives of the pharmaceutical companies that supplied the drugs did not comment on the manuscript before submission.

RESULTS

Study Sample

We screened 182 patients between April 2000 and March 2002 and enrolled 100 eligible patients (Figure 2). Most baseline characteristics were well matched between the 2 groups (Table 1). However, although all studied patients met inclusion criteria at the screening visit, several patients had either normal ALT levels or levels that were greater than 5 times the upper limit of normal at the baseline randomization visit. The median ALT levels of the combination group were higher than those of the lamivudine group (P = 0.02), but the proportion of patients (P > 0.2) in the different ALT categories (1 to 2 times, 2 to 5 times, or >5 times the upper limit of normal) did not differ between groups.

Forty-eight patients (96%) in each treatment group completed treatment. No patients received any additional antiviral or immunomodulator treatment other than the study drugs. Also, no patients assigned to lamivudine monotherapy received pegylated interferon during the study.

Virologic Response

On intention-to-treat analysis, 30 of 50 patients (60%) in the combination treatment group and 14 of 50

HBeAg, and HBV DNA level less than 500 000 copies/mL and biochemical response as normalization of ALT level. We assessed both the virologic and biochemical end-of-treatment responses (week 52 for monotherapy and week 60 for combination therapy) and the sustained responses at 24 weeks after the end of treatment.

The primary end point was sustained virologic response. Secondary end points included sustained biochemical response, end-of-treatment virologic and biochemical responses, reduction in HBV DNA levels, and histologic improvement. We scored histologic necroinflammation by using the Knodell scoring system (score range, 0 to 10) and liver fibrosis by using the Ishak scoring system (score range, 0 to 6). The necroinflammation score was the sum of 3 histologic components: severity of periportal necrosis (score range, 0 to 10), intralobular necrosis (score range, 0 to 4), and portal inflammation (score range, 0 to 4). We considered a reduction of at least 2 points to be a clinically meaningful indicator of histologic changes. We regarded liver biopsy results with at least 3 portal tracts as sufficient for histologic scoring. One histopathologist, who was blinded to treatment assignments or the times at which the specimens were obtained, assessed all histologic specimens. The proportion of patients developing lamivudine-resistant mutant in the 2 treatment groups was also compared.

Statistical Analysis

On the basis of the literature, the proportion of patients receiving lamivudine monotherapy who achieve sustained virologic response was around 15%. Since the reported sustained virologic response for combination of conventional interferon and lamivudine was up to 33% (17), we anticipated that the use of pegylated interferon and lamivudine combination treatment had superior efficacy and that the rate of sustained virologic response was 30% higher than that of lamivudine monotherapy (that is, sustained virologic response in combination groups vs. lamivudine monotherapy group, 45% vs. 15%). Given these estimated response rates, we calculated that 94 patients would be required to provide a power of 80% at an α level of 0.05, allowing for a dropout rate of 10%

We used a modified intention-to-treat analysis to assess treatment responses. In the analysis, we included all randomly assigned patients who received at least 1 dose of the study medication. For the assessment of virologic or biochemical response, we considered patients who withdrew before the end of treatment (week 52 for monotherapy and week 60 for combination therapy) or for whom data were missing at the end of treatment to have failed response. For analyses of sustained response, we regarded patients who were classified as having a failed response at the end of treatment due to early withdrawal or missing data as having treatment failure. Among patients assigned to combination therapy who stopped pegylated interferon treatment prematurely, we assessed virologic and biochemical responses if lamivudine treatment was continued to the end of treatment. For the analysis of adverse events, we included all patients who were randomly assigned to either treatment group and who received at least 1 dose of study medication.

We performed all statistical tests by using SPSS, version 11.0 (SPSS, Inc., Chicago, Illinois). We expressed continuous variables as the median (range). Hepatitis B virus DNA was logarithmically transformed for analysis. We compared continuous variables, including patient age, liver biochemistry, log_{10} HBV DNA levels, and histologic scores, by using the Mann–Whitney U test. We compared categorical variables and the proportions of patients with virologic and biochemical responses, histologic improvements, lamivudine-resistant mutants, and adverse events by using the Pearson chi-square test or Fisher exact test, as appropriate. We compared the timing of HBeAg seroconversion by using Kaplan–Meier survival analysis. We used a logistic regression model to compare virologic response, with adjustment for baseline imbalance in ALT levels. A P value less than 0.05 was considered statistically significant. All statistical tests were 2-sided.

Role of Funding Source

Schering-Plough Corp. supplied pegylated interferon-α2b, and GlaxoSmithKline supplied lamivudine. The authors developed the study protocol and were responsible for data collection and progress of the study. The authors had full access to all of the study’s data files and were responsible for statistical analysis, reporting of data, and manuscript submission. Representatives of the pharmaceutical companies that supplied the drugs did not comment on the manuscript before submission.

RESULTS

Study Sample

We screened 182 patients between April 2000 and March 2002 and enrolled 100 eligible patients (Figure 2). Most baseline characteristics were well matched between the 2 groups (Table 1). However, although all studied patients met inclusion criteria at the screening visit, several patients had either normal ALT levels or levels that were greater than 5 times the upper limit of normal at the baseline randomization visit. The median ALT levels of the combination group were higher than those of the lamivudine group (P = 0.02), but the proportion of patients (P > 0.2) in the different ALT categories (1 to 2 times, 2 to 5 times, or >5 times the upper limit of normal) did not differ between groups.

Forty-eight patients (96%) in each treatment group completed treatment. No patients received any additional antiviral or immunomodulator treatment other than the study drugs. Also, no patients assigned to lamivudine monotherapy received pegylated interferon during the study.

Virologic Response

On intention-to-treat analysis, 30 of 50 patients (60%) in the combination treatment group and 14 of 50
patients (28%) in the lamivudine monotherapy group showed virologic response at the end of treatment (absolute difference, 32 percentage points [95% CI, 14 to 50 percentage points]; \(P = 0.001\)). After adjustment for baseline ALT levels, the absolute difference in predicted probabilities of combination treatment and lamivudine monotherapy for end-of-treatment response was 31 percentage points (CI, 10 to 49 percentage points; \(P = 0.003\)). All patients who had lost HBeAg developed antibodies to HBeAg. One patient receiving combination treatment had lost HBsAg by the end of treatment. No patient receiving lamivudine monotherapy became HBsAg-negative. Five patients receiving combination therapy and 2 patients receiving lamivudine monotherapy became HBV DNA–negative by polymerase chain reaction assay. Among patients with baseline ALT levels less than 5 times the upper limit of normal, the end-of-treatment virologic response of the combination treatment group was still statistically signifi-
treatment in both groups, 11 additional patients receiving combination treatment and 11 patients receiving lamivudine monotherapy had HBeAg seroconversion. More patients in the combination treatment group had HBeAg seroconversion in the last 28 weeks of extended lamivudine treatment (10 of 30 [33%] remaining patients) than in the lamivudine monotherapy group (3 of 39 [8%] remaining patients) (absolute difference, 25 percentage points [CI, 7 to 45 percentage points]). Among patients who developed HBeAg seroconversion during treatment, the median time for HBeAg seroconversion after commencement of treatment was 24 weeks (range, 8 to 60 weeks) in the combination treatment group and 24 weeks (range, 0 to 40 weeks) in the lamivudine monotherapy group (P = 0.13).

At 24 weeks after treatment, 18 patients (36%) in the combination treatment group and 7 patients (14%) in the lamivudine monotherapy group had sustained virologic response (absolute difference, 22 percentage points [CI, 6 to 38 percentage points]; P = 0.011). After adjustment for the baseline ALT levels, the absolute difference in predicted probabilities of combination treatment and lamivudine monotherapy for sustained virologic response was 22 percentage points (CI, 3 to 47 percentage points; P = 0.015). Among the subgroup of patients with baseline ALT levels less than 5 times the upper limit of normal, the predicted probability for sustained virologic response was still higher in the combination treatment group (15 of 42 patients [36% (CI, 21% to 50%)] than in the lamivudine monotherapy group (6 of 44 patients [14% (CI, 4% to 24%)]) (absolute difference, 22 percentage points [CI, 4 to 40 percentage points]). Twelve of 30 (40%) patients who received combination treatment and 7 of 14

### Table 1. Baseline Characteristics of the Studied Patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Combination Therapy Group (n = 50)</th>
<th>Lamivudine Group (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>32 (19–57)</td>
<td>34 (21–65)</td>
</tr>
<tr>
<td>Men, n (%)</td>
<td>31 (62)</td>
<td>36 (72)</td>
</tr>
<tr>
<td>Body weight, kg</td>
<td>64 (41–90)</td>
<td>68 (42–89)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>22 (16–33)</td>
<td>25 (18–32)</td>
</tr>
<tr>
<td>ALT level, n (%)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2 (4)</td>
<td></td>
</tr>
<tr>
<td>1–2 times the upper limit of normal</td>
<td>13 (26)</td>
<td>21 (42)</td>
</tr>
<tr>
<td>2–5 times the upper limit of normal</td>
<td>27 (54)</td>
<td>20 (40)</td>
</tr>
<tr>
<td>5 times the upper limit of normal</td>
<td>8 (16)</td>
<td>6 (12)</td>
</tr>
<tr>
<td>ALT level, U/L</td>
<td>144 (48–1179)</td>
<td>119 (36–461)</td>
</tr>
<tr>
<td>HBeAg-positive, %</td>
<td>100</td>
<td>98‡</td>
</tr>
<tr>
<td>HBV DNA level &gt; 500 000 copies/mL, %</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>HBV DNA level, log₁₀ copies/mL</td>
<td>8.04 (5.91–9.74)</td>
<td>7.67 (5.74–9.49)</td>
</tr>
<tr>
<td>HBV genotype, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>15 (30)</td>
<td>16 (32)</td>
</tr>
<tr>
<td>C</td>
<td>32 (64)</td>
<td>31 (64)</td>
</tr>
<tr>
<td>B and C</td>
<td>3 (6)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Histology§</td>
<td>Necroinflammation score 5 (1–11)</td>
<td>5 (1–12)</td>
</tr>
<tr>
<td>Fibrosis score</td>
<td>1 (0–6)</td>
<td>1 (0–5)</td>
</tr>
</tbody>
</table>

§ Continuous variables are shown as the median (range). Patients met entry criteria at screening but not at baseline (continued on study as per protocol). ALT = alanine aminotransferase; BMI = body mass index; HBeAg = hepatitis B e antigen; HBV = hepatitis B virus.
† Combination group: ALT level < 1.3 times upper limit of normal (n = 3 [6%]), ALT > 5 times upper limit of normal (n = 8 [16%]). Lamivudine group: ALT level < 1.3 times the upper limit of normal (n = 9 [18%]), ALT > 5 times the upper limit of normal (n = 6 [12%]).
Figure 3. Serial median log₁₀ hepatitis B virus (HBV) DNA reduction (top) and serial median alanine aminotransferase (ALT) levels (bottom) among patients who received pegylated interferon and lamivudine combination treatment (solid line) versus patients who received lamivudine monotherapy (dotted line) from baseline to 24 weeks after treatment.

All randomly assigned patients were included in the analysis. The numbers of patients receiving open-label lamivudine treatment for hepatitis B relapse after cessation of studied drugs are presented. One patient in the combination treatment group and 2 patients in the lamivudine monotherapy group developed severe reactivation of hepatitis while receiving open-label lamivudine treatment and did not attend the follow-up visit at week 84 and week 76, respectively.

Biochemical Response

Figure 3 shows the ALT levels for patients in the 2 treatment groups. At the end of treatment, 45 patients (90%) receiving combination treatment and 39 patients (78%) receiving lamivudine monotherapy had normalization of ALT levels (absolute difference, 12 percentage points [CI, −2 to 26 percentage points]). At 24 weeks after treatment, more patients receiving combination treatment had sustained ALT level normalization compared with those receiving lamivudine monotherapy (50% vs. 30% [absolute difference, 20 percentage points (CI, 1 to 39 percentage points)]). Among those with sustained virologic response, all except 1 patient who received combination treatment had sustained normalization of ALT levels (ALT level 1.1 times the upper limit of normal).

Histologic Response

Paired liver biopsy specimens were available in 40 patients receiving combination treatment and 44 patients receiving lamivudine monotherapy. Eight patients had insufficient liver tissue for accurate grading in 1 of the paired biopsy specimens (7 patients at baseline and 1 patient at week 52), 3 patients declined liver biopsy after treatment because of personal reasons, 4 patients withdrew prematurely from the study (1 patient also had insufficient liver tissue at baseline biopsy), and 2 post-treatment liver biopsies were cancelled because of the hospital outbreak of severe acute respiratory syndrome in March 2003. Among the 16 patients without evaluable paired liver biopsy specimens, 6 patients (4 receiving combination treatment and 2 receiving lamivudine monotherapy) had end-of-treat-
ment virologic response and 2 patients (both receiving combination treatment) had sustained virologic response.

Twenty-four (60%) patients receiving combination treatment and 26 (59%) patients receiving lamivudine monotherapy had at least a 2-point increase in necroinflammatory score (absolute difference, 1 percentage point [CI, −20 to 22 percentage points]). Four (10%) patients receiving combination treatment and 4 (9%) patients receiving lamivudine monotherapy had at least a 2-point decrease in fibrosis scores (absolute difference, 6 percentage points [CI, −8 to 20 percentage points]), while 4 (10%) and 2 (5%) patients, respectively, had at least a 2-point increase in fibrosis scores (absolute difference, 5 percentage points [CI, −6 to 17 percentage points]).

Drug-Resistant Mutants

At the end of treatment, a higher proportion of patients receiving lamivudine monotherapy developed a lamivudine-resistant mutant (19 of 48 [40%] patients) than did those receiving combination treatment (10 of 48 [21%] patients) (absolute difference, 19 percentage points [CI, 8 to 37 percentage points]). Among patients receiving monotherapy, 7 had a lamivudine-resistant mutant and 12 had both wide-type and lamivudine-resistant mutants. Among patients receiving combination therapy, 5 had a resistant mutant and 5 had mixed wide-type and resistant mutants. Patients who developed lamivudine resistance tended to have higher pretreatment HBV DNA (Table 2). A higher proportion of patients who developed lamivudine resistance (9 of 29 [31%] patients) had HBV DNA greater than 500 000 copies/mL at the end of treatment than those who did not develop lamivudine resistance (7 of 67 [10%] patients) (absolute difference, 21 percentage points [CI, 2 to 39 percentage points]). However, the development of lamivudine resistance did not seem to mitigate the virologic, biochemical, and histologic response (Table 2).

Table 2. Virologic, Biochemical, and Histologic Variables of Patients and the Development of Lamivudine-Resistant Mutation*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lamivudine-Resistant Mutant Present (n = 29)</th>
<th>Lamivudine-Resistant Mutant Absent (n = 67)</th>
<th>Adjusted P Value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of patients with genotype B to those with genotype C, n,n</td>
<td>12:17</td>
<td>19:42‡</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Baseline HBV DNA, log10 copies/mL</td>
<td>8.60 (6.46–9.77)</td>
<td>8.00 (5.74–10.20)</td>
<td>0.009</td>
</tr>
<tr>
<td>Baseline ALT level, U/L</td>
<td>135 (48–461)</td>
<td>126 (36–1179)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Baseline necroinflammatory score</td>
<td>5 (2–12)</td>
<td>5 (1–11)</td>
<td>0.096</td>
</tr>
<tr>
<td>Baseline fibrosis score</td>
<td>1 (0–5)</td>
<td>1 (0–6)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>End-of-treatment virologic response, n (%)</td>
<td>12 (41)</td>
<td>32 (48)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>End-of-treatment biochemical response, n (%)</td>
<td>24 (83)</td>
<td>60 (90)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Sustained virologic response, n (%)</td>
<td>8 (28)</td>
<td>17 (25)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Sustained biochemical response, n (%)</td>
<td>8 (28)</td>
<td>32 (48)</td>
<td>&gt;0.2</td>
</tr>
</tbody>
</table>

* Continuous variables are shown as the median (range).
† P value adjusted for treatment group by logistic regression model.
‡ 6 patients had mixed genotypes B and C.

Safety

Most adverse symptoms and events were transient and were related to the use of pegylated interferon-α2b (Table 3). Four (8%) patients receiving combination treatment had serious adverse events. One patient developed bipolar disorder requiring antidepressant therapy (week 21), 1 developed pulmonary tuberculosis with right pleural effusion requiring antituberculosis treatment (week 11), 1 developed thyrotoxicosis requiring propylthiouracil treatment (week 17), and 1 developed severe local reaction at the injection sites (week 8) that resolved spontaneously. Peglated interferon-α2b treatment was stopped in all 4 cases. Lamivudine treatment was continued in the first 3 cases until week 60, and we evaluated the treatment responses as per protocol. The fourth patient, who had received only 7 doses of pegylated interferon, withdrew from the study and was considered to have treatment failure. Five (10%) patients required reduction of dosage of pegylated interferon-α2b to 50 µg per week (if body weight > 65 kg) or 1.0 µg/kg per week (if body weight < 65 kg) as per protocol because of anemia (1 patient), neutropenia (3 patients), and/or thrombocytopenia (4 patients). One patient had pegylated interferon withheld for 2 doses at weeks 4 and 5 because of severe hepatitis flare-up (ALT level, 1762 U/L) and resumed at full dose at week 6 when ALT level decreased to a lower level (242 U/L). No patient who received lamivudine monotherapy developed serious adverse event during treatment, and no patient had dosage adjustment for lamivudine.

Five (10%) patients in the combination group and 11 (23%) patients in the lamivudine monotherapy group developed severe post-treatment relapse of chronic hepatitis B (absolute difference, −13 percentage points [CI, −27 to 2 percentage points]). Two patients in lamivudine monotherapy group had post-treatment relapse leading to elevation of serum bilirubin levels to 82 µmol/L (4.8 mg/dL) and 153 µmol/L (9.0 mg/dL), respectively. Lamivudine treatment was resumed, and all patients responded. No patient died or required liver transplantation.
DISCUSSION

The antiviral effect of lamivudine, a nucleoside analogue, for treating chronic HBV infection is suboptimal. The drug is associated with a low HBeAg seroconversion rate, frequent post-treatment relapses, and development of drug resistance with extended treatment. Combination therapy with lamivudine and other nucleoside analogues (for example, adefovir dipivoxil and telbivudine) may not improve virologic response (28, 29), and combination therapy with lamivudine and interferon shows conflicting results (Table 4). In our study, we administered a combination of pegylated interferon-α2b and lamivudine in a staggered manner. Our rationale for staggered treatment was that initial treatment with pegylated interferon-α2b probably enhances the immune clearance of intrahepatic HBV, including the closed covalent circular DNA, and that adding lamivudine at a later phase theoretically suppresses HBV replication and prevents reinfection of the hepatocytes. This particular hypothesis requires confirmation with studies of viral kinetics.

We found that patients with HBeAg-positive chronic hepatitis B and moderately elevated ALT levels had a higher rate of virologic response at the end of treatment with the staggered combination regimen (60%) than with monotherapy (28%). The rate of sustained HBeAg seroconversion 24 weeks after cessation of treatment was also higher with the staggered combination regimen (36%) than with monotherapy (14%). The end-of-treatment se-

Table 3. Common Adverse Events and Severe or Life-Threatening (Grade 3 or 4) Laboratory Toxicity during Treatment and at 24-Week Follow-up*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combination Therapy Group (n = 50)</th>
<th>Lamivudine Group (n = 50)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common adverse events, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper respiratory tract symptoms†</td>
<td>37 (74)</td>
<td>19 (38)</td>
<td>0.001</td>
</tr>
<tr>
<td>Fever</td>
<td>36 (72)</td>
<td>2 (4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Alopecia</td>
<td>24 (48)</td>
<td>2 (4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Abdominal discomfort</td>
<td>22 (44)</td>
<td>13 (26)</td>
<td>0.093</td>
</tr>
<tr>
<td>Malaise</td>
<td>22 (44)</td>
<td>7 (14)</td>
<td>0.002</td>
</tr>
<tr>
<td>Headache</td>
<td>21 (42)</td>
<td>2 (4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Myalgia</td>
<td>13 (26)</td>
<td>2 (4)</td>
<td>0.006</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>12 (24)</td>
<td>2 (4)</td>
<td>0.01</td>
</tr>
<tr>
<td>Reduced appetite</td>
<td>12 (24)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>Local erythematous reaction</td>
<td>12 (24)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>Allergic rashes</td>
<td>9 (18)</td>
<td>1 (2)</td>
<td>0.02</td>
</tr>
<tr>
<td>Dizziness</td>
<td>8 (16)</td>
<td>1 (2)</td>
<td>0.036</td>
</tr>
<tr>
<td>Vomiting or diarrhea</td>
<td>7 (14)</td>
<td>3 (6)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Weight loss (&gt;10%)</td>
<td>7 (14)</td>
<td>1 (2)</td>
<td>0.065</td>
</tr>
<tr>
<td>Laboratory toxicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased ALT level</td>
<td>8 (16)</td>
<td>12 (24)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Decreased phosphate level</td>
<td>2 (4)</td>
<td>1 (2)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Decreased neutrophil count</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Increased creatine kinase level</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Increased alkaline phosphatase level</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>&gt;0.2</td>
</tr>
</tbody>
</table>

* Common adverse events were defined as those occurring in 5% or more of the patients in either group. ALT = alanine aminotransferase.
† Upper respiratory tract symptoms included cough, running nose, and sore throat.

Table 4. Summary of Previous Randomized Trials on Interferon and Lamivudine Combination Treatment in Hepatitis B e Antigen–Positive Chronic Hepatitis B Virus

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Study Location</th>
<th>Combination Treatment</th>
<th>Comparison Treatment</th>
<th>Response in Combination Treatment Group, n/n (%)*</th>
<th>Response in Comparison Treatment Group, n/n (%)*</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hason et al., 2003 (32)</td>
<td>Middle East</td>
<td>Interferon-α2a × 16 wk + lamivudine × 48 wk†</td>
<td>Lamivudine × 48 wk</td>
<td>2/32 (6.2)</td>
<td>0/29 (0)</td>
<td>&gt;0.2</td>
</tr>
<tr>
<td>Yalcin et al., 2003 (33)</td>
<td>Middle East</td>
<td>Interferon-α2b + lamivudine × 12 mo</td>
<td>Interferon-α2b × 12 mo</td>
<td>15/33 (45)</td>
<td>3/16 (19)</td>
<td>0.13</td>
</tr>
<tr>
<td>Barbaro et al., 2001 (17)</td>
<td>Mediterranean</td>
<td>Interferon-α2b + lamivudine × 24 wk</td>
<td>Lamivudine × 52 wk</td>
<td>25/76 (33)</td>
<td>11/75 (15)</td>
<td>0.014</td>
</tr>
<tr>
<td>Schalm et al., 2000 (11)</td>
<td>North Europe</td>
<td>Interferon-α2b × 16 wk</td>
<td>Lamivudine × 24 wk‡</td>
<td>20/68 (29)</td>
<td>12/64 (19)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>North America</td>
<td></td>
<td></td>
<td>14/80 (18)§</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Response defined as sustained hepatitis B e antigen seroconversion and suppressed hepatitis B virus DNA/total patients.
† Lamivudine started 4 weeks after the commencement of interferon therapy.
‡ Interferon started 8 weeks after the commencement of lamivudine therapy.
§ Response assessed at the end of lamivudine treatment for lamivudine monotherapy group.
Pegylated Interferon and Lamivudine in Chronic Hepatitis B

Article

roconversion rate for the staggered combination regimen was better than that reported for other antiviral treatment regimens (9–11, 17, 28–30). Although a previous study had suggested that pegylated interferon might lead to higher rates of sustained seroconversion than conventional interferon (31), we found a rate of sustained post-treatment seroconversion similar to that reported in patients using either a conventional interferon and lamivudine combination (Table 4) or conventional monotherapy (34). Because these studies involved various patient populations with different HBV genotypes and used varying treatment regimens and definitions of virologic response, head-to-head comparisons are still needed to establish whether combination therapy with pegylated interferon and lamivudine leads to similar or higher rates of sustained response than either pegylated interferon alone or conventional interferon with or without lamivudine.

We also found a relatively high rate (40%) of post-treatment viral relapse among patients with initial end-of-treatment responses. The high relapse rate could be related to short durations of either pegylated interferon-α2b or lamivudine treatment. Extended lamivudine treatment after HBeAg seroconversion might improve the sustainability of virologic response (15, 35). However, despite suppression of HBV DNA in most patients and HBeAg seroconversion in almost half of the cases, we found lamivudine-resistant mutations in 21% of patients receiving combination treatment and 40% of patients receiving lamivudine monotherapy. The amount of lamivudine resistance that we found is higher than that reported in some series of Asian patients, perhaps because we used a highly sensitive line probe assay that could detect very low levels of mutants in a mixture with wild type (9–11, 27).

We found that patients in both treatment groups had biochemical and histologic improvements similar to those shown in previous studies with lamivudine treatment (9–11). We based histologic measurements on results of repeated liver biopsies that were performed in a sample of patients at the end of treatment. Histologic measurements taken several months after treatment might differ, particularly since patients in the lamivudine monotherapy group had higher post-treatment relapse rates than those in the combination therapy group.

Our trial has several limitations. First, we compared a staggered regimen of 60-week combination treatment with pegylated interferon and lamivudine versus 52-week treatment of lamivudine monotherapy. Under this study design, patients receiving combination treatment received a longer duration of antiviral treatment, which might affect the difference in clinical outcome. However, the antiviral effect of combination treatment at week 48 was still superior to that of lamivudine monotherapy and was similar to that of combination treatment at week 60. The extended therapy duration to 60 weeks did not seem to be a major factor affecting the virologic and clinical responses.

Second, since we did not include a pegylated interferon-α2b monotherapy group, we did not evaluate the additive benefit of lamivudine over pegylated interferon alone. The final results of a multicenter trial that compares the pegylated interferon and lamivudine combination with pegylated interferon monotherapy will shed light on this important issue (36). Third, the trial was not double-blinded since we did not use a placebo. However, no patients received antiviral or immunomodulator therapy other than the assigned drugs, and the end points were laboratory values that were probably not affected by patients’ perceptions. Fourth, patients assigned to combination treatment had higher median ALT levels at baseline than patients assigned to monotherapy. We thought it unlikely that these relatively small differences in ALT levels accounted for the differences between groups in clinical outcomes, particularly since we confirmed a beneficial effect of combination treatment among the subgroup of patients with baseline ALT levels less than 5 times the upper limit of normal. Fifth, although we evaluated virologic response 24 weeks after treatment, chronic hepatitis B is a persistent disease. Studies with even longer follow-ups are needed to assess clinical outcomes and the durability of response to treatment.

In conclusion, we found that, in patients with HBeAg-positive chronic hepatitis B and moderately elevated ALT levels, combination treatment with pegylated interferon-α2b and lamivudine was associated with higher rates of end-of-treatment and post-treatment HBeAg seroconversion, an increased potency of HBV suppression, and a lower incidence of lamivudine resistance than lamivudine monotherapy. We now need head-to-head comparisons to see whether this combination treatment leads to similar or higher rates of sustained response compared with either pegylated interferon alone or conventional interferon with or without lamivudine.

From the Chinese University of Hong Kong and Alice Ho Miu Ling Nethersole Hospital, Hong Kong, China.

Grant Support: Schering-Plough Corp. supplied pegylated interferon-α2b, and GlaxoSmithKline supplied lamivudine.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Joseph J.-Y. Sung, MD, PhD, Department of Medicine and Therapeutics, 9/F Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong SAR, China; e-mail, joesung@cuhk.edu.hk.

Current author addresses and author contributions are available at www.annals.org.

References

Pegylated Interferon and Lamivudine in Chronic Hepatitis B

[PMID: 8328741]


Current Authors Addresses: Drs. H.L.-Y. Chan, F.K.-L. Chan, Lee, and Sung; Mr. Hui; Mr. Wong; Ms. Chim; and Mr. Hung: Departments of Medicine and Therapeutics, 9/F Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong SAR, China. Dr. Leung: Department of Medicine, Alice Ho Miu Ling Nethersole Hospital, 11 Chuen On Road, Tai Po, Hong Kong SAR, Hong Kong. Dr. Liew: Department of Anatomical and Cellular Pathology, 1/F Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong SAR, China. Dr. Tam: Department of Microbiology, 1/F Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong SAR, China. Dr. Lam: Department of Chemical Pathology, 1/F Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, Hong Kong SAR, China.

A Cost-Effectiveness Analysis of Combination Antiplatelet Therapy for High-Risk Acute Coronary Syndromes: Clopidogrel plus Aspirin versus Aspirin Alone

Mark D. Schleinitz, MD, MS, and Paul A. Heidenreich, MD, MS

Background: Although clopidogrel plus aspirin is more effective than aspirin alone in preventing subsequent vascular events in patients with unstable angina, the cost-effectiveness of this combination has yet to be examined in this high-risk population.

Objective: To determine the cost-effectiveness of clopidogrel plus aspirin compared with aspirin alone.

Design: Cost–utility analysis.

Data Sources: Published literature.

Target Population: Patients with unstable angina and electrocardiographic changes or non–Q-wave myocardial infarction.

Time Horizon: Lifetime.

Perspective: Societal.

Interventions: Combination therapy with clopidogrel, 75 mg/d, plus aspirin, 325 mg/d, for 1 year, followed by aspirin monotherapy, was compared with lifelong aspirin, 325 mg/d.

Outcome Measures: Lifetime costs, life expectancy in quality-adjusted life-years (QALYs), and the incremental cost-effectiveness ratio.

Results of Base-Case Analysis: Patients treated with aspirin alone lived 9.51 QALYs after their initial event and incurred expenses of $127 700; the addition of clopidogrel increased life expectancy to 9.61 QALYs and costs to $129 300. The incremental cost-effectiveness ratio for clopidogrel plus aspirin compared with aspirin alone was $15 400 per QALY.

Results of Sensitivity Analyses: The analysis of 1 year of therapy was robust to all sensitivity analyses. In the probabilistic sensitivity analysis, fewer than 3% of simulations resulted in cost-effectiveness ratios over $50 000 per QALY. The cost-effectiveness of longer combination therapy depends critically on the balance of thrombotic event rates, durable efficacy, and the increased bleeding rate in patients taking clopidogrel.

Limitations: This analysis may not apply to patients with severe heart failure, those undergoing long-term anticoagulant therapy, those recently managed with revascularization, or those undergoing short-term treatment with glycoprotein IIb/IIIa inhibitors.

Conclusions: In patients with high-risk acute coronary syndromes, 1 year of therapy with clopidogrel plus aspirin results in greater life expectancy than aspirin alone, at a cost within the traditional limits of cost-effectiveness. The durable efficacy of clopidogrel relative to the risk for hemorrhage should be further explored before more protracted therapy can be recommended.


See also:

Print
Editors’ Notes ........................................... 252
Summary for Patients ............................... I-53

Web-Only
Conversion of figures and table into slides

© 2005 American College of Physicians
over time; and consideration of populations with varying risk.

Target Population

The target population was analogous to that of the CURE trial (1): patients with an acute coronary syndrome characterized by electrocardiographic changes or elevated serum cardiac markers in association with chest pain. Patients who had prolonged ST-segment elevation, who had undergone revascularization in the previous 3 months, who were at risk for severe bleeding or heart failure, or who had been treated with oral anticoagulants or glycoprotein IIb/IIIa inhibitors in the preceding 3 days were excluded. On the basis of the average age of CURE patients, we considered a 64-year-old as our base case and explored other ages in sensitivity analysis. We compared treatment costs and clinical outcomes for 2 types of antiplatelet therapy: 1) aspirin alone, 325 mg/d, and 2) clopidogrel, 75 mg/d, plus aspirin, 325 mg/d, for 1 year, followed by aspirin mono-therapy.

Model Structure

In our model, we included vascular events: myocardial infarction, stroke, vascular death and revascularization, intracerebral and gastrointestinal hemorrhagic events, and clopidogrel-associated thrombotic thrombocytopenic purpura. We included age-related mortality, correcting for events explicitly included in the model. We did not directly include procedures or outcomes, such as congestive heart failure, that CURE did not specifically address (1). We indirectly accounted for the cost of such events by using age-adjusted annual health care costs (7). This assumes equal probability of such events in each arm of the analysis. We allowed for multiple events, including multiple events of a particular type, within a given monthly cycle.

We created Markov states for conditions that changed quality of life, annual cost, or probability of future events. For events with temporary decrements in quality of life, we assessed a utility toll proportional to the duration of hospitalization required (8). We also modeled each combination of 2 events. When more than 2 events occurred, we

Figure 1. Schematic representation of the decision model.

The rectangular node depicts the treatment decision. Representative Markov states are outlined. The remaining 3 subtrees, 1 each for thrombotic events, hemorrhagic events, and side effect, make up the monthly cycle of the model. Multiple events, including multiple events within a subtree, are possible. At the completion of each cycle, patients return to the Markov state appropriate for both the Markov state in which they began the cycle and the events occurring during the cycle. Not shown is the risk for age-related mortality, present in each cycle.

—The Editors
used the Markov state that combined the 2 events with the lowest utility (Figure 1).

**Model Inputs**

**Probabilities**

We derived probabilities for vascular and hemorrhagic events over the first year of our analysis from the CURE trial (1). Rates of cardiovascular outcomes were highest in the month immediately following the acute coronary syndrome (1). To account for temporal variation in risk, we calculated probabilities for the first month separately from those for subsequent months. Beyond the time frame of the CURE trial, we empirically calculated declining probability functions for myocardial infarction, cardiovascular death, and revascularization on the basis of proportional decline in risk over time in survivors of non–Q-wave myocardial infarction (9).

We assumed that all bleeding events, other than intracerebral hemorrhage, were gastrointestinal. Because intracerebral hemorrhage was equally likely for the 2 treatment strategies (1), all of the excess bleeding in the clopidogrel plus aspirin arm was attributed to gastrointestinal causes. The rate of thrombotic thrombocytopenic purpura associated with clopidogrel was derived from an observational study (10, 11). We represented uncertainty in event rates through β distributions based on the number of events that occurred and the number of patient-years at risk. We obtained age-specific mortality from life tables (12) and corrected for death rates of events explicitly included in our model (Table).

**Efficacy of Combination Therapy**

Despite the variation in risk, the relative efficacy of clopidogrel plus aspirin did not vary between the first and subsequent months (1). In modeling beyond the time frame of the CURE trial, we assumed in our base case that efficacy remained constant, on the basis of data for clopidogrel monotherapy (13). We explored waning efficacy over time in sensitivity analysis. We used 3 variables to describe loss of efficacy, duration of constant efficacy (1 year to lifetime), duration of efficacy deterioration (1 to 10 years), and the extent of efficacy decline relative to aspirin monotherapy (none to complete loss).

**Costs**

We derived costs for each event and for chronic care of disabled patients from the literature. For medications, we used the average wholesale price in the United States as our base value (35) and considered prices negotiated by a large-volume purchaser in constructing the price range. Acute care costs for clinical events represent the direct costs of all medical care incurred during hospitalization. Chronic care costs represent direct expenditures for medications, procedures, and nursing care specific to the condition in question and were assessed for each month in the Markov state. For Markov states representing survival after either severe stroke or intracerebral hemorrhage and another event, we assumed 20% of the long-term cost of the additional condition to account for overlapping therapy. We accounted for other health care costs by using age-adjusted annual health expenditures (7). We updated costs to 2002 U.S. dollars with a gross domestic product deflator (44) and generated log-normal distributions for each estimate for use in sensitivity analyses (Table).

**Utilities**

We used published population-based utilities, representing either time-tradeoff or standard-gamble techniques. For Markov states representing 2 events, we combined utilities with multiplication (8). For events and procedures that did not result in durable changes in the health state of the individual, we used disutility tolls based on the average duration of hospitalization (8). We modeled uncertainty in utility estimates by calculating β distributions based on the range of utility estimates in the literature (Table).

**Sensitivity Analyses**

To assess the degree to which variation in any variable altered our results, we performed 1-way sensitivity analyses for each model input by analyzing the results at both extremes of its 95% CI. In evaluating patients of different ages, we adjusted both age-related mortality and estimates of annual health care cost accordingly. To better understand the distribution of the cost-effectiveness ratio for clopidogrel plus aspirin and the potential value of further research, we performed probabilistic sensitivity analysis using Monte Carlo simulation (45, 46). In each of 1000 simulations, the value for each model input was randomly selected from its distribution. We constructed a cost-effectiveness acceptability curve by calculating the average net monetary benefit for each strategy in each simulation over cost-effectiveness thresholds ranging from no additional expenditure for the least expensive therapy to $100,000 for each quality-adjusted life-year (QALY) gained. We then determined the proportion of simulations for which clopidogrel plus aspirin resulted in the greater net monetary benefit at each cost-effectiveness threshold. We also assessed varying the duration of clopidogrel therapy from 1 month to 1 year in monthly increments. We considered prolonged therapy, up to patient lifetime, in yearly increments, simultaneously assessing potential decline in the efficacy of clopidogrel.

To better characterize the role of risk in determining the cost-effectiveness of clopidogrel with aspirin, we conducted a 2-way sensitivity analysis on the annual probability of vascular events and the proportion of events that were cerebrovascular. We based this analysis on steady-state event rates from the CURE trial, excluding those events in the first month, and on the durable efficacy of clopidogrel. To determine the aggregate effect of adding clopidogrel to aspirin after an acute coronary syndrome, we...
Table. Model Inputs*

<table>
<thead>
<tr>
<th>Definition</th>
<th>Base-Case Value (Range)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event probabilities for patients treated with aspirin (first month)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal myocardial infarction</td>
<td>0.0079 (0.0059 to 0.014)</td>
<td>1</td>
</tr>
<tr>
<td>Nonfatal myocardial infarction</td>
<td>0.024 (0.020 to 0.028)</td>
<td>1</td>
</tr>
<tr>
<td>Revascularization</td>
<td>0.227 (0.217 to 0.237)</td>
<td>1</td>
</tr>
<tr>
<td>Vascular death</td>
<td>0.016 (0.013 to 0.019)</td>
<td>1</td>
</tr>
<tr>
<td>Bleeding</td>
<td>0.015 (0.012 to 0.018)</td>
<td>1</td>
</tr>
<tr>
<td>Event probabilities for patients treated with aspirin (months 2 to 12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal myocardial infarction</td>
<td>0.0011 (0.00081 to 0.0014)</td>
<td>1</td>
</tr>
<tr>
<td>Nonfatal myocardial infarction</td>
<td>0.0036 (0.0031 to 0.0041)</td>
<td>1</td>
</tr>
<tr>
<td>Revascularization</td>
<td>0.018 (0.017 to 0.019)</td>
<td>1</td>
</tr>
<tr>
<td>Vascular death</td>
<td>0.0023 (0.0019 to 0.0027)</td>
<td>1</td>
</tr>
<tr>
<td>Bleeding</td>
<td>0.011 (0.009 to 0.014)</td>
<td>1</td>
</tr>
<tr>
<td>Event probabilities for all patients treated with aspirin (monthly)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal stroke</td>
<td>0.00049 (0.00033 to 0.00069)</td>
<td>1</td>
</tr>
<tr>
<td>Nonfatal stroke</td>
<td>0.0010 (0.00079 to 0.0013)</td>
<td>1</td>
</tr>
<tr>
<td>Fatal intracerebral hemorrhage</td>
<td>0.0001 (0.00089 to 0.0018)</td>
<td>13</td>
</tr>
<tr>
<td>Nonfatal intracerebral hemorrhage</td>
<td>0.0001 (0.00094 to 0.0018)</td>
<td>13</td>
</tr>
<tr>
<td>Event probabilities for all patients treated with clopidogrel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal thrombotic thrombocytopenic purpura</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First month on therapy (monthly)</td>
<td>0.0000001</td>
<td>10, 11</td>
</tr>
<tr>
<td>After first month (annual)</td>
<td>0.0000001</td>
<td>10, 11</td>
</tr>
<tr>
<td>Nonfatal thrombotic thrombocytopenic purpura</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First month on therapy (monthly)</td>
<td>0.0000006</td>
<td>10, 11</td>
</tr>
<tr>
<td>After first month (annual)</td>
<td>0.0000007</td>
<td>10, 11</td>
</tr>
<tr>
<td>Efficacy of clopidogrel with aspirin (relative risk reduction), %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thrombosis prevention</td>
<td>20 (10 to 28)</td>
<td>1</td>
</tr>
<tr>
<td>Bleeding</td>
<td>~38 (~67 to 0)</td>
<td>1</td>
</tr>
<tr>
<td>Revascularization (first month)</td>
<td>9 (0 to 18)</td>
<td>1</td>
</tr>
<tr>
<td>Costs, $</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal stroke</td>
<td>16 061 (5000 to 39 000)</td>
<td>14, 15</td>
</tr>
<tr>
<td>Severe stroke</td>
<td>16 295 (5000 to 40 000)</td>
<td>14, 16–23</td>
</tr>
<tr>
<td>Moderate stroke</td>
<td>11 760 (2500 to 35 000)</td>
<td>14, 19, 21, 24</td>
</tr>
<tr>
<td>Mild stroke</td>
<td>5865 (1500 to 16 000)</td>
<td>14, 16, 19, 21–23, 25</td>
</tr>
<tr>
<td>Fatal myocardial infarction</td>
<td>19 689 (5000 to 54 000)</td>
<td>14, 16, 26</td>
</tr>
<tr>
<td>Nonfatal myocardial infarction</td>
<td>17 452 (5000 to 45 000)</td>
<td>14, 16, 26</td>
</tr>
<tr>
<td>Coronary angioplasty</td>
<td>12 485 (5000 to 26 000)</td>
<td>27–29</td>
</tr>
<tr>
<td>Coronary artery bypass surgery</td>
<td>28 100 (15 000 to 48 000)</td>
<td>27–30</td>
</tr>
<tr>
<td>Fatal intracerebral hemorrhage</td>
<td>21 358 (9000 to 43 000)</td>
<td>15</td>
</tr>
<tr>
<td>Nonfatal intracerebral hemorrhage</td>
<td>27 106 (7500 to 71 000)</td>
<td>15, 22, 31</td>
</tr>
<tr>
<td>Gastrointestinal bleeding</td>
<td>57 311 (1500 to 15 000)</td>
<td>14, 32</td>
</tr>
<tr>
<td>Thrombotic thrombocytopenic purpura</td>
<td>35 542 (5000 to 128 000)</td>
<td>33</td>
</tr>
<tr>
<td>Vascular death</td>
<td>7500 (1000 to 28 000)</td>
<td>Assumed</td>
</tr>
<tr>
<td>Other death</td>
<td>5000 (1000 to 15 000)</td>
<td>Assumed</td>
</tr>
<tr>
<td>Annual care after severe stroke</td>
<td>33 687 (10 000 to 84 000)</td>
<td>19, 21, 34</td>
</tr>
<tr>
<td>Annual care after moderate stroke</td>
<td>20 214 (5000 to 56 000)</td>
<td>19</td>
</tr>
<tr>
<td>Annual care after mild stroke</td>
<td>96 165 (0 to 36 000)</td>
<td>19</td>
</tr>
<tr>
<td>Annual care for coronary disease</td>
<td>1180 (0 to 7600)</td>
<td>16, 26</td>
</tr>
<tr>
<td>Annual care after intracerebral hemorrhage</td>
<td>18 543 (5000 to 49 000)</td>
<td>22, 31, 34</td>
</tr>
<tr>
<td>Annual care, age 50 y</td>
<td>2330 (1500 to 3100)</td>
<td>7</td>
</tr>
<tr>
<td>Annual care, age 65 y</td>
<td>3040 (2160 to 3920)</td>
<td>7</td>
</tr>
<tr>
<td>Annual care, age 80 y</td>
<td>3750 (2760 to 4740)</td>
<td>7</td>
</tr>
<tr>
<td>Daily cost of clopidogrel</td>
<td>3.80 (1.80 to 7.10)</td>
<td>35, VA pharmacy</td>
</tr>
<tr>
<td>Daily cost of aspirin</td>
<td>0.02 (0.01 to 0.04)</td>
<td>35, VA pharmacy</td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe stroke</td>
<td>0.11 (0 to 0.35)</td>
<td>36–40</td>
</tr>
<tr>
<td>Moderate stroke</td>
<td>0.39 (0.25 to 0.55)</td>
<td>36–38, 40</td>
</tr>
<tr>
<td>Mild stroke</td>
<td>0.76 (0.55 to 0.95)</td>
<td>36–38, 40, 41</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>0.87 (0.80 to 0.95)</td>
<td>42</td>
</tr>
<tr>
<td>Intracerebral hemorrhage</td>
<td>0.30 (0 to 0.60)</td>
<td>40, 43</td>
</tr>
<tr>
<td>Disutility tolls, QALY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal bleeding</td>
<td>0.005 (0 to 0.01)</td>
<td>Assumed</td>
</tr>
<tr>
<td>Thrombotic thrombocytopenic purpura</td>
<td>0.027 (0 to 0.055)</td>
<td>Assumed</td>
</tr>
<tr>
<td>Other input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discount rate, %</td>
<td>3 (0 to 5)</td>
<td>6</td>
</tr>
</tbody>
</table>

* Probabilities are monthly unless otherwise specified. When disabilities coexisted, we combined utilities by multiplication. Costs are expressed in 2002 U.S. dollars and represent one-time charges for acute events, annual costs of chronic care for the specified condition, and daily costs of clopidogrel and aspirin. QALY = quality-adjusted life-year; VA = Veterans Affairs.
first estimated the number of people in the United States who would meet the inclusion and exclusion criteria on an annual basis (47). We projected the lifetime mean health and economic outcomes for this cohort.

RESULTS

Base-Case Analysis

Patients treated with aspirin lived an average of 9.51 QALYs after their initial event, incurring costs of $127 700. Those treated with clopidogrel plus aspirin lived an average of 9.61 QALYs (that is, an additional 0.10 QALY), with a lifetime cost of $129 300, or $1600 more. The incremental cost-effectiveness ratio for clopidogrel plus aspirin relative to aspirin alone was $15 400 per QALY.

Sensitivity Analyses

One-Way Analyses

Our evaluation was robust to all one-way analyses. At the lowest efficacy estimate for clopidogrel plus aspirin, a 10% relative risk reduction, the incremental cost-effectiveness ratio was $45 900 per QALY. At a relative risk reduction of 28%, the cost-effectiveness ratio decreased to $9800 per QALY compared with aspirin alone. The cost of clopidogrel was less important. The average discounted lifetime cost for combination therapy increased by $1100 when clopidogrel cost $7.10 per day, resulting in an incremental cost-effectiveness ratio of $26 000 per QALY. At a daily cost of $1.80, this decreased to $8900 per QALY.

Variation in the probability of hemorrhage altered both the effectiveness, in quality-adjusted life expectancy, and the cost-effectiveness of adding clopidogrel to aspirin. The 38% increase in the relative risk for hemorrhage in the base case decreased average life expectancy by 0.032 QALY and decreased costs by $100. When the relative risk for hemorrhage was 67% greater with clopidogrel plus aspirin than with aspirin alone, the addition of clopidogrel to aspirin cost $18 800 for each QALY gained. When the increase in the bleeding risk was only 13%, the cost-effectiveness ratio decreased to $13 400 per QALY compared with aspirin.

Probabilistic Analysis

At a cost-effectiveness threshold of zero, the less expensive therapy was selected regardless of efficacy. Aspirin alone was less expensive in 99.7% of the simulations. As we increased the cost-effectiveness threshold, that is, society’s willingness to pay for improved quantity or quality of life, the combination of clopidogrel and aspirin became progressively more attractive. At a cost-effectiveness threshold of $14 600 per QALY, clopidogrel with aspirin was optimal in 50% of simulations. Combination therapy was optimal in 97.2% of simulations at a threshold of $50 000 per QALY (Figure 2).

Duration of Therapy

When clopidogrel was used only for the first month after the initial event, patients lived longer (9.55 vs. 9.51 QALYs) and incurred an additional $54 ($1200 per QALY) in lifetime costs compared with patients treated with aspirin alone. Each additional month of therapy added approximately 0.005 QALY in life expectancy, at an incremental cost of $140 ($26 100 per QALY) relative to the next shorter duration of therapy (Figure 3).

Treatment beyond 1 year was progressively less cost-effective. Five years of combination therapy yielded the greatest quality-adjusted life expectancy (9.67 QALYs), with the hemorrhagic risk of longer therapy outweighing the benefit. The marginal cost of the second year of therapy ($31 600 per QALY) was similar to that of the first, but the costs of the third, fourth, and fifth years of therapy ($61 300 per QALY, $136 500 per QALY, and $730 000 per QALY, respectively) were less financially attractive. The incremental cost-effectiveness ratio of the third year of therapy relative to the first 2 years was more than $100 000 per QALY in simulations in which the efficacy of clopidogrel decreased by more than 25% before the end of the third year.

Risk of the Population

Variation in the risk of the population, in conjunction with the proportion of events that were cerebrovascular, also determined the cost-effectiveness of adding clopidogrel to aspirin (Figure 4). In the base case, the cumulative annual probability of a vascular event was 10.1%. Strokes accounted for 18.2% of events. As the probability of any vascular event decreased, the proportion of stroke-related events required to maintain the same cost-effectiveness ratio increased. For populations with a low risk for vascular
events, particularly if the majority of that risk was attributable to coronary events, treatment with aspirin alone resulted in greater quality-adjusted life expectancy.

**Aggregate Results**

We estimated that 250,000 people in the United States would meet the inclusion and exclusion criteria of the CURE trial annually. Adding clopidogrel to aspirin therapy for 1 year in this cohort resulted in a societal gain of 25,500 QALYs at a cost of $392 million over the cohort’s lifetime. The first month of combination therapy provided 11,000 additional QALYs at a cost of $13.5 million. Each additional month added 1300 QALYs and cost $35 million.

**Figure 3.** Sensitivity analysis on duration of therapy.

![Graph showing cost-effectiveness ratios for different durations of combination therapy with clopidogrel plus aspirin followed by aspirin alone.](image)

Shown are the results of lifetime treatment with aspirin and the results of varying duration of combination therapy with clopidogrel plus aspirin followed by aspirin alone. Lines depict the incremental cost-effectiveness ratio for the next longer duration. Quality-adjusted life expectancy was greatest with 5 years of combination therapy. QALY = quality-adjusted life-year.

**Figure 4.** Two-way sensitivity analysis on annual event rate and the proportion of events attributable to cerebrovascular accidents.

![Graph showing sensitivity analysis on annual event rate and the proportion of events attributable to cerebrovascular accidents.](image)

The black circle indicates the base case. QALY = quality-adjusted life-year.
DISCUSSION

For patients with high-risk acute coronary syndromes, adding clopidogrel to aspirin for 1 month increased life expectancy by 0.04 QALY, more than 2 weeks of perfect health, at excellent value. One year of combination therapy added more than 1 month to quality-adjusted life expectancy and cost $15 400 for each QALY gained relative to aspirin, or $26 100 per QALY relative to 1 month of combination treatment. Our analysis was robust to all sensitivity analyses. In probabilistic sensitivity analysis, fewer than 3% of simulations favored aspirin monotherapy at a cost-effectiveness threshold of $50 000 per QALY, indicating that further trials of combination therapy for this indication and duration would provide little value (46). The incremental benefit and value of 1 year of combined therapy with clopidogrel and aspirin are almost identical to those of therapy with glycoprotein IIb/IIIa inhibitors for a similar population receiving a coronary stent (48) and represent better value than implantation of a cardioverter defibrillator in patients with ventricular arrhythmias (49).

Longer courses of combination treatment after an acute coronary syndrome, particularly beyond 2 years, are less likely to be cost-effective. This finding is critically dependent on 3 variables: the ongoing risk of the treated population, the durable efficacy of clopidogrel added to aspirin, and the bleeding risk conferred by adding clopidogrel. For persons who have an annual risk for a vascular event of less than 4% and a low risk for stroke, aspirin alone is more effective than combined clopidogrel and aspirin (Figure 3). For such patients, the benefit of fewer vascular events is outweighed by the increased risk for bleeding. The progressive decrease in risk explains why the cost-effectiveness of combination therapy wanes over time.

Although our analysis of 1 year of therapy was robust to variation in the efficacy estimate of clopidogrel plus aspirin, this was not true for longer treatment. Combination therapy must maintain a threshold of absolute risk reduction to overcome the risk for bleeding and the increased cost of adding clopidogrel to aspirin. Decreasing risk leads to smaller allowances in relative efficacy to maintain this threshold, meaning the precision of the efficacy estimate becomes more important as risk decreases.

Our analysis was less affected by the cost of clopidogrel. Even when clopidogrel was free, patients receiving 1 year of clopidogrel treatment incurred average lifetime costs that were $300 greater than those for patients treated with aspirin alone. These findings were due to the additional cost of medical care over a longer lifetime and to the cost of hemorrhages. Thus, medication cost was responsible for 80% of the incremental cost of combination therapy. Although increasing age commonly decreases the cost-effectiveness of preventive strategies, age was less important in identifying patients for whom the addition of clopidogrel would be effective. This is due to the immediate reduction in thrombotic event rates conferred by clopidogrel.

Our analysis shares the limitations of the CURE trial. Patients who were treated with glycoprotein IIb/IIIa inhibitors within the preceding 3 days or who had undergone revascularization within the preceding 3 months were excluded from CURE (1). In a randomized comparison of 1 year versus 1 month of combined clopidogrel and aspirin therapy in patients who underwent elective percutaneous coronary intervention (many of whom received glycoprotein IIb/IIIa inhibitors), relative and absolute risk reductions for cardiovascular outcomes, as well as risk for bleeding, were similar to those seen in the CURE trial (50). This suggests that combination therapy may be similarly cost-effective in such patients. Our analysis may not apply to patients receiving anticoagulants or those at risk for severe bleeding or heart failure, who were also excluded from CURE (1). We assumed that all excess bleeding events were gastrointestinal and were potentially more costly than procedural hemorrhage. Thus, the true incremental cost-effectiveness ratio for the addition of clopidogrel to aspirin may be lower than in our base case.

Our estimation of decline in the event rate over time is based on data from the Framingham Heart Study (9). Short-term prognosis for survivors of myocardial infarction has substantially improved over time, suggesting that we may have underestimated future events. This would bias our analysis against longer treatment with clopidogrel. It does not change the need for data on long-term therapy.

The applicability of this work to other groups of patients is not certain. Our results cannot be generalized to all patients with coronary disease. We and others have shown that clopidogrel, alone or in combination with aspirin, is less effective than aspirin or is financially unattractive in other groups with coronary disease (3, 51). Our current analysis not only confirms this but also suggests that the increased risk for hemorrhage relative to decreased thrombotic risk explains this finding. Our 2-way analysis of population risk and the proportion of events attributable to strokes suggests that the addition of clopidogrel to aspirin may be equally cost-effective in other high-risk populations, particularly patients with previous stroke. This extrapolation assumes that clopidogrel plus aspirin is equally effective in other populations, an assumption not supported in a trial of clopidogrel monotherapy (13). Furthermore, at least 2 strategies, clopidogrel monotherapy and aspirin plus dipyridamole, are more effective than, and cost-effective in comparison with, aspirin in patients with previous stroke (51, 52). This must be confirmed by current trials in patients with cerebrovascular disease (53) and must be compared with all available treatment strategies.

Because the decision about optimal therapy for acute coronary syndromes will involve many patients each year, the aggregate impact of the addition of clopidogrel may also be relevant. Assuming that our analysis applies to 250 000 patients per year in the United States, adding
clopidogrel to aspirin for 1 year increases lifetime costs by $392 million. The first month of therapy provides 43% of the potential gain (11,000 QALYs) but only 3.4% of the incremental costs ($13.5 million). Although prolonged therapy represents good value according to traditional limits of cost-effectiveness, the value of other potential medical or nonmedical uses of the $35 million required for each month of therapy must also be considered.

In patients with high-risk unstable angina or non-Q-wave myocardial infarction, the addition of 1 year of clopidogrel therapy to aspirin therapy increases quality-adjusted life expectancy by 4 weeks at a cost that is comparable to other accepted therapies. The balance between reduced risk for vascular events and increased risk for hemorrhage is critical in determining the relative efficacy, and ultimately the cost-effectiveness, of this strategy compared with aspirin therapy alone. Further assessment of the benefits and risks of combination therapy beyond 1 year is warranted before it can be recommended.

From Brown University and Rhode Island Hospital, Providence, Rhode Island; and Stanford University Medical Center and VA Palo Alto Healthcare System, Palo Alto, California.

Grant Support: Dr. Schleinitz was supported in part by a Building Interdisciplinary Research Careers in Women’s Health (BIRCHW) career development grant (HD43447) from the National Institutes of Health Office of Research on Women’s Health, administered through Women and Infants’ Hospital, Providence, Rhode Island. Dr. Heidenreich was supported by a Career Development Award from the Veterans Administration Health Services Research and Development Service (CD98326).

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Mark D. Schleinitz, MD, MS, Division of General Internal Medicine, Brown University and Rhode Island Hospital, 593 Eddy Street, MPB-1, Providence, RI 02903; e-mail, Mark_Schleinitz@Brown.edu.

Current author addresses and author contributions are available at www.annals.org.

References


Current Author Addresses: Dr. Schleinitz: Brown University and Rhode Island Hospital, Division of General Internal Medicine, 593 Eddy Street, MPB-1, Providence, RI 02903.
Dr. Heidenreich: VA Palo Alto Healthcare System, 3801 Miranda Avenue, Palo Alto, CA 94304.

Author Contributions: Conception and design: M.D. Schleinitz, P.A. Heidenreich.
Analysis and interpretation of the data: M.D. Schleinitz, P.A. Heidenreich.
Drafting of the article: M.D. Schleinitz.
Critical revision of the article for important intellectual content: M.D. Schleinitz, P.A. Heidenreich.
Final approval of the article: M.D. Schleinitz, P.A. Heidenreich.
Statistical expertise: M.D. Schleinitz.
Collection and assembly of data: M.D. Schleinitz.
Systematic Review: The Relationship between Clinical Experience and Quality of Health Care

Niteesh K. Choudhry, MD; Robert H. Fletcher, MD, MSc; and Stephen B. Soumerai, ScD

Background: Physicians with more experience are generally believed to have accumulated knowledge and skills during years in practice and therefore to deliver high-quality care. However, evidence suggests that there is an inverse relationship between the number of years that a physician has been in practice and the quality of care that the physician provides.

Purpose: To systematically review studies relating medical knowledge and health care quality to years in practice and physician age.

Data Sources: English-language articles in MEDLINE from 1966 to June 2004 and reference lists of retrieved articles.

Study Selection: Studies that provided empirical results about knowledge or a quality-of-care outcome and included years since graduation or physician age as explanatory variables.

Data Extraction: We categorized studies on the basis of the nature of the association between years in practice or age and performance.

Data Synthesis: Overall, 32 of the 62 (52%) evaluations reported decreasing performance with increasing years in practice for all outcomes assessed; 13 (21%) reported decreasing performance with increasing experience for some outcomes but no association for others; 2 (3%) reported that performance initially increased with increasing experience, peaked, and then decreased (concave relationship); 13 (21%) reported no association; 1 (2%) reported increasing performance with increasing years in practice for some outcomes but no association for others; and 1 (2%) reported increasing performance with increasing years in practice for all outcomes. Results did not change substantially when the analysis was restricted to studies that used the most objective outcome measures.

Limitations: Because of the lack of reliable search terms for physician experience, reports that provided relevant data may have been missed.

Conclusions: Physicians who have been in practice longer may be at risk for providing lower-quality care. Therefore, this subgroup of physicians may need quality improvement interventions.


For author affiliations, see end of text.
METHODS

We searched MEDLINE (Ovid Technologies, 1966 to June 2004; English language) for terms describing physician experience (keywords: physician age, clinician age, physician experience, clinician experience), physician demographic characteristics (keywords: physician characteristics, clinician characteristics), practice variation (subject heading: physician’s practice patterns), and performance in various domains (subject headings: clinical competence, health knowledge, attitudes and practice, outcomes assessment [health care]; keywords: knowledge, guideline adherence, appropriateness, outcomes). We retrieved potentially relevant articles and reviewed their reference lists to identify studies that our search strategy may have missed (Figure 1). We also searched our personal archives to identify additional studies. We included studies if they 1) were original reports providing empirical results; 2) measured knowledge, guideline adherence, mortality, or some other quality-of-care process or outcome; and 3) included years since graduation from medical school, years since certification, or physician age as a potential explanatory variable. We excluded studies if they described practice variation that is not known to affect quality of care (for example, assessed test-ordering behavior in clinical situations where optimal practice is unknown) or evaluated the performance of fewer than 20 physicians. For studies that examined several different endpoints, we included only those outcomes that are linked to knowledge or quality of care.

We used a standardized data extraction form to obtain data on study design and relevant results. We categorized studies into 4 groups on the basis of whether they evaluated knowledge (for example, knowledge of indications for blood transfusion), adherence to standards of care for diagnosis, screening, or prevention (for example, adherence to preventive care guidelines), adherence to standards of care for therapy (for example, appropriate prescribing), or health outcomes (for example, mortality). We classified the results of each study into 6 groups on the basis of the nature of the association between length of time in practice or age and performance: consistently negative, partially negative, no effect, mixed effect, partially positive, and consistently positive. “Consistently negative” studies were those for which all reported outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age. “Partially negative” studies showed decreasing performance with increasing experience for some outcomes and no association for others. We used similar definitions for “consistently positive” and “partially positive” studies. “Concave” studies found performance to initially improve with years in practice or age, then peak, and subsequently decrease.

We did not use formal meta-analytic techniques because the included studies used many different effect measures and some did not report parameter estimates. Since studies based on self-reported practice may suffer from social desirability bias (21), we explored the effect of study quality on results by subcategorizing studies according to whether they measured outcomes with self-reports (that is, using surveys and interviews) or observed practice (that is, using chart audits or administrative data review). We also compared studies according to whether they performed multivariable modeling to adjust for patient and physician covariates. We used the Fisher exact test to compare the observed frequencies. We conducted all analyses with SAS, version 8.2 (SAS Institute, Inc., Cary, North Carolina).

Role of the Funding Source

The Harvard Pilgrim Health Care Foundation supported this study. It had no role in the design, conduct, or reporting of the study or in the decision to submit the manuscript for publication.

RESULTS

Fifty-nine articles that reported data on 62 groups of relevant outcomes formed the basis of our analysis. Overall, 32 of the 62 evaluations (52%) demonstrated a negative association between increasing experience and performance (that is, performance decreased as experience increased) for all outcomes assessed; 13 (21%) reported a negative association for some outcomes but no association for other outcomes; 2 (3%) reported a concave relationship (that is, performance initially increased as experience increased, then peaked, and subsequently decreased); 13 (21%) reported no association; 1 (2%) reported a positive association (that is, performance increased as experience increased) for some outcomes but no association for other outcomes; and 1 (2%) reported a positive association for all outcomes assessed (Figure 2).
Knowledge

Twelve studies assessed the knowledge of practicing physicians, and all studies reported a negative association between knowledge and experience (Table 1). Studies by Ayanian and colleagues (7) and Salem-Schatz and colleagues (22) had large sample sizes, high response rates, and good sampling methods; used rigorous criteria to evaluate knowledge; and performed multivariate analysis.

Ayanian and colleagues (7) surveyed physicians to assess their beliefs about the survival benefit of therapies for acute myocardial infarction; the appropriate use of these therapies has been well-established in randomized, controlled trials. Specialists were more knowledgeable than generalists; however, after adjustment for this and other variables, physicians younger than 40 years of age were more likely to correctly believe in the value of therapies that improve survival (for example, thrombolytic agents, aspirin, and β-blockers). They were also significantly less likely to believe in the value of therapies that have been disproved (for example, prophylactic lidocaine) \((P < 0.05)\).

Salem-Schatz and colleagues (22) interviewed surgeons and anesthesiologists to assess their knowledge of the risks associated with and indications for the transfusion of blood products. They found a highly significant negative association between knowledge and the number of years the physicians had been in practice \((P < 0.001)\).

Adherence to Standards of Practice for Diagnosis, Screening, and Prevention

Twenty-four studies have assessed the appropriateness of physician use of diagnostic and screening tests, as well as preventive health care (Table 2). Overall, 15 (63%) of these studies demonstrated that physicians in practice for
more years were less likely to adhere to standards of practice in this domain.

In the largest of these studies, Czaja and colleagues (33) surveyed participants to assess their adherence to cancer screening guidelines endorsed by the American Cancer Society and the National Cancer Institute. Physicians who had graduated more than 20 years before the survey were consistently less likely to adhere to recommended practices (odds ratio, 0.62 to 0.72; \( P < 0.05 \)).

Using more objective measures of guideline adherence, Aubin and colleagues (17) assessed the practice of 21 physicians and found that after adjustment for patient and physician covariates, younger physicians were more likely to appropriately screen for hypertension (odds ratio, 1.11 [95% CI, 1.06 to 1.15]).

Several other studies provide contrary results. Streja and Rabkin (47) assessed the use of recommended preventive care measures and found that after adjustment for other physician covariates (such as specialty, practice style, and number of diabetic patients in their practice), older physicians were more likely than younger physicians to test for proteinuria (odds ratio, 2.62 [CI, 1.61 to 4.37]) and to refer their patients for screening ophthalmology assessments (odds ratio, 1.48 [CI, 1.01 to 2.18]). However, older physicians were no more likely to order a high-density lipoprotein cholesterol level test. Their analysis did not adjust for any patient variables, such as the presence of macrovascular and renal disease. Rhee (12) evaluated the performance of 454 physicians treating patients in 15 different medical and surgical diagnostic categories and found a concave relationship between years in practice and adherence to standards of practice. Physicians in practice for 6 to 15 years provided the most appropriate care, whereas physicians with more or fewer years of experience provided less appropriate care.

### Adherence to Standards of Appropriate Therapy

Table 3 presents the 19 studies that have assessed the influence of physician age and years in practice on adherence to standards of therapy. Of these studies, 14 (74%) found a partially or consistently negative association between physician age and adherence to standards of appropriate use of therapy.

A large and well-designed study by Beaulieu and colleagues (64) examined the prescribing behavior of physicians caring for patients with stable angina. After multivariate adjustment in a hierarchical model, older physicians were significantly less likely to prescribe aspirin (odds ratio for physicians in practice for \( > 20 \) years compared with those in practice \( < 10 \) years, 0.58). Age did not affect use of β-blockers or lipid-lowering agents.

### Outcomes

Seven studies present data on the relationship between number of years in practice and actual health outcomes (Table 4). The strongest of these was conducted by Nocinic and colleagues (14), who analyzed mortality for 39,007 hospitalized patients with acute myocardial infarction managed by 4546 cardiologists, internists, and family practitioners. After controlling for a patient’s probability of death, hospital location and practice environment, physician specialty, board certification, and the volume of patients seen, these researchers observed a 0.5% (SE, 0.27%) increase in mortality for every year since the treating physician had graduated from medical school.

Hartz and colleagues (11) specifically assessed the association between experience and mortality rates for sur-

---

*Figure 2. Distribution of study results relating physician age to clinical performance in various domains.*

| Studies in which length of time in practice or age was associated with lower performance for all outcomes. | Studies in which length of time in practice or age was associated with lower performance for some outcomes; no effect was found for other outcomes. | Studies in which there was a concave relationship between length of time in practice or age and performance. | Studies in which no association was found between length of time in practice or age and performance. | Studies in which length of time in practice or age was associated with higher performance for some outcomes; no effect was found for other outcomes. | Studies in which length of time in practice or age was associated with higher performance for all outcomes. |
Table 1. Studies Relating Length of Time in Practice or Physician Age to Knowledge*

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Knowledge Being Assessed</th>
<th>Persons Studied, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall Effect†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salem-Schatz et al., 1990 (22)</td>
<td>Blood product transfusion; orthopedic surgeons, and anesthesiologists</td>
<td>122 general surgeons,</td>
<td>Strong inverse association between years in practice and knowledge of transfusion risks and indications (P = 0.0001)</td>
<td>Yes</td>
<td>Knowledge assessment–based medical literature and NIH consensus conference; response rate, 91%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Golden et al., 2001 (23)</td>
<td>Emergency contraception</td>
<td>233 pediatricians</td>
<td>Younger physicians and more recent graduates (P = 0.02) were more likely to identify FDA-approved methods of emergency contraception (age categorized as &lt;40 y, 41–50 y, or &gt;50 y)</td>
<td>No</td>
<td>Response rate, 24%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Meskauskas and Webster, 1975 (24)</td>
<td>General medical knowledge</td>
<td>3356 interns certified ≥8 y earlier</td>
<td>Inverse relationship between age and ABIM recertification examination scores (age categorized as &lt;40 y, 40–44 y, 45–49 y, 50–54 y, 55–59 y, 60–64 y, or &gt;65 y)</td>
<td>No</td>
<td>Participants self-selected; tests of significance not presented</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Nocini et al., 1985 (25)</td>
<td>General medical knowledge</td>
<td>1947 interns</td>
<td>Inverse relationship between ABIM recertification examination scores and age (age categorized as &lt;40 y, 40–49 y, 50–59 y, or ≥60 y)</td>
<td>No</td>
<td>Participants volunteered; tests of significance not presented</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Ramsey et al., 1991 (26)</td>
<td>General medical knowledge</td>
<td>289 interns certified 5 to 15 y earlier</td>
<td>Significant inverse correlation (r = −0.3) between score on ABIM examination questions and years since certification</td>
<td>Yes</td>
<td>Participants partially self-selected, but sample was representative of population</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Cruft et al., 1981 (27)</td>
<td>General surgical knowledge</td>
<td>478 surgeons certified ≥7 y earlier</td>
<td>Inverse relationship between age and performance on American Board of Surgery recertification examination (age categorized as 40–45 y, 46–50 y, 51–55 y, 56–60 y, or 61–73 y)</td>
<td>No</td>
<td>Participants self-selected; tests of significance not presented</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Gemerson et al., 1991 (28)*</td>
<td>HIV</td>
<td>473 interns, family practitioners, general practitioners, and obstetrician–gynecologists</td>
<td>Younger physicians had significantly more knowledge about AIDS (P = 0.01)</td>
<td>Yes</td>
<td>Knowledge assessed by using questions from National Center for Health Statistics survey and others devised by investigators; response rate, 63%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Lewis et al., 1987 (29)*</td>
<td>HIV</td>
<td>1000 general practitioners, family physicians, and general internists</td>
<td>Younger physicians had greater AIDS-related knowledge (consistent across variables assessed, although P values not reported)</td>
<td>Unclear</td>
<td>Measures of “competence” used were defined by group of expert clinicians at UCLA; response rate, 60%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Shapiro, 1989 (30)</td>
<td>HIV</td>
<td>1271 general practitioners</td>
<td>Knowledge of HIV and AIDS decreased as years since graduation increased (P = 0.008)</td>
<td>Yes</td>
<td>Knowledge assessed with 6 questions designed by author; response rate, 70%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Evans et al., 1984 (31)</td>
<td>Hypertension</td>
<td>56 family physicians</td>
<td>Highly significant inverse correlation between test scores and years since graduation (r = −0.55; P &lt; 0.001)</td>
<td>No</td>
<td>Questionnaire validated to discriminate among physicians of different levels of training and specialty; response rate, 78%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Ayanian et al., 1994 (7)</td>
<td>MI</td>
<td>1211 cardiologists, internists, and family practitioners</td>
<td>Physicians &lt;40 y of age had greater knowledge of evidence-based therapies (P &lt; 0.05)</td>
<td>Yes</td>
<td>All physicians had served as the attending for at least 1 patient with MI within the preceding 3 mo; response rate, 61%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Schroen et al., 2000 (32)</td>
<td>Non–small-cell lung cancer</td>
<td>1010 pulmonologists and thoracic surgeons</td>
<td>Physicians trained before 1980 more likely to underestimate survival (P &lt; 0.001) and less likely to believe in value of chemotherapy in situations that have been well-established</td>
<td>No</td>
<td>Response rate, approximately 50%</td>
<td>Consistently negative</td>
</tr>
</tbody>
</table>

* ABIM = American Board of Internal Medicine; FDA = U.S. Food and Drug Administration; MI = myocardial infarction; NIH = National Institutes of Health; UCLA = University of California, Los Angeles.
† “Consistently negative” studies were those for which all outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age.
‡ “Partially negative” studies were those that showed statistically significant decreasing performance with increasing years in practice or age for some outcomes and no effect for others. Similar definitions were used for consistently positive and partially positive studies. “Concave studies” found performance to initially improve with years in practice or age then peak and subsequently decline.
‡ Also reported results on adherence to standards of diagnosis, screening, and preventive health care.


<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Disease or Condition</th>
<th>Physician Group Studied</th>
<th>Sample Size, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Patient Covariates?</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall Effect†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-reported practice (surveys or interviews)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czaja et al., 1994 (33)</td>
<td>Cancer screening</td>
<td>Family physicians, internists, general surgeons, and gynecologists</td>
<td>3436 physicians</td>
<td>Physicians who graduated &gt;20 y ago less likely to adhere to screening practices (OR, 0.62–0.72; ( P &lt; 0.05 ))</td>
<td>No</td>
<td>Yes</td>
<td>Only considered adherence with guidelines (results presented separately); response rate, 67%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Cook et al., 2001 (34)</td>
<td>Chlamydia screening</td>
<td>Family physicians, internists, gynecologists, and pediatricians</td>
<td>1600 physicians</td>
<td>No effect of length of time in practice and likelihood of screening ((&lt;10 \text{ y or } &gt;10 \text{ y in practice})</td>
<td>No</td>
<td>Yes</td>
<td>Guidelines established by CDC and USPSTF</td>
<td>No effect</td>
</tr>
<tr>
<td>Richards et al., 1998 (35)</td>
<td>Colon cancer screening for women</td>
<td>Primary care providers</td>
<td>508 physicians</td>
<td>Older physicians more likely to recommend screening contrary to national guidelines (OR, 3.42–3.79)</td>
<td>No</td>
<td>Yes</td>
<td>Response rate, 42%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Epstein et al., 2001 (15)‡</td>
<td>Depression</td>
<td>Psychiatrists</td>
<td>278 physicians</td>
<td>Physicians in practice for fewer years significantly more likely to correctly diagnose depression (OR, 0.59 [95% CI, 0.43–0.81], for a 10-y increase in age or practice)</td>
<td>No</td>
<td>Yes</td>
<td>Appropriateness defined by consensus agreement of 4 national experts; analyses adjusted for medical comorbidity but not severity of depression; response rate, 53%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Jacques et al., 1991 (36)</td>
<td>Diabetes</td>
<td>General practitioners, family physicians, and general internists</td>
<td>610 physicians</td>
<td>Physicians who had graduated more recently more likely to appropriately use glucose self-monitoring, hemoglobin A1c measurements, ophthalmology examinations (( P &lt; 0.001 )); no effect was observed for blood pressure and weight assessment, foot examination, glyemic education; year of graduation categorized in 4 groups</td>
<td>No</td>
<td>Yes</td>
<td>Guidelines established by American Diabetes Association; response rate, 31%</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Kenny et al., 1993 (37)</td>
<td>Diabetes</td>
<td>Primary care physicians</td>
<td>1434 physicians</td>
<td>Younger physicians more likely to adhere to preventive care guidelines for 6 of 8 procedures surveyed</td>
<td>No</td>
<td>Yes</td>
<td>Effect estimates not reported but results based on logistic regression; guidelines established by American Diabetes Association</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Marrero et al., 1991 (38)</td>
<td>Diabetes</td>
<td>Primary care physicians</td>
<td>212 physicians</td>
<td>Younger physicians more likely to obtain a hemoglobin A1c measurement (OR for every 10-y change in graduation date, 1.53; ( P = 0.0017 )); no relationship for use of glucose self-monitoring</td>
<td>No</td>
<td>No</td>
<td>Guidelines established by American Diabetes Association; response rate, 31%</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Schwartz et al., 1997 (39)</td>
<td>Disease prevention and health promotion</td>
<td>Members and Fellows of ACP</td>
<td>1349 physicians</td>
<td>Appropriate use of health promotion and disease prevention practices decreased with increasing age (( P ) value not presented)</td>
<td>No</td>
<td>Yes</td>
<td>Guidelines endorsed by several national agencies; response rate, 75%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Sherman and Honishman, 1993 (40)</td>
<td>Exercise counseling</td>
<td>Primary care physicians</td>
<td>422 physicians</td>
<td>Physicians &gt; 40 y of age more likely to counsel patients on exercise (OR, 3.08 [CI, 1.33–7.15])</td>
<td>No</td>
<td>Yes</td>
<td>Justification for exercise counseling based on research evidence; response rate, 61%</td>
<td>Consistently positive</td>
</tr>
<tr>
<td>Zerr et al., 1999 (41)</td>
<td>Fever in infants</td>
<td>Pediatricians and emergency department and family physicians</td>
<td>474 physicians</td>
<td>Physicians who graduated longer ago less likely to adhere to guidelines (OR, 0.93 [CI, 0.91–0.96], per year since graduation)</td>
<td>No</td>
<td>Yes</td>
<td>Physicians provided with guidelines; adherence assessed by using clinical scenarios that presented children of different ages; response rate, 36%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Gerson et al., 1991 (28)</td>
<td>HIV</td>
<td>Internists, family doctors, general practitioners, and obstetrician–gynecologists</td>
<td>473 physicians</td>
<td>Inverse relationship between years since graduation and adherence to New York State Department of Health AIDS prevention recommendations (( P &lt; 0.01 ))</td>
<td>No</td>
<td>Yes</td>
<td>Study also assessed knowledge (results presented separately); response rate, 63%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Heath et al., 1997 (42)</td>
<td>HIV</td>
<td>Family physicians and specialists who treat HIV</td>
<td>868 physicians</td>
<td>Inverse relationship between physician age and use of appropriate preventive care strategies (( P &lt; 0.001–0.004 ))</td>
<td>No</td>
<td>Yes</td>
<td>Response rate, 38.2%–50%; guidelines were issued by provincial agency</td>
<td>Consistently negative</td>
</tr>
</tbody>
</table>

Continued on following page
### Table 2—Continued

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Disease or Condition</th>
<th>Physician Group Studied</th>
<th>Sample Size, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Patient Covariates?</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall Effect†‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lewis et al., 1987 (29)</td>
<td>HIV</td>
<td>Primary care physicians</td>
<td>1000 physicians</td>
<td>Inverse relationship between years in practice and appropriateness of diagnostic work-up (consistent across variables assessed, although P values not reported)</td>
<td>No</td>
<td>Unclear</td>
<td>Measures of “competence” used were defined by group of expert clinicians at UCLA; study also assessed knowledge (results presented separately); response rate, 66%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Skotniski et al., 1996 (43)</td>
<td>HIV</td>
<td>Primary care physicians</td>
<td>480 physicians</td>
<td>No significant association between physician age and likelihood of testing a high-risk patient</td>
<td>No</td>
<td>Unclear</td>
<td>Response rate, 50%; older physicians were more likely to test any patient who asked to be tested (not entirely in keeping with guidelines but unclear)</td>
<td>No effect</td>
</tr>
<tr>
<td>Roetzheim et al., 1991 (16)</td>
<td>Mammography</td>
<td>Primary care physicians</td>
<td>565 physicians</td>
<td>Physicians &lt; 50 y of age were more likely than older physicians (72% vs. 49%; P &lt; 0.001) to fully adhere to American Cancer Society recommendations</td>
<td>No</td>
<td>No</td>
<td>Response rate, 42%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Ely et al., 1998 (44)</td>
<td>Preventive care guidelines</td>
<td>Primary care physicians</td>
<td>146 physicians</td>
<td>Physician age or year of graduation not associated with preventive care practices</td>
<td>No</td>
<td>Yes</td>
<td>Appropriateness defined by recommendations from the USPSTF; response rate, 70%</td>
<td>No effect</td>
</tr>
<tr>
<td>Rattay et al., 2004 (45)</td>
<td>Weight counseling</td>
<td>Pediatricians</td>
<td>813 physicians</td>
<td>Physician age not associated with frequency of weight counseling</td>
<td>No</td>
<td>Yes</td>
<td>Age categorized as &lt;45 y or &gt;45 y</td>
<td>No effect</td>
</tr>
<tr>
<td><strong>Observed practice (chart audit)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ford et al., 1987 (46)</td>
<td>Breast, rectal, and small-cell lung cancer</td>
<td>Physicians in community hospital oncology programs</td>
<td>Not reported (2052 patients)</td>
<td>Physicians with fewer years in practice more likely to adhere to guidelines for breast and rectal cancer staging and consultation (P &lt; 0.01–P &lt; 0.001); no age effect was observed for small-cell lung cancer</td>
<td>No</td>
<td>No</td>
<td>Effect of age persisted regardless of how involved physicians were in guideline creation process</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Streja and Rabkin, 1999 (47)</td>
<td>Diabetes</td>
<td>Primary care physicians</td>
<td>22 physicians (519 patients)</td>
<td>Physicians with &gt;15 y experience more likely to test for proteinuria (OR, 2.62 [CI, 1.61–4.37]) and refer for ophthalmology (OR, 1.48 [CI, 1.01–2.18]) but not more likely to order an HDL cholesterol test (OR, 1.04 [CI, 0.97–1.06])</td>
<td>No</td>
<td>Yes</td>
<td>Did examine effect of patient characteristics on appropriate screening, but did not enter these variables into the same model as physician characteristics</td>
<td>Partially positive</td>
</tr>
<tr>
<td>Anis et al., 2004 (48)</td>
<td>Dietary and exercise counseling</td>
<td>Primary care physicians</td>
<td>38 physicians (4344 patients)</td>
<td>No effect of length of time in practice and likelihood of counseling</td>
<td>Yes</td>
<td>Partial</td>
<td>Physician covariates not significant on univariate analysis and not included in multivariate analysis</td>
<td>No effect</td>
</tr>
<tr>
<td>Aubin et al., 1994 (17)</td>
<td>Hypertension</td>
<td>Family physicians</td>
<td>21 physicians (847 patients)</td>
<td>Younger physicians more likely to appropriately screen for hypertension (OR, 1.11 [CI, 1.06–1.15])</td>
<td>Yes</td>
<td>Yes</td>
<td>Adjusted for patient age, sex, number of visits, type of visit, but not patient comorbidity; did not specify threshold for older vs. younger</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Hulka et al., 1976 (49)</td>
<td>Several (4 conditions)</td>
<td>Family physicians, internists, gynecologists, and pediatricians</td>
<td>61 physicians (1258 patients)</td>
<td>Physicians in practice for fewer years more likely to appropriately manage infants (P &lt; 0.01). No difference observed for managing pregnancy, diabetes, or heart failure</td>
<td>Yes</td>
<td>Yes</td>
<td>Performance scores were developed on the basis of consensus panel discussions, all involving at least 4 family physicians as well as other physicians</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Rhee, 1976 (12)</td>
<td>Several (15 diagnostic categories)</td>
<td>Physicians in Hawaii</td>
<td>454 physicians (2571 patient discharges)</td>
<td>Inverted “V” relationship between years in practice and adherence with standards of practice (P = 0.01)—physicians in practice 6–15 y provided the most appropriate care; physicians with more or fewer years in practice provided less appropriate care</td>
<td>No</td>
<td>Yes</td>
<td>Performance scores were developed on the basis of norms established by a “panel of physicians”; criteria not fully presented but seem to focus largely on diagnostic evaluation</td>
<td>Concave</td>
</tr>
<tr>
<td>Saraiya et al., 2002 (50)</td>
<td>Tuberculosis screening for foreign-born persons</td>
<td>Physicians who performed screening</td>
<td>491 physicians (5739 patients)</td>
<td>No consistent effect of number of years in practice on adherence with CDC screening recommendations</td>
<td>No</td>
<td>No</td>
<td>Did not have demographic data on 30% of physicians; 75% of physicians were primary care providers</td>
<td>No effect</td>
</tr>
</tbody>
</table>

* ACP = American College of Physicians; CDC = Centers for Disease Control and Prevention; HDL = high-density lipoprotein; OR = odds ratio; USPSTF = U.S. Preventive Services Task Force.

† “Consistently negative” studies were those for which all outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age. “Partially negative” studies were those that showed statistically significant decreasing performance with increasing years in practice or age for some outcomes and no effect for others. Similar definitions were used for consistently positive and partially positive studies. “Concave studies” found performance to initially improve with years in practice or age then peak and subsequently decline.

‡ Also reported results on adherence to standards of therapy—results presented separately.
found that physicians who have been in practice longer had higher operative mortality rates ($P < 0.001$). In contrast, Burns and Wholey’s (69) large study of patients hospitalized for various conditions found no difference in mortality rates between experienced and newer physicians.

### Table 3. Studies Relating Length of Time in Practice or Physician Age to Adherence to Standards of Appropriate Therapy*

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Disease or Condition</th>
<th>Physician Group Studied</th>
<th>Sample Size, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Patient Covariates?</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall Effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported practice (surveys or interviews)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McFall et al., 1994 (51)</td>
<td>Breast cancer</td>
<td>Family physicians, internists, gynecologists, and general surgeons</td>
<td>1460 physicians</td>
<td>Physicians in practice for ≥20 y chose therapy less consistent with NIH recommendations for 3 of 6 treatments studied (OR, 0.56–0.78); no age effect was observed for other therapies</td>
<td>No</td>
<td>Yes</td>
<td>Physicians were not oncologists but reported participating in decision making about treatment and referral; response rate, 71%</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Epstein et al., 2001 (15)</td>
<td>Depression</td>
<td>Psychiatrists</td>
<td>278 physicians</td>
<td>Physician age did not influence appropriate prescribing of an antidepressant</td>
<td>Limited</td>
<td>Yes</td>
<td>Appropriateness defined by consensus of 4 national experts; analyses adjusted for medical comorbidity but not severity of depression; response rate, 53%</td>
<td>No effect</td>
</tr>
<tr>
<td>Epstein et al., 1996 (20)</td>
<td>Depression and anxiety</td>
<td>Psychiatrists</td>
<td>38 physicians</td>
<td>“Accuracy score” (reflecting agreement with expert consensus) decreased as a function of years in practice ($P &lt; 0.001$)</td>
<td>No</td>
<td>Yes</td>
<td>Appropriateness established by expert consensus; response rate, 19%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Montaner et al., 1996 (52)</td>
<td>HIV</td>
<td>Physicians in British Columbia</td>
<td>463 physicians</td>
<td>Physicians &lt; 45 y of age significantly more likely to appropriately use antiretroviral therapy ($P = 0.004$); no age effect observed for other management areas</td>
<td>No</td>
<td>Yes</td>
<td>Appropriateness assessed by using provincial guidelines; overall response rate, 14%</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Roy-Byrne et al., 2002 (53)</td>
<td>Panic disorder</td>
<td>Primary care physicians</td>
<td>37 physicians (38 patients)</td>
<td>Length of time in practice did not predict appropriate prescribing</td>
<td>Yes</td>
<td>Yes</td>
<td>Based on patient reports (all enrolled in clinical trial); appropriateness based on previously published algorithm</td>
<td>No effect</td>
</tr>
<tr>
<td>Observing practice (chart audit or administrative data review)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stolley et al., 1972 (13)</td>
<td>Prescribing of 5 specific drugs (ritalin, equagesic, chloromycetin, vitamin B12, and oral contraceptives)</td>
<td>Primary care physicians</td>
<td>37 physicians</td>
<td>Appropriateness decreased as years in practice increased ($P &lt; 0.01$)</td>
<td>No</td>
<td>No</td>
<td>Appropriateness assessed by at least 13 experts who rated any given drug, and the total panel consisted of 33 individuals; response rate, 84%</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Hynes, 1994 (54)</td>
<td>Breast cancer</td>
<td>Physicians treating breast cancer</td>
<td>Not reported (3972 patients)</td>
<td>Appropriateness of surgical care increased with increasing years in practice but decreased after 14 y of experience ($P &lt; 0.01$); physicians in practice for more years were less likely to provide postmastectomy rehabilitation therapy</td>
<td>Yes</td>
<td>Yes</td>
<td>Concave</td>
<td></td>
</tr>
<tr>
<td>Becker et al., 1971 (55)</td>
<td>Chloramphenicol use</td>
<td>Primary care physicians</td>
<td>37 physicians</td>
<td>Likelihood of prescribing chloramphenicol increased as years since graduation increased ($P &lt; 0.01$)</td>
<td>No</td>
<td>Yes</td>
<td>All prescriptions of chloramphenicol were judged to be inappropriate given limited indications</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Ray et al., 1976 (56)</td>
<td>Chloramphenicol use</td>
<td>Physicians in Tennessee caring for Medicaid patients</td>
<td>3409 physicians</td>
<td>Number of years since medical school graduation did not predict use of chloramphenicol</td>
<td>No</td>
<td>Yes</td>
<td>All prescriptions of chloramphenicol were judged to be inappropriate given limited indications</td>
<td>No effect</td>
</tr>
<tr>
<td>Moride et al., 2002 (57)</td>
<td>Depression</td>
<td>General practitioners and psychiatrists</td>
<td>1527 physicians</td>
<td>Graduation from medical school before 1970 associated with increased odds of suboptimal treatment duration (OR, 1.12 [95% CI, 1.01–1.24])</td>
<td>Yes</td>
<td>Yes</td>
<td>Patient covariates include age, sex, and health status but not illness severity</td>
<td>Consistently negative</td>
</tr>
</tbody>
</table>

Continued on following page
“Consistently negative” studies were those for which all outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age.

“Partially negative” studies were those that showed statistically significant decreasing performance with increasing years in practice or age for some outcomes and no effect for others. Similar definitions were used for consistently positive and partially positive studies. “Concave studies” found performance to initially improve with years in practice and subsequently decline.

### Table 3—Continued

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Disease or Condition</th>
<th>Physician Group Studied</th>
<th>Sample Size, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Patient Covariates?</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1997 (19)</td>
<td>Drug prescribing in elderly patients</td>
<td>Physicians in British Columbia</td>
<td>6344 physicians (819 269 drug claims)</td>
<td>Physicians &lt; 45 y of age had significantly lower rates of inappropriate drug selection for all 4 classes of drugs studied (P &lt; 0.001, for most analyses)</td>
<td>No</td>
<td>Yes</td>
<td>Used the same criteria as Beers et al. (58) to define appropriateness</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Beers et al., 1993 (58)</td>
<td>Drug prescribing in elderly patients</td>
<td>Physicians practicing in nursing homes in Los Angeles</td>
<td>309 physicians</td>
<td>Physician age or years since graduation not significantly related to appropriate prescribing</td>
<td>No</td>
<td>Yes</td>
<td>Appropriateness defined on the basis of consensus of expert opinion</td>
<td>No effect</td>
</tr>
<tr>
<td>Dhalla et al., 2002 (59)</td>
<td>Drug prescribing in elderly patients</td>
<td>Physicians in Ontario</td>
<td>2424 physicians (19 911 patients)</td>
<td>Patients of physicians &gt; 50 y of age had a higher odds of receiving an inappropriate medication (OR, 1.14 [CI, 1.05–1.23]; P = 0.002)</td>
<td>Limited</td>
<td>Yes</td>
<td>Adjusted for patient age and sex only; used the same criteria as Beers et al. (58) to define appropriateness</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Geller et al., 1996 (60)</td>
<td>Hysterectomy</td>
<td>Physicians performing hysterectomy</td>
<td>339 physicians (36 104 patients)</td>
<td>Physicians in practice for 15–19 y and 25–29 y perform more hysterectomies than physicians in practice for 0–4 y (P &lt; 0.05); no effect observed for physicians of other ages</td>
<td>Yes</td>
<td>Yes</td>
<td>Also controlled for sociodemographic and financial patient factors in addition to clinical covariates</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Payne et al., 1984 (61)</td>
<td>Several (10 conditions)</td>
<td>Physicians in the Midwest</td>
<td>1135 physicians (3163 patients)</td>
<td>Physicians in practice for 0–9 y provided more appropriate care than other physicians</td>
<td>No</td>
<td>Yes</td>
<td>Tests of significance not presented; criteria for appropriateness defined by consensus; no difference between physicians with 10–19 y and &gt;20 y of experience</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Sanazaro and Worth, 1985 (18)</td>
<td>Several (10 conditions)</td>
<td>Internists</td>
<td>66 physicians</td>
<td>Number of cases treated inappropriately increased with number of years since graduation (P &lt; 0.05)</td>
<td>No</td>
<td>No</td>
<td>Appropriateness judged by panel appointed by ACP and ASIM; participants were all volunteers</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Fehrenbach et al., 2001 (62)</td>
<td>Myocardial infarction</td>
<td>Physicians treating acute myocardial infarction</td>
<td>473 physicians (578 patients)</td>
<td>Physicians trained before 1980 less likely to prescribe β-blockers (P &lt; 0.05); in multivariate adjustment, OR of receiving β-blocker for patients of physicians trained before 1980 was 0.66 (CI, 0.40–1.03)</td>
<td>Yes</td>
<td>Yes</td>
<td>Borderline-significant results after multivariate adjustment; patients all belonged to 1 large national HMO</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Willson et al., 2000 (63)</td>
<td>Myocardial infarction</td>
<td>Physicians treating acute myocardial infarction</td>
<td>1452 physicians</td>
<td>Physicians &gt; 50 y of age less likely to prescribe aspirin to eligible patients (P &lt; 0.001); relationship did not persist after multivariate adjustment; no effect observed for thrombolysis</td>
<td>Yes</td>
<td>Yes</td>
<td>Also adjusted for hospital volume</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Beaulieu et al., 2001 (64)</td>
<td>Stable angina</td>
<td>Physicians in Quebec</td>
<td>3293 physicians (11 141 patients)</td>
<td>Older physicians significantly less likely to prescribe aspirin (OR for &lt; 10 y in practice, 1.7 compared with physicians in practice &gt; 20 y; P &lt; 0.05); no effect seen for β-blockers or lipid-lowering agents</td>
<td>Yes</td>
<td>Yes</td>
<td>Prescribed on study results, we stratified the 43 reports pertaining to adherence to standards of practice on the basis of whether outcomes were assessed by using self-reported data or more objective measures (that is, use of chart audits or administrative databases). Overall, 30 (70%) of these studies demonstrated a consistently or partially negative association between length of time in practice or physician age and adherence to standards of care. While the proportion of studies that found a consistently or partially negative association was slightly larger for self-reported studies than for</td>
<td></td>
</tr>
</tbody>
</table>

*ACP = American College of Physicians; ASIM = American Society of Internal Medicine; HMO = health maintenance organization; NIH = National Institutes of Health; OR = odds ratio.

† “Consistently negative” studies were those for which all outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age. “Partially negative” studies were those that showed statistically significant decreasing performance with increasing years in practice or age for some outcomes and no effect for others. Similar definitions were used for consistently positive and partially positive studies. “Concave studies” found performance to initially improve with years in practice or age then peak and subsequently decline.
Table 4. Studies Relating Number of Years in Practice or Physician Age to Health Care Outcomes

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Disease or Condition</th>
<th>Physician Group Studied</th>
<th>Sample Size, n</th>
<th>Results</th>
<th>Multivariate Adjustment for Patient Covariates?</th>
<th>Multivariate Adjustment for Physician Covariates?</th>
<th>Other Comments</th>
<th>Overall Effect*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norcini et al., 2000 (14)</td>
<td>Acute myocardial infarction</td>
<td>Family practitioners, internists, and cardiologists</td>
<td>4546 physicians</td>
<td>Mortality for patients admitted with acute myocardial infarction increased by 0.5% for every year since physician graduated from medical school (P = 0.05)</td>
<td>Yes</td>
<td>Yes</td>
<td>Also corrected for hospital factors (e.g., access to advanced cardiac care)</td>
<td>Consistently negative</td>
</tr>
<tr>
<td>Blanc et al., 2003 (65)</td>
<td>Asthma</td>
<td>Physicians treating asthma</td>
<td>147 physicians (317 patients)</td>
<td>Number of years since medical school graduation not related to patients' self-reported health status or asthma-specific quality of life</td>
<td>Yes</td>
<td>Yes</td>
<td>Surveyed patients and physicians separately and then linked results using hierarchical regression models; sample included very few younger physicians</td>
<td>No effect</td>
</tr>
<tr>
<td>O’Neill et al., 2000 (66)</td>
<td>Carotid endarterectomy</td>
<td>Surgeons</td>
<td>507 physicians (12,725 patients)</td>
<td>Mortality of patients undergoing endarterectomy increased with years since licensure (P &lt; 0.001); no relationship between length of time in practice and combined “bad outcome” (death or morbidity)</td>
<td>Yes</td>
<td>Yes</td>
<td>Data for surgeon age available for 440 physicians; years since licensure was strongest predictor of mortality</td>
<td>Partially negative</td>
</tr>
<tr>
<td>Hartz et al., 1999 (11)</td>
<td>Coronary bypass surgery</td>
<td>Surgeons</td>
<td>275 physicians (83,547 patients)</td>
<td>More years in practice significantly associated with higher mortality ratios (C = 0.22; P &lt; 0.001)</td>
<td>Yes</td>
<td>Yes</td>
<td>Consistently negative</td>
<td></td>
</tr>
<tr>
<td>Katon et al., 2000 (67)</td>
<td>Depression</td>
<td>Primary care physicians</td>
<td>63 physicians (1559 patients)</td>
<td>Physician age not related to patients having persistent or residual depressive symptoms after initiation of antidepressant medication or other quality-of-care measures</td>
<td>Yes</td>
<td>Yes</td>
<td>Study did not find any significant physician characteristics to explain variability in quality of care</td>
<td>No effect</td>
</tr>
<tr>
<td>Davidson et al., 1995 (68)</td>
<td>Drug prescribing in elderly patients</td>
<td>General practitioners</td>
<td>366 physicians</td>
<td>No age effect on mortality or hip fracture rate</td>
<td>Limited</td>
<td>No</td>
<td>Adjusted for patient age by using analysis of covariance</td>
<td>No effect</td>
</tr>
<tr>
<td>Burns and Wholey, 1991 (69)</td>
<td>Several medical and surgical conditions</td>
<td>Attending physicians for hospitalized patients</td>
<td>54,571 discharges</td>
<td>Physicians in practice for more years had significantly longer lengths of stay after adjustment for patient comorbid condition for 9 of 16 diagnoses evaluated (P &lt; 0.05); years in practice did not predict length of stay for other diagnoses or for mortality</td>
<td>Yes</td>
<td>Yes</td>
<td>Results adjusted for multiple covariates, suggesting increased length of stay may be unnecessary</td>
<td>Partially negative</td>
</tr>
</tbody>
</table>

* "Consistently negative" studies were those for which all outcomes demonstrated a statistically significant decrease in performance with increasing years in practice or age. "Partially negative" studies were those that showed statistically significant decreasing performance with increasing years in practice or age for some outcomes and no effect for others. Similar definitions were used for consistently positive and partially positive studies. “Concave studies” found performance to initially improve with years in practice or age then peak and subsequently decline.

those studies that used objective measures (71% vs. 62%), these differences were not statistically significant (P > 0.2).

Stratifying studies on the basis of whether they performed a multivariable analysis yielded similar results: 71% of the studies that adjusted for patient covariates found a consistently or partially negative association compared with 74% of studies that did not adjust for these factors, and 68% of the studies that adjusted for physician covariates found a consistently or partially negative association compared with 67% of studies that did not.

**Discussion**

Although based on heterogeneous studies, our systematic review of empirical studies evaluating the relationship between clinical experience and performance suggests that physicians who have been in practice for more years and older physicians possess less factual knowledge, are less likely to adhere to appropriate standards of care, and may also have poorer patient outcomes. These effects seem to persist in those studies that adjusted for other known predictors of quality, such as patient comorbidity and physician volume or specialization. The results are somewhat paradoxical since it is generally assumed that clinical experience enhances knowledge and skill and, therefore, leads to better patient care.

Our findings have many possible explanations. Perhaps most plausible is that physicians’ “toolkits” are created during training and may not be updated regularly (70). Older physicians seem less likely to adopt newly proven therapies (71, 72) and may be less receptive to new standards of care (73). In addition, practice innovations that involve theoretical shifts, such as the use of less aggressive
surgical therapy for early-stage breast cancer or protocols for reducing length of stay, may be harder to incorporate into the practice of physicians who have trained long ago than innovations that add a procedure or technique consistent with a physician’s preexisting knowledge (74).

Our findings may also reflect the substantial environmental changes that have occurred in medicine over the past several decades; evidence-based medicine has been widely adopted, and quality assurance techniques, such as disease management and performance evaluation, are frequently used. More experienced physicians may have less familiarity with these strategies and may be less accepting of them. Given this, our results may represent a cohort effect; that is, when the current generation of more recently trained physicians has been in practice for a longer time, there may be smaller differences between their practice and those of their younger colleagues than our data would suggest.

Our study has several limitations. First, although we attempted to systematically review the literature on the association between number of years in practice or physician age and performance, our search strategy may have missed reports. This reflects the limited attention to this issue and the lack of consistent search terms to identify clinical experience. In addition, studies that were specifically designed to assess the relationship between experience and performance but found no association may have been less likely to be submitted or accepted for publication, and published studies that included number of years in practice or age among other physician characteristics may not have presented non–statistically significant results for these particular variables. Therefore, while we have no reason to suspect that we were more likely to identify studies showing decreasing performance with age, our findings are still potentially subject to an under-reporting bias.

Second, few reports included in this review were designed to specifically evaluate length of time in practice as their primary characteristic of interest. Consequently, our results may have been due to chance arising from multiple testing. However, we believe this is unlikely given the relative consistency of the results in several different domains, their “dose–response” relationship, and their overall plausibility. Moreover, restricting our analysis to the 32 studies that considered a broader set of physician characteristics, including number of years in practice or age as the focus of their investigation (that is, excluding those studies that considered physician age or number of years in practice only as confounders), does not change our results: 21 of the 32 (66%) studies reported a consistently or partially negative association between physician age and performance, whereas only 1 study demonstrated a partially positive association.

Third, disagreements may exist between clinical practice guidelines (33), and, thus, establishing appropriate norms may be difficult. As a result, assessing performance on the basis of guideline adherence may not reliably assess health care quality. Despite this, some studies included in our review used norms that had been adopted by several professional associations and that consequently reflect widely accepted standards of practice. Even for these studies, we observed age effects.

Finally, length of time in practice may be associated with other dimensions of quality that are not captured by the outcome measures that we evaluated. While we identified studies that assessed various conditions and aspects of performance, the relationship between age and performance may be different for other diseases and outcomes. For example, older physicians may be more effective at delivering the humanistic, rather than the technical, aspects of medical care. If this were true, one would expect that the patients of older physicians would report higher satisfaction, which has been demonstrated in some studies (75, 76) but not others (77, 78). Alternatively, physicians who have been in practice for a longer time may have better clinical judgment and may thus provide better care in complex cases or may be better diagnosticians. These outcomes have not been rigorously assessed.

Despite these limitations, our results are troubling. Although it is difficult to draw firm conclusions about the performance of older physicians in managing specific conditions or clinical scenarios, our results do suggest that older physicians may need quality improvement interventions that are generally applicable to all physicians. In addition, the requirements that are imposed on physicians to keep up to date and to demonstrate continuing competence should be further considered. Widely adopted continuing medical education techniques, such as the distribution of printed materials and lectures, are largely ineffective even in experimental conditions (79). Our results reinforce this. Moreover, many experienced physicians are exempt from the recertification requirements to which their more recently trained colleagues must adhere. For example, the American Board of Internal Medicine only requires physicians who received initial Board certification in or after 1990 to appear for periodic recertification examinations.

Our results also have implications for further research. The link between experience and performance should be further evaluated with studies that are designed a priori to specifically measure this association. These studies should use objective and widely accepted measures of performance; should be disease- or process-specific; and should be replicated for physicians of different specialties, demographic characteristics (such as sex), and different environment practices. The effect of age for physicians who routinely collaborate with other physicians, who frequently engage in evidence-based discussions, or whose practices are influenced by disease management, performance feedback, and computerized reminder systems may be different from that for physicians who practice in relative isolation or in more traditional settings.

An optimal study would follow a particular cohort of physicians over time. However, this is not practical and...
may be confounded by other secular trends in health care provision. Alternative designs would be similar to those of the highest quality included in our review and would adequately control for patient comorbidity, other physician factors, and the clustering of patients within physicians. These studies should also model the nature of the relationship between experience and performance since performance may improve during the initial phases of independent practice, plateau for some period of time, and then decrease. Finally, the ability of behavior change strategies to reduce the disparities in quality created by physician age should be evaluated in well-controlled clinical trials.

In summary, our results suggest that physicians with more experience may paradoxically be at risk for providing lower-quality care. The extent, magnitude, and nature of these results must be clarified, and added attention should be given to this subgroup of physicians who may need quality improvement interventions.

From Harvard Medical School and Harvard Pilgrim Health Care, Brigham and Women’s Hospital, and the Harvard PhD Program in Health Policy, Boston, Massachusetts.

Grant Support: By the Harvard Pilgrim Health Care Foundation. Dr. Choudhry is also supported by the Harvard Medical School Fellowship in Pharmaceutical Policy Research, a Frank Knox Scholarship from Harvard University, and a Canadian Institutes of Health Research Postdoctoral Fellowship. Dr. Soumerai is an investigator in the HMO Research Network Center for Education and Research in Therapeutics, funded by the U.S. Agency for Healthcare Research and Quality.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Niteesh K. Choudhry, MD, Department of Ambulatory Care and Prevention, Harvard Medical School, 133 Brookline Avenue, 6th Floor, Boston, MA 02215; e-mail, choudhry@fas.harvard.edu.

Current author addresses are available at www.annals.org.

References
Clinical Experience and Quality of Health Care

Volume 142  Number 4


www.annals.org
have higher thresholds for recommending coronary arteriography than family physicians? Health Serv Res. 1987;22:623-35. [PMID: 3692862]


Current Author Addresses: Drs. Choudhry, Fletcher, and Soumerai: Department of Ambulatory Care and Prevention, Harvard Medical School, 133 Brookline Avenue, 6th Floor, Boston, MA 02215.
An Evidence-Based Guide to Writing Grant Proposals for Clinical Research

Sharon K. Inouye, MD, MPH, and David A. Fiellin, MD

The competition for funds to conduct clinical research is intense, and only a minority of grant proposals receive funding. In particular, funding for patient-oriented research lags behind that allocated for basic science research. Grant writing is a skill of fundamental importance to the clinical researcher, and conducting high-quality clinical research requires funds received through successful grant proposals. This article provides recommendations for the grant-writing process for clinical researchers. On the basis of observations from a National Institutes of Health study section, we describe types and sources of grant funds, provide key recommendations regarding the process of grant writing, and highlight the sections of grants that are frequently scrutinized and critiqued. We also provide specific recommendations to help grant writers improve the quality of areas commonly cited as deficient. Application of this systematic approach will make the task more manageable for anyone who writes grants.


For author affiliations, see end of text

Background Information on Grants

Types of Grants

In general, grants are of 2 major types: project and career development grants. Project grants generally support a specific research project and usually include only a small proportion of the principal investigator’s salary, typically about 20% to 25% (but sometimes as much as 35% to 40%). These can be small grants for pilot work or preliminary studies, or larger grants, such as for investigator-initiated projects. Career development grants (9) generally provide mostly salary support (75% to 100%) and relatively little project support. Examples of these include NIH K awards and foundation-based career development awards.

Funding Sources

Many sources are available to help with locating information on grants for clinical research (Table 1). One of the most useful is the NIH Office of Extramural Research home page, which outlines all NIH grant mechanisms (http://grants.nih.gov/grants/oer.htm). The NIH
Guide Archive (http://grants.nih.gov/grants/guide/index.html) provides NIH program announcements and requests for applications. Researchers can establish customizable searches for grant information in specific areas of interest on the Web sites of the Community of Science (http://fundingopps.cos.com/), Grantsnet (http://www.grantsnet.org/), or the Illinois Researcher Information Service (IRIS) (www.library.uiuc.edu/iris/). Information on foundation grants can be found on the Web site of the Foundation Center (http://fdncenter.org/), an independent nonprofit organization that provides comprehensive, up-to-date information on foundations and corporate giving programs. Other valuable sources of grant information are institutional grants or development offices, research mentors and other experienced investigators, and foundation files or resource libraries.

Understand the Reviewers’ Perspective

The reviewers are probably successful, busy clinical researchers who will be reviewing the grant in time “borrowed” from other activities. Assume that the reviewers are intelligent, are savvy about research, have a broad fund of scientific knowledge, but perhaps have little in-depth experience in your area of interest. While the reviewers are probably committed to being thorough and fair, it is important to remember that they will often not be in your exact field and will be reviewing your grant in a few hours at the most. Given these circumstances, the urgency of focus, conciseness, conceptual clarity, and transparent language becomes apparent. The grant writer must help the busy reviewer understand the project by making its significance crystal clear, avoiding jargon and topic-specific abbreviations or terminology, and not expecting reviewers to search references. The proposal should be completely self-contained. The quality of the review, however, may vary depending on the experience and skill of your assigned reviewers in research, mentoring, and the grant-review process. Examining the NIH Review Criteria can be helpful (10, 11). See Appendix 1 (available at www.annals.org) for more information on the NIH review process.

Seek Guidance from the Program Officer or Grants Administrator

Contact the program officer or grants administrator (if available) for the grant before and during the grant-writing process as questions arise. Their guidance can be invaluable in this process. However, their encouragement does not represent endorsement by the review committee.

Review Successful Grant Applications

If successfully funded grant applications are available for the particular funding mechanism, these can serve as useful models for the application. For NIH grants, information on funded grants (including principal investigator and abstract) can be obtained from the CRISP (Computer Retrieval of Information on Scientific Projects) database (http://crisp.cit.nih.gov/). Knowing what studies are being conducted can help you identify unanswered questions, avoid duplication, and gauge the priorities of the funding agency. Obtain a full grant proposal by contacting the principal investigator directly or the NIH under the Freedom of Information Act (FOIA). For information, contact the NIH Freedom of Information Office Coordinator for the appropriate NIH Institute (www.nih.gov/icd/od/foia/coord.htm) or the NIH Freedom of Information Office (301-496-5633). There will be a processing and copying fee. For foundation grants, contact the foundation for a listing of recently funded grants, and then contact the principal investigators directly.

Know Your Audience

Find out in advance as much as possible about the potential reviewers. For the NIH, the membership of study sections (Integrated Review Groups) is available on the Centers for Scientific Review Web site (www.csr.nih.gov/committees/rosterindex.asp). Search the literature to determine the potential reviewers’ areas of expertise. The Web sites of foundations, or their staff, may provide the composition of review committees. In these organizations, trustees or board members and foundation staff may also review the grant, so it is imperative that key sections and significance be understandable to lay reviewers.

Table 1. Clinical Research Funding Sources for New Investigators*

<table>
<thead>
<tr>
<th>Source</th>
<th>Details about Funding Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH/AHRQ</td>
<td>K08 (Mentored Clinical Scientist Development Award)</td>
</tr>
<tr>
<td></td>
<td>K23 (Mentored Patient-Oriented Research Career Development Award)</td>
</tr>
<tr>
<td></td>
<td>R01 (Investigator Initiated Research Grant)</td>
</tr>
<tr>
<td></td>
<td>R03 (Small Grant)</td>
</tr>
<tr>
<td></td>
<td>R21 (Exploratory/Developmental Research Grant)</td>
</tr>
<tr>
<td></td>
<td>Other federal funding sources include: CDC, CMS, VA</td>
</tr>
<tr>
<td>Foundations/national organizations</td>
<td>Specific to clinical or research area, such as Alzheimer’s Association, American Cancer Society, American Diabetes Association, March of Dimes</td>
</tr>
<tr>
<td></td>
<td>Offer career, project, pilot grants</td>
</tr>
<tr>
<td>Industry/pharmaceutical companies</td>
<td>Some unrestricted educational or research grants</td>
</tr>
<tr>
<td></td>
<td>Grants may be linked to specific drug or product</td>
</tr>
<tr>
<td>Local/community/intramural</td>
<td>Local or community foundations, local organizations, hospital auxiliaries</td>
</tr>
<tr>
<td></td>
<td>Source for small research projects, clinical demonstration projects, quality improvement initiatives, service delivery enhancements</td>
</tr>
<tr>
<td></td>
<td>Intramural programs (e.g., pilot grants from Center grants, such as General Clinical Research Center, Older Americans Independence Center, Diabetes Center)</td>
</tr>
</tbody>
</table>

* AHRQ = Agency for Healthcare Research and Quality; CDC = Centers for Disease Control and Prevention; CMS = Centers for Medicare & Medicaid Services; NIH = National Institutes of Health; VA = Department of Veterans Affairs.
Stress the Significance

The importance of the proposed study should be communicated clearly and should be readily apparent to someone outside the field. Present the burden of the problem in quantitative terms, and demonstrate the impact of the proposed research on the field. The grant writer needs to convey why this research is important—innovative, incremental knowledge or paradigm shift—and its implications.

Follow All the Rules

Obtain and follow all of the guidelines exactly. Grant proposals that do not meet guidelines are commonly returned without review. Even if they proceed through the review process, the score may be adversely affected. Determine the requirements for documentation of Institutional Review Board (IRB) approval because these vary by funding agency and mechanism. Most applications involving human participants require a description of the proposed methods for their protection. Some grants allow the investigator to provide evidence of IRB approval after peer review has been completed but before funding is awarded. Determine the procedures of the specific funding agency to which you are applying. Follow all instructions regarding font size (characters per inch and lines per vertical inch), margins, format, and content. Grammatical and typographical errors annoy reviewers and convey a sense of carelessness that does not reflect favorably on the skills of the grant writer. Indicate the principal investigator’s name and grant number on every page, and number each page. Do not expect any flexibility about submission dates.

Get Advice and Input from a Biostatistician

Input on the methods and analyses from an experienced biostatistician will enhance the success of your proposal. Seek advice early on for input about study design, data analysis plans, and sample size calculations.

Allow Enough Time for Prereview and Revision

Set yourself an internal deadline for completing the application 4 to 6 weeks before the actual deadline to allow time for prereview by mentors and colleagues, as well as by an experienced investigator outside of your field. Give colleagues at least 2 weeks for their review. Allowing enough time to revise the proposal in response to this feedback before submission will maximize the chances for success. It is important that you prepare the application carefully and convincingly. A high-quality product is more important than meeting a funding deadline, particularly when the application could be submitted for a future funding cycle or a different funding mechanism. In addition, allow ample time to refine budgets and subcontracts, and to obtain letters of support. See Figure 1 for a proposed timeline for tasks related to grant preparation.

Write the Abstract Carefully

The abstract should distill the essential elements of the research project into short, concise, and clear statements. The abstract will guide the assignment of a grant to a particular study section and is generally the first portion of the grant that reviewers read. Therefore, the abstract must engage the reviewers’ interest immediately and sustain their interest throughout. The abstract should highlight the nature of the problem, the need for the research, the hypo-

![Figure 1. Grant-writing timeline: example 1.](image-url)
esis to be tested, the methods to be used, and the significance and unique features of the research. It is good practice to write the abstract early and revise it throughout the process instead of writing it just before the grant is submitted.

**Avoid Use of Appendices**

Never put any vital information into appendices. The grant should stand alone, and appendices should only provide supporting materials. The reviewers may not receive or read the appendices.

**Major Review Issues Identified in NIH Grant Proposals**

**Study Approach**

To base our recommendations on evidence from the actual grant review process, we examined the review sheets (formerly called “pink sheets”) from 66 R01 applications submitted to 1 clinical research study section. This informal study was intended to describe the representative types of problems commonly raised in NIH review of grants for patient-oriented research. We categorized major areas of critique raised by reviewers on each of the grants (Appendix Table 1, available at www.annals.org). This study was not intended to comprehensively or systematically examine the NIH review process or to thoroughly describe the grant-writing process. More detailed information on grant-writing can be found elsewhere (12–21).

**Results**

See Appendix Table 1 (www.annals.org) for the major review issues identified in 66 NIH applications by grant section. In general, from the reviewers’ perspective, the most important sections of the grant are specific aims/hypotheses, methods, and preliminary work. Thus, these sections should receive the appropriate amount of time and space in the grant-writing process.

**Specific Aims/Hypotheses**

This is the most important section of the grant. Common critiques from reviewers are that the specific aims and hypotheses are poorly focused, undeveloped, or overly ambitious (Appendix Table 1 at www.annals.org). Grant writers should spend considerable time and energy on fully conceptualizing and articulating the key elements of the research questions and hypotheses. We advise getting careful input from mentors, colleagues, and collaborators to refine this section. Begin the section with a concise, accurate synopsis of the research (study design, sample size, study groups, and primary outcomes) so that reviewers can tell what is planned in the research proposal; no additions or surprises should appear later. Follow the synopsis with clearly worded primary and secondary aims and related hypotheses. They should be focused, clearly conceptualized, and feasible. See Appendix 2 and Appendix 3 for examples (Appendix 3 is available at www.annals.org).

**Background and Significance**

This section justifies and builds the case for the project, but it is important to focus on the proposed specific aims and highlight the need for the proposed study. This section puts the project into context by providing essential background information for the content area, showing how the proposed project builds on previous work, and identifying gaps in previous knowledge. Common critiques from reviewers are that this section did not justify the need for the study, provided too much extraneous background information, or overstated the significance of the study (see Appendix Table 1 at www.annals.org). After reviewing the literature for pertinent areas, the grant writer should strive for balance in setting the context for the grant. For each background area presented, it is important to show exactly how the background directly links with the proposed project. This section should naturally progress from the description of the current state of knowledge to the gap that the proposed research will fill. The following types of closing sentences on the paragraphs can be helpful to guide the reviewer:

- “Thus, these studies demonstrate the importance of this area [elaborate here]”
- “These studies provide the important background for this study in...”
- “The proposed project will build on this previous work [or address limitations in the previous work by]...”

For each area covered, explicitly state the relationship to the proposed project. Avoid the common mistake of making this section too long; be sure to leave adequate room for a fully developed Methods section. Having a “Significance” paragraph at the end of the Background section can help to frame the current status of the work in the field and explain how the proposed project will make a contribution. Specify in strong but realistic terms how the proposed project will contribute to the field. Use this section to justify the study and the methods used. Make the case for the proposed project, but be careful not to overstate its significance.

**Preliminary Studies and Pilot Work**

Reviewers are particularly interested in detailed description of preliminary or pilot work that is directly linked to the proposed study. Common critiques from reviewers are that preliminary or pilot work was lacking, was inadequately described, or lacked clear linkage to the proposed study. This section should summarize the principal investigator’s (or co-investigators’) previous work related to the proposed project. The principal investigator (or co-investigators) on the grant should be an author on the studies presented, and the references should be provided (manuscripts or abstracts). This section allows the investigator to convince the reviewers that 1) he or she has the expertise and experience to carry out this work, 2) the work is feasible, and 3) suitable groundwork has been done. This
section shows the reviewer that the investigator knows how to do research; shows the thoughtfulness, rigor, and preparation needed for the study; and gives important preliminary data for the proposed project. This important section warrants space and detail. For each preliminary or pilot study indicated, present the specific objectives, methods, results (with brief description and data), and significance (provide direct linkage with the proposed study—one quarter to one half a single-spaced page for each) (for examples, see Appendix 3 at www.annals.org).

Indicate which studies provided experience with the proposed methods (for example, design, intervention, assessment instruments, and enrollment strategies) in the current study, even if they are on a different topic. Pilot work to assure availability of study participants is key. The presentation of pilot data on a proposed intervention strategy—its feasibility, reproducibility, and standardization—is crucial here. This is one section where “more is better,” as long as the contributions and linkage to the proposed project are clear.

General Issues

Reviewers often raise issues on the layout and formatting of the grant, such as comments about typographical errors, small font sizes, formats that were difficult to read, excessive use of topic-specific jargon or abbreviations, and information presented in the wrong sections (for example, background information in the Methods section, new aims in the Analysis section).

Reviewers are unlikely to be convinced that the principal investigator is a good researcher if the grant is sloppily written or poorly laid out. Give time and attention to proofreading and making the grant easy to read. Provide spaces between paragraphs and between sections. Address study limitations thoroughly and realistically. For revised grants, the reviewers will focus on the degree of responsiveness to previous critiques. Provide an itemized, cordial, thoughtful response to each reviewer comment.

Methods

This section represents the heart of the grant, and all of the grants reviewed had comments on the Methods section (Appendix Table 1 at www.annals.org). The most common general issue is that the methods were underdeveloped. We recommend that grant writers devote at least 50% of the page allowance of the grant to methods, with particular attention to the specific issues raised in the following paragraphs.

Design and Setting. Describe the study design in detail. If randomization is involved, describe the procedure. Describe the method for blinding of participant allocation to treatment groups. If applicable for observational studies, describe how you will select case-patients and controls. Will you enroll a representative sample of the target population? If not, will there be any potential biases? How will you handle them? Describe the setting or settings in enough detail so that reviewers can understand how this setting would compare or extrapolate to other study settings.

Study Sample: Inclusion Criteria. The major purpose of this section is to carefully describe and justify the choice of the study sample. Common critiques from reviewers are that the study sample is potentially biased or nonrepresentative, or that the inclusion criteria are poorly described or not well-justified. We advise addressing any potential biases and assuring that these will not invalidate the study results. The NIH requires inclusion of women, minorities, and children; grant writers must justify the exclusion of these populations.

Study Sample: Exclusion Criteria. Exclusion of partici-
Table 2. Checklist for the Grant-Writing Process

<table>
<thead>
<tr>
<th>Section and Topic</th>
<th>Questions To Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific aims/hypotheses</td>
<td>Are the aims well focused and fully conceptualized? Are the hypotheses clearly articulated? Do the aims appear balanced—not overly ambitious or unrealistic?</td>
</tr>
<tr>
<td>Background/significance</td>
<td>Is the significance/importance of the work evident? Is the work innovative? Does it contribute substantially to previous work in the field? Is the need for the study (or all aspects of the study) well-justified? Is the significance overstated? Is there extraneous information?</td>
</tr>
<tr>
<td>Preliminary/pilot studies</td>
<td>Are preliminary studies well described and their contributions to the proposed project clear? Is there sufficient pilot work? Is availability of subjects assured? Are enrollment and/or intervention procedures tested and feasible?</td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Study sample</td>
<td>Are inclusion and exclusion criteria fully described and well-justified? Are the reasons for selecting this sample clear, not merely convenience? Are there too many exclusions that are not well justified, or are important exclusions overlooked? Is availability of adequate numbers of participants from the sampling frame assured? Are there enough participants in the setting to do this study as described?</td>
</tr>
<tr>
<td>Data collection/procedures</td>
<td>Are procedures well-described? Are there quality assurance measures for data collectors? Is there adequate description of study instruments/measures? Are standardized, validated measures used? Are all important study variables described and collected? Are there extraneous variables that are never used in subsequent analyses?</td>
</tr>
<tr>
<td>Outcome</td>
<td>Is the outcome adequately described, defined, and specified? Are the validity, reliability, and performance characteristics of the outcome measure provided? Are the outcome data collected by researchers who are blinded to the study hypotheses and study group assignment?</td>
</tr>
<tr>
<td>Intervention (if applicable)</td>
<td>Does the intervention appear potent (that is, is it likely to be effective as described) Is the intervention well-described—can you understand what was done, or is it a “black box”? Is the protocol standardized so that it is likely to be reproducible in other settings? Is the intervention administered by a separate individual/group not involved in outcome assessment? Is there blinded administration of the intervention protocol (e.g., double-blind drug trial)? Is there randomization to study groups? Is there likely to be potential bias in the way the patients were allocated to treatment groups or received the intervention? Will adherence to the intervention be monitored? Will the effects of nonadherence be considered? Are safety issues regarding the intervention addressed? Is an appropriate control group selected? Are issues of contamination or co-interventions in the control group addressed?</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Have you consulted a biostatistician? Are the analytic approach and structure of analyses adequately described? Will an intention-to-treat approach be used? Is there adequate attention to potential confounders? Are there sample size or power calculations? Are attrition rates/losses provided? Do they appear realistic/justified? Do anticipated losses threaten the validity of the study? How will missing data and nonresponses be handled in analyses?</td>
</tr>
<tr>
<td>Summary</td>
<td>Are the strengths and weaknesses of the grant presented? How do the weaknesses affect the validity or interpretation of the study results? Are potentially fatal flaws unaddressed? Are the implications of the work discussed?</td>
</tr>
</tbody>
</table>

Patients may be required for feasibility or safety reasons, but the grant writer should remember that any exclusion will make the study less generalizable. Common critiques are that the reasons for exclusion were not well justified, that the exclusions would result in important bias in the sample, or that in some cases important exclusions were overlooked. We advise that each exclusion criterion be well-justified. Address any important potential biases, and assure that these will not invalidate the study results or their applicability to more general samples.

**Availability of Participants.** A common critique from reviewers is that the availability of participants for the proposed study is not assured. Provide data and assurance that adequate numbers of patients will be available for the study in the proposed setting, given the inclusion and exclusion criteria. The strongest evidence is pilot work in the pro-
proposed study population (even early descriptive work); if this is not possible, provide data from previous related studies.

**Data Collection/Procedures.** This section assures quality in the data collection procedures. Common critiques by reviewers are inadequate description of the proposed study instruments or variables and concerns about validity or reliability of the data collection methods. We advise describing all study procedures and instruments. A tabular format can help provide information on standardized and validated instruments, including references and performance characteristics, such as sensitivity, specificity, and reliability statistics. Describe all study variables (that is, measurements or data elements), and indicate how each variable will be used in subsequent analyses. List and define all variables; tabular formats can be useful (for examples, see Appendix Tables 2 and 3, available at www.annals.org). Be inclusive, but do not include any variables that will not be used in the analyses. Outline the screening and enrollment procedures, along with subsequent assessment and follow-up procedures. Elucidate the interviewer training and standardization or reliability assessments, as well as ongoing procedures for quality assurance of data collection.

**Outcomes.** This section provides a detailed description, including the operational definition and specification, of each study outcome. Common critiques by reviewers include concerns about the lack of or inadequate blinding of outcome assessment, inadequate description or specification of the outcome measure, or concerns about validity or reliability of the outcome. The grant writer must fully define the outcome or outcomes and describe the performance characteristics of the measures used for each outcome. Ideally, the outcomes should be assessed by trained research staff who are not involved in the intervention (if applicable) and who are blinded to the study hypotheses and to the intervention status of the participants. Describe how blinding will be achieved and maintained; address any potential threats to maintaining blinding. Describe how equal surveillance for outcomes will be assured in all study groups (that is, that the outcomes will be equally likely to be detected in all study groups). The study should be adequately powered to evaluate all of the primary outcomes in the study.

**Intervention/Controls (If Applicable).** This section should comprehensively describe the intervention strategy and how it will be implemented. Common critiques by reviewers are that grant proposals poorly describe the intervention, present an unstandardized intervention or one of questionable potency, do not adequately describe plans to monitor adherence, and do not address contamination or co-intervention in the control group. Another common reviewer concern is that inadequate randomization procedures or unblinded administration of the intervention may lead to potential bias in allocation to the intervention. We advise the grant writer to describe the intervention strategy and standardized protocols in sufficient detail so that the intervention is not a “black box” and might be replicated in other settings. Describe the interventionists, their proficiency, and any training required. Give details on how you will track adherence to interventions. Detail the quality assurance methods for the interventionists. Report how you will monitor potential sources of contamination or co-intervention in the control group during the study.

**Data Analysis/Sample Size Calculations.** This section describes all data analysis issues, including data management procedures, analytic approach, and sample size/power calculations. Common critiques by reviewers are insufficient description of the analytic approach, lack of an intention-to-treat analytic strategy, inadequate control for potential confounders, insufficient description of the handling of missing data, and not enough consideration of attrition. We advise early and ongoing involvement of a biostatistician in the grant-writing process to ensure statistical input in the study design, data management, analysis, and sample size calculations. Working with the biostatistician, the grant writer should fully describe data management and quality assurance procedures, such as double entry of data, error and validity checks, and training of staff who will handle data management procedures. Lay out proposed analyses here for each specific aim or hypothesis. Specify the outcome variable and the independent variables and covariates to be examined in each analysis. Laying out the framework for the analyses is of paramount importance; however, use caution in specifying only one statistical approach, since locking oneself into a particular statistical method (for example, logistic regression) may raise concerns. Discuss alternate strategies considered and why you decided on your approach. Carefully address how nonresponses and missing data will be handled in analyses.

Work with a biostatistician to provide relevant sample size and power calculations for primary outcomes applying best estimates for effect sizes from pilot work or previous studies. Estimate realistic attrition rates, and account for these in the calculations.

**Advantages and Limitations of Current Approach**

Realistically assess the strengths and weaknesses of the proposed project. State how you will address the limitations, and assure reviewers that the limitations will not invalidate the study results.

**Tentative Timetable**

A timetable for the proposed study is invaluable for reviewers to understand the study procedures and duration. Many graphical or tabular formats are available. For examples, see Figure 2 and the Appendix Figure (the Appendix Figure is available at www.annals.org).

**Summary and Conclusions**

Table 2 summarizes key questions to address in any grant proposal. This summary was based on the evidence
and recommendations in this article and can assist you with writing future grants. Following this approach will ensure that you have addressed key areas of concern from the reviewers’ point of view. We based our recommendations directly on primary evidence (reviewers’ comments) gathered from an NIH Study Section; thus, we provide a real-world representation of review issues. While results of only 1 study section are represented, a sizable number of grants were included and the comments capture critical issues that have been stressed in other reports (14–21).

We believe that application of this systematic approach to grant-writing may help to make the task less onerous and more enjoyable for new clinical investigators, and for all persons embarking on writing grants.

Appendix 2: Example of Specific Aims Section
(Note: Appendix 1 is available at www.annals.org.)

The following is an example of a specific aims section. The principal investigator gave permission for inclusion of this example.

Delirium, or acute confusional state, is a common, serious, and potentially preventable problem for hospitalized older patients. Our previous study, the Delirium Prevention Trial which involved 852 patients, documented the effectiveness of a multicomponent targeted intervention strategy (MTI) for substantially reducing the risk of development of delirium during hospitalization, compared with usual care (UC). The overall objectives of the current renewal application are to extend the analyses of the Delirium Prevention Trial to examine cost-effectiveness, secondary short-term and long-term outcomes, and effects of adherence on intervention effectiveness. These investigations will more fully establish the effectiveness of our intervention strategy, including its overall cost-effectiveness and the lasting nature of the benefits.

A.1. Specific Aim 1
To examine the cost-effectiveness of the multicomponent targeted intervention (MTI) strategy for delirium prevention compared with usual care (UC). This aim involves the following sub-aims:

1. To measure the direct health care costs of the MTI strategy, as compared with UC.
2. To perform a net cost analysis for health care costs for the MTI strategy for short-term (hospitalization and one-month follow-up) and long-term (six-month and twelve-month follow-up) periods.
3. To estimate cost-effectiveness ratios of MTI compared with UC.

The associated hypotheses are:

Hypothesis 1a: In the short term, the MTI strategy may result in increased costs compared with the usual care group because of the costs associated with the intervention itself.

Hypothesis 1b: In the long term, the MTI strategy will prove cost-effective since the intervention is effective and long-term cost-savings from reduction of delirium and its associated sequelae (e.g., institutionalization, rehospitalization, and increased use of home care) will offset—in whole or part—the costs of the intervention.

A.2. Specific Aim 2
To examine the effectiveness of the intervention strategy in the Delirium Prevention Trial relative to usual care on improving secondary outcomes in short-term (hospitalization and one-month follow-up) and in six-month and one-year follow-up, including functional status, cognitive status, depression, subjective health rating, independent living ability, mortality, and health care utilization (i.e., rehospitalization, emergency department visits, physician visits, formal home health care, rehabilitation stay, and nursing home placement).

Hypothesis 2: The intervention strategy will result in improved short-term and long-term secondary outcomes compared with usual care.

A.3. Specific Aim 3
To measure the impact of level of adherence on effectiveness of the interventions in the Delirium Prevention Trial.

Hypothesis 3: The effectiveness of the intervention strategy will increase as the level of adherence with the interventions increases.

From Yale University School of Medicine, New Haven, Connecticut.

Acknowledgments: The authors thank the many students and trainees at Yale University School of Medicine, particularly the Robert Wood Johnson Clinical Scholars Program and students in the Investigative Medicine Program, who have attended our grant-writing courses and who gave us useful feedback that influenced this paper. They thank Dr. Ralph Horwitz for manuscript review and feedback and Ms. Patricia Fugal for assistance with manuscript preparation. This work is dedicated to Jordan and Benjamin Helfand.

Grant Support: In part by grants R01AG12551 (Dr. Inouye) and K24AG00949 (Dr. Inouye) from the National Institute on Aging. Dr. Inouye is Director of the Research Career Development Core for the Yale Claude D. Pepper Older Americans Independence Center (P30AG21342), is funded by the National Institute on Aging, and is the recipient of a Donaghue Investigator Award (DP98-105) from the Patrick and Catherine Weldon Donaghue Medical Research Foundation. Dr. Fiellin is a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar.

Potential Financial Conflicts of Interest: None disclosed.

Corresponding Author: Sharon K. Inouye, MD, Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street (DC013K), PO Box 208025, New Haven, CT 06520-8025.

Current author addresses are available at www.annals.org.

References


11. NIH announces updated criteria for evaluating research grant applications.
APPENDIX 1: THE NIH REVIEW PROCESS

Several federal NIH Web sites describe the NIH grant review process:


Briefly, on a thrice-yearly cycle the NIH Center for Scientific Review assigns received grants, roughly 16,000 per round, to an integrated review group. Within the review group each grant is assigned to a study section that typically includes 20 or more researchers. Within the study section, your grant is assigned to a primary and secondary reviewer, who thoroughly review and critique the grant on the basis of the project’s significance, approach, innovation, and the strengths of the investigator and research environment. The primary and secondary reviewers of all grants for a study section present the research proposal to the integrated review group, which is followed by a 10- to 15-minute discussion by all review group members, most of whom have focused primarily on the abstract, significance, and specific aims sections. Within 6 to 8 weeks of the review you will receive a summary statement, which includes a priority score and a percentile rank. Based on the score, percentile rank, and the priorities of the institute, your grant may or may not be funded during the cycle.

APPENDIX 3: GRANT SECTION EXAMPLES

Following are examples of specific aims and preliminary studies sections. Appendix Tables 2 and 3 provide examples of study variables tables, and the Appendix Figure shows a sample timeline. Principal investigators gave permission for inclusion of these examples.

Specific Aims Section: Example 1

The Insulin Resistance Intervention after Stroke Trial (IRIS) is a randomized, double-blind, placebo-controlled trial that will test the hypothesis that reducing insulin resistance and its sequelae with thiazolidinedione therapy will prevent stroke and myocardial infarction (MI) among patients with a recent ischemic stroke. Eligible subjects are men and women over 44 years of age without diabetes mellitus who have insulin resistance and a recent non-disabling ischemic stroke. During 3 years of recruitment, 3136 patients will be randomly assigned to pioglitazone, a thiazolidinedione (TZD), or placebo. The specific aims are:

Primary Aim

1. To determine if pioglitazone, compared to placebo, will reduce the overall risk for fatal or non-fatal stroke or fatal or non-fatal MI among non-diabetic men and women over age 44 years with insulin resistance and a recent ischemic stroke.

Among non-diabetics with insulin resistance, we hypothesize that pioglitazone will reduce the occurrence of any primary endpoint (fatal or non-fatal stroke or MI) within four years from 27% to 22%. The basis of this hypothesis is research showing that insulin

<table>
<thead>
<tr>
<th>Area*</th>
<th>Grants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific aims/hypothesis</td>
<td>30 (45)</td>
</tr>
<tr>
<td>Goals overstated, overly ambitious or unrealistic</td>
<td>12 (18)</td>
</tr>
<tr>
<td>Poorly focused or inadequately conceptualized</td>
<td>10 (15)</td>
</tr>
<tr>
<td>Hypotheses not clearly articulated</td>
<td>8 (12)</td>
</tr>
<tr>
<td>Background/significance</td>
<td>24 (36)</td>
</tr>
<tr>
<td>Need for study not well justified</td>
<td>19 (29)</td>
</tr>
<tr>
<td>Too much background, insufficient room for methods, extraneous information</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Overstatement of significance of study</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Preliminary/pilot studies</td>
<td>33 (50)</td>
</tr>
<tr>
<td>More pilot work needed</td>
<td>27 (41)</td>
</tr>
<tr>
<td>Studies cited with no clear link to proposed study</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Inadequate description of preliminary studies</td>
<td>2 (3)</td>
</tr>
<tr>
<td>General issues</td>
<td>24 (36)</td>
</tr>
<tr>
<td>Layout poor (editing/typographical/grammatical errors, inconsistencies, too-small font, omitted lines or tables, poor photocopy, difficult to read)</td>
<td>13 (20)</td>
</tr>
<tr>
<td>Use of jargon, abbreviations, undefined terms</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Information presented in wrong sections</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Limitations not adequately discussed</td>
<td>2 (3)</td>
</tr>
<tr>
<td>(For revision) Inadequately responsive to previous reviewers' comments</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Methods</td>
<td>66 (100)</td>
</tr>
<tr>
<td>Generally underdeveloped</td>
<td>10 (15)</td>
</tr>
<tr>
<td>Study sample</td>
<td>46 (70)</td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>36 (54)</td>
</tr>
<tr>
<td>Flawed sample (nonrepresentative, potential bias)</td>
<td>24 (36)</td>
</tr>
<tr>
<td>Poorly described</td>
<td>12 (18)</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>23 (35)</td>
</tr>
<tr>
<td>Reasons for exclusion not well justified</td>
<td>12 (18)</td>
</tr>
<tr>
<td>Important exclusions overlooked</td>
<td>7 (11)</td>
</tr>
<tr>
<td>Postenrollment exclusions (potential bias)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Availability of study participants not assured</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Data collection/procedures</td>
<td>18 (27)</td>
</tr>
<tr>
<td>Inadequate description of study instruments or variables</td>
<td>9 (14)</td>
</tr>
<tr>
<td>Concerns about validity or reliability of data collection methods</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Important variables omitted</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Many study variables not used in analyses</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Outcome</td>
<td>40 (66)</td>
</tr>
<tr>
<td>Concerns about adequate blinding of outcome assessment</td>
<td>24 (36)</td>
</tr>
<tr>
<td>Outcome measure inadequately described, defined, or specified</td>
<td>15 (23)</td>
</tr>
<tr>
<td>Concerns about validity or reliability of outcome measure</td>
<td>9 (14)</td>
</tr>
<tr>
<td>Intervention</td>
<td>16 (24)</td>
</tr>
<tr>
<td>Inadequate description of how adherence will be monitored or analyzed</td>
<td>9 (14)</td>
</tr>
<tr>
<td>Method of randomization not described or potential bias in selection process</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Concerns about potency of intervention</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Poorly described or unstandardized protocol</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Unblinded administration of intervention</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Concerns about unaddressed safety issues</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Controls</td>
<td>7 (11)</td>
</tr>
<tr>
<td>Issue of contamination or co-intervention</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Lack of or inadequate description of control group</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Data analysis</td>
<td>42 (66)</td>
</tr>
<tr>
<td>Inadequate control for important confounders</td>
<td>21 (32)</td>
</tr>
<tr>
<td>Insufficient description of analytic approach</td>
<td>16 (24)</td>
</tr>
<tr>
<td>Intention-to-treat analytic strategy needed</td>
<td>7 (11)</td>
</tr>
<tr>
<td>Inadequate description of handling missing data or nonresponses</td>
<td>7 (11)</td>
</tr>
<tr>
<td>Sample size/power</td>
<td>28 (42)</td>
</tr>
<tr>
<td>Lack of or inadequate description of sample size or power calculations</td>
<td>17 (26)</td>
</tr>
<tr>
<td>Estimates of attrition rates not provided, too low, or require justification</td>
<td>13 (20)</td>
</tr>
<tr>
<td>Anticipated attrition or losses to follow-up that threaten validity of study</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

*Note: Within general topic areas, individual grants may have been categorized according to more than one issue; thus, the sum of the individual issues for the topic area may exceed 100%.
resistance is associated with increased risk for stroke, myocardial infarction, and pathologic processes that promote vascular disease. Insulin resistance is associated with vascular endothelial dysfunction, vascular inflammation, dyslipidemia, impaired fibrinolysis, and atherosclerosis. Markers of vascular inflammation have recently been shown to be related to increased risk for stroke and heart disease. Pioglitazone markedly reduces insulin resistance and vascular inflammation. It also improves endothelial function, lipid metabolism and fibrinolysis. By these and other mechanisms we hypothesize that pioglitazone will protect patients with ischemic stroke and insulin resistance against recurrent vascular events.

**Secondary Aims**

1. To determine if pioglitazone, compared to placebo, will reduce the risk for recurrent stroke.

   We hypothesize that pioglitazone will reduce the occurrence of recurrent fatal or non-fatal stroke as a discrete outcome.

2. To determine if pioglitazone, compared to placebo, will reduce the risk for acute coronary syndromes (acute MI or unstable angina).

   Unstable angina is an important clinical event because it identifies individuals at high risk for MI who need urgent diagnostic and therapeutic intervention. We hypothesize that pioglitazone will reduce the overall occurrence of acute coronary syndromes.
3. To determine if pioglitazone, compared to placebo, is effective in preventing progression to overt diabetes among patients with cerebrovascular disease and insulin resistance.

Insulin resistance is the principal risk factor for type II diabetes. We hypothesize that pioglitazone, by sensitizing cells to insulin’s action, will prevent progres-
4. To determine if pioglitazone, compared to placebo, will reduce the risk for all-cause mortality.

We hypothesize that pioglitazone will reduce all-cause mortality because of its potent vaso-protective effects.

Specific Aims Section: Example 2

The proposed study will test the hypothesis that the efficacy for reducing illicit drug use and improving buprenorphine adherence of physician management (PM) plus cognitive-behavior therapy (CBT) is greater than PM alone during the initial 12 weeks of maintenance treatment, and during 12 weeks of follow-up. The study will explore potential patient predictors of differential treatment response identified in early studies (early abstinence achievement, cocaine abuse or dependence, prescription opiate (versus heroin) dependence) and help identify patient subgroups for whom CBT leads to the greatest differential improvement and is most cost effective. The study will also expand upon our prior work exploring important service delivery questions regarding costs, spillover effects, and patient and staff experiences regarding benefits and problems.

The specific aims of this grant are:

1. To determine the effect of the addition of CBT to PM compared to PM alone on illicit drug use in opioid dependent patients receiving buprenorphine maintenance in a primary care office-based setting.

Hypothesis 1a is that the addition of CBT to PM will lead to decreased illicit drug use during the time CBT is provided. Hypothesis 1b is that CBT will lead to decreased illicit drug use in the 12 weeks follow-up period after completion of CBT.

2. To determine the effect of the addition of CBT to PM compared to PM alone on adherence to buprenorphine.

Hypothesis 2 is that the addition of CBT to PM will lead to greater adherence to buprenorphine.

3. To compare the cost-effectiveness of the addition of CBT to PM to PM alone.

Hypothesis 3 is that the addition of CBT to PM will demonstrate cost-effectiveness overall.

4. To conduct exploratory analyses regarding patient-treatment matching and evaluate whether some patient subgroups (e.g., lack of achievement of early abstinence, cocaine abuse or dependence, and heroin (versus prescription opiate) dependence) specifically benefit from the addition of CBT to PM.

Preliminary Studies Section: Example 1

Secondary Outcomes of Delirium: Does Delirium Contribute to Poor Hospital Outcomes? A Three-Site Epidemiologic Study*

In three prospective cohorts totaling 727 patients aged 65 years and older, delirium was found to be an important independent prognostic determinant of hospital outcomes at discharge and 3-month follow-up, including new nursing home placement, death or new nursing home placement, and functional decline—even after controlling for age, gender, dementia, illness severity, and functional status. These cohorts were observational (non-interventional), and the populations were distinct from the Delirium Prevention Trial subjects. The adjusted odds ratios for delirium associated with each secondary outcome were as follows. Delirium was a significant predictor of new nursing home placement at both hospital discharge (adjusted odds ratio (OR) for delirium = 3.0, 95% confidence interval (CI) 1.4-6.2) and at 3-month follow-up (adjusted OR = 3.0, CI 1.5-6.0). Although not a significant independent predictor of death (which was an
infrequent outcome), delirium was a significant predictor for the combined outcome of death or new nursing home placement (adjusted OR = 2.1, CI 1.1-4.0 at discharge and adjusted OR = 2.6, CI 1.4-4.5 at 3-month follow-up). Moreover, delirium was a significant independent predictor of functional decline at both hospital discharge (adjusted OR 3.0, CI 1.6-5.8) and 3-month follow-up (adjusted OR 2.7, CI 1.4-5.2).

**Significance**

This study documents that delirium itself serves as an important independent prognostic predictor for secondary outcomes at hospital discharge and 3-month follow-up. The Delirium Prevention Trial—with its successful reduction of delirium rates—may well improve these secondary outcomes. Thus, this preliminary study provides convincing evidence justifying the need for the proposed detailed examination of secondary short-term and long-term outcomes in the current application. The current application will allow completion of follow-up data collection (telephone interviews and medical record reviews), tracking of health care utilization data and other secondary outcomes (nursing home placement, rehospitalization), tracking of all losses to follow-up, and verification of mortality status—all of which are critical to conduct the proposed analyses.

* Cite pertinent abstracts or references for preliminary work.

**Preliminary Studies Section: Example 2**

**The Study of Insulin Resistance after Stroke-II (SIRS-II)**

This study was a critical preliminary investigation seeking proof of the principle that TZDs would reduce insulin resistance in non-diabetic patients with cerebrovascular disease. The aim of SIRS-II was to examine the effectiveness of pioglitazone 45mg daily, compared with placebo, for improving insulin sensitivity among non-diabetic patients with insulin resistance and a recent TIA or ischemic stroke.

SIRS-II was a randomized, double-blind, placebo controlled trial. Eligible patients were recruited from SIRS-I if they had an insulin sensitivity index ≥ 2.5 and no history of New York Heart Association class III or IV heart failure (a contraindication to therapy). Consenting patients underwent a repeat oral glucose tolerance test before randomization if their prior test was over 30 days old. Patients were randomized according to a pre-printed schedule that was prepared by the investigational pharmacist at Yale-New Haven Hospital. Randomization was blocked in groups of 4 subjects to assure balanced treatment allocation. Investigators did not have access to the randomization schedule. After randomization, a research associate contacted subjects by telephone weekly for one month to supervise dose escalation, monitor for side effects, and encourage adherence. The study pill was dispensed as a 15mg tablet of pioglitazone or matching placebo. Dosage was increased by one pill weekly to a total of 3 pills per day by week three. After the first month, the telephone contact was reduced to every two weeks. At two months, blood was obtained from all subjects to monitor for liver toxicity (serum aspartate transaminase). At 3 months, all subjects underwent a repeat oral glucose tolerance test, measurement of serum aspartate transaminase, and physical examination, and treatment was stopped.

From 3/20/01-6/1/01, 20 patients were enrolled, with mean age of 67 in the placebo group (n = 10) and 66 in pioglitazone group (n = 10). As shown in Appendix Table 4, the insulin sensitivity index declined 0.1 units (-1%) among patients who received placebo and increased 1.1 units (62%) among patients who received pioglitazone (p = 0.0003), indicating a significant reduction in insulin resistance for pioglitazone recipients.

**Appendix Table 4. Mean Change in Insulin Sensitivity Index according to Treatment**

<table>
<thead>
<tr>
<th>Time of test</th>
<th>Placebo (N = 10)</th>
<th>Pioglitazone (N = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td>1.8 ± 0.5</td>
<td>2.1 ± 0.6</td>
</tr>
<tr>
<td>Exit</td>
<td>1.7 ± 0.6</td>
<td>3.2 ± 1.2</td>
</tr>
<tr>
<td>Change</td>
<td>−0.1 ± 0.6</td>
<td>1.1 ± 1.1</td>
</tr>
<tr>
<td>Proportional change</td>
<td>−1% ± 30%</td>
<td>62% ± 38%</td>
</tr>
</tbody>
</table>

Other measures of insulin sensitivity also indicated that pioglitazone had a significant effect. Fasting insulin declined 29% among patients who received pioglitazone (from 17 μU/ml to 11 μU/ml) compared with a 7% increase among patients who received placebo (from 18 μU/ml to 19 μU/ml). The homeostasis model assessment of insulin sensitivity (HOMA), an alternative measure to the Insulin Sensitivity Index, decreased 29% among patients who received pioglitazone compared with a 10% increase among placebo recipients (a lower HOMA score indicates greater insulin sensitivity). In a secondary analysis using stored blood, mean C-reactive protein declined 33% among patients who received pioglitazone compared to an increase of 10% among persons who received placebo (p = 0.06 for comparison of mean change).

Pioglitazone therapy was well tolerated. No patients were withdrawn from therapy because of side effects, and 6 out of 10 patients on active therapy said they would take their assigned treatment again if asked. All patients were interviewed every two weeks regarding specific side effects. Compared to patients on placebo, patients assigned to pioglitazone were more likely to report nausea (3 subjects compared to none) and weight gain (3 subjects compared to none). Despite the more frequent report of weight gain among subjects taking pioglitazone, actual mean weight gain was less among subjects receiving pioglitazone (-0.8 pounds) than among those receiving placebo (+2.2 pounds).

In summary, these findings demonstrate that pioglitazone is effective in reducing insulin resistance and safe for use among elderly non-diabetic patients with a recent TIA or ischemic stroke.

* Cite pertinent abstracts or references for preliminary work.

**Preliminary Studies Section: Example 3**

**Methadone Maintenance for Stabilized Patients in Primary Care**

To determine the feasibility and efficacy of office-based methadone maintenance provided by primary care physicians for...
stable opioid dependent patients, we performed a randomized controlled open clinical trial in the offices of primary care internists and an opioid treatment program (OTP). Of 87 eligible patients, 47 opioid dependent patients on methadone without evidence of illicit drug use for one year and without significant untreated psychiatric comorbidity were randomized to office-based methadone maintenance from primary care physicians or usual care at an opioid treatment program. Thirteen of the twenty-two (59% CI: 38%-80%) patients in office-based care compared with 11/24 (46% CI: 26%-66%) of OTP patients had evidence of any illicit drug use by self-report, urine or hair toxicology testing during the six-month treatment period \( (p = 0.37) \). Ongoing illicit drug use meeting criteria for clinical instability occurred in 4/22 (18%, CI: 2%-34%) of office-based care patients compared with 5/24 (21%, CI: 5%-37%) of OTP patients \( (p = 0.82) \). Sixteen of the 22 (73%, CI: 54%-92%) office-based patients compared with 3/24 (13%, CI: 0%-26%) of the OTP patients felt the quality of care was excellent \( (p < .001) \). There were no differences over time within or between treatments in functional status, or the use of health, legal or social services. These results support the feasibility and efficacy of transferring stable methadone maintained opioid dependent patients to physician offices for continuing treatment.

**Relevance**

This study supports the feasibility and efficacy of opioid agonist maintenance in primary care and demonstrates our ability to conduct detailed assessments of opioid dependent patients receiving this treatment in a clinical trial. The study also points to the fragility of abstinence even among stable, methadone maintained patients and suggests the potential importance of providing patients cognitive behavioral therapy (CBT) to prevent resumption of illicit drug use.

* Cite pertinent abstracts or references for preliminary work.

**Current Author Addresses:** Drs. Inouye and Fiellin: Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street (DC013K), PO Box 208025, New Haven, CT 06520-8025.
Update in Pulmonary Diseases

Martin J. Tobin, MD

This year’s Update in Pulmonary Diseases incorporates articles on the following topics: mechanical ventilation, obstructive lung disease, and pulmonary infection.

Mechanical Ventilation

Physician Training Did Not Provide Adequate Education on Mechanical Ventilation


Only 37% of critically ill patients receive care from an intensive care specialist. National surveys show that 67% of patients admitted to intensive care units remain in the care of their primary physicians and that 59% of general internists in practice use mechanical ventilation.

Cox and colleagues wanted to determine the level of understanding of mechanical ventilation among internal medicine residents. They administered a 19-question examination to 259 residents at 31 internal medicine residency programs. Of the participants, 46% were residents in a university hospital and 54% were residents in a community hospital. About half of the residents had more than 4 months of intensive care unit experience during training.

The average score on the test was 74% correct (range, 37% to 100%). More than one third of the residents answered fewer than 70% of the questions correctly, and 10% of the residents answered fewer than half correctly. Most residents correctly identified tension pneumothorax (86% correct) and clinical findings suggestive of severe hypotension secondary to auto-positive end-expiratory pressure (93% correct). High rates of incorrect answers were for questions on setting tidal volume in patients with the acute respiratory distress syndrome (48% incorrect), identifying whether a patient was ready for a weaning trial (38% incorrect), and recognizing an indication for noninvasive ventilation (27% incorrect).

Higher scores were associated with a closed-unit versus open-unit organization (76% vs. 71% correct), resident perception of greater versus lesser knowledge (79% vs. 71% correct), and graduation from a U.S. versus foreign medical school (75% vs. 69% correct). Only 46% of program directors thought residents had adequate knowledge of ventilators by graduation.

In summary, the data suggested that many senior residents in internal medicine are not gaining the knowledge needed to appropriately care for patients who receive mechanical ventilation. This finding is concerning because general internists are responsible for the care of most critically ill patients.

Prescribed Tidal Volumes Decreased before the Release of Randomized Trial Results


Research in the 1980s showed that patients with the acute respiratory distress syndrome were vulnerable to injury caused by high tidal volume. Studies in the mid- to late 1990s found that mortality was lower when patients were ventilated with a tidal volume less than 6 mL/kg of body weight as compared with 12 mL/kg. In this study, Weinert and colleagues wanted to see whether this knowledge was affecting ventilator care in the acute respiratory distress syndrome, particularly in response to the Acute Respiratory Distress Syndrome (ARDS) Clinical Network study released in March 1999. They examined records of 398 patients hospitalized for acute lung injury at 3 University of Minnesota hospitals, tracking the set tidal volume prescribed at 8 a.m. on the first 3 days of treatment. They then performed linear analysis (developed a segmental model) to see whether the setting changed between 1994 and 2001.

In the first 5 years (1994–1995 to 1998–1999), the average tidal volume was 9.5 mL/kg. The tidal volume decreased in mid- to late 1998, before the release of the ARDS Clinical Network study. For the last 2 years of the study (1999–2000 to 2000–2001), tidal volume decreased further to 8.9 mL/kg. Physicians prescribed a tidal volume of 6 mL/kg or less, which was recommended in the ARDS Clinical Network study, in just 0.9% (2 of 222) of patients. Also, between the first and third days of mechanical ventilation treatment, tidal volume decreased by only 33 mL, indicating that physicians generally don’t change tidal volume once they set it on the first day. Patients with less severe lung injury received higher tidal volumes.

When physicians set the tidal volume in patients requiring mechanical ventilation, they typically check for the amount of distention in the chest by measuring the plateau pressure (the ventilator pressure during the pause at the end of an inspiration). Ideal plateau pressure should be less than 32 cm H₂O, since pressures that exceed that level cause alveolar overdistention; this, in turn, places a patient at risk for developing ventilator-induced lung injury. Dur-
Early Noninvasive Pressure Support Ventilation Accelerated Resolution of Cardiogenic Pulmonary Edema


Physicians know that, in patients with acute respiratory failure secondary to chronic obstructive pulmonary disease, noninvasive ventilation decreases mortality and the need to intubate. Also, in patients with acute respiratory failure secondary to cardiogenic pulmonary edema, continuous positive airway pressure, which delivers positive end-expiratory pressures during spontaneous breathing, decreases the rate of intubation but not mortality. Furthermore, combining continuous positive airway pressure with pressure support enables patients to breathe with less effort as compared with continuous positive airway pressure alone. Despite this evidence favoring noninvasive ventilation, a study in patients with cardiogenic pulmonary edema suggested that pressure support increased the risk for myocardial infarction. As a result of that study (1), treatment for acute respiratory failure secondary to cardiogenic pulmonary edema has become controversial.

In this study, Nava and colleagues wanted to determine whether noninvasive ventilation benefited patients with cardiogenic pulmonary edema. The multicenter, randomized trial involved 130 patients with acute respiratory failure secondary to cardiogenic pulmonary edema. Inclusion criteria were severe acute respiratory failure (ratio of PaO2 to fraction of inspired oxygen [FiO2] < 250), ability to breathe less than 10 L of oxygen per minute for 15 minutes or more (the time needed to stabilize patients and make a diagnosis), sudden-onset dyspnea with respiratory rate greater than 30 breaths/min, and typical physical signs of pulmonary edema. Exclusion criteria were immediate need for intubation, disturbance of consciousness, shock, ventricular arrhythmia, life-threatening hypoxia, acute myocardial infarction, severe chronic renal failure, and pneumothorax. Patients were randomly assigned to noninvasive ventilation or usual medical care. They received noninvasive ventilation through a standardized protocol with a full face mask. Pressure support was set at 10 cm H2O and increased in increments of 2 cm to the maximum tolerated level (14.5 ± 21.1 cm H2O). Positive end-expiratory pressure was set at 5 cm H2O and was increased until the physician observed a brisk increase in oxygen saturation on pulse oximetry (Spo2) (6.1 ± 3.2 cm H2O). The mean (±SD) duration of noninvasive ventilation was 11.4 ± 3.6 hours. The primary end point was the need for intubation, and the secondary end points were in-hospital mortality and changes in some physiologic variables.

Nava and colleagues found that, compared with medical therapy and supplemental oxygen, noninvasive ventilation with pressure support achieved more rapid improvements in dyspnea, respiratory rate, and PaO2–FiO2 ratio. Standard medical therapy and pressure support ventilation had similar overall rates of intubation (25% and 20%, respectively) and hospital mortality (14% and 8%, respectively). In the subgroup of patients with hypercapnia (PCO2 > 45 mm Hg), noninvasive ventilation achieved faster improvement in PCO2 than standard therapy and decreased the rate of intubation (6% vs. 29%; P = 0.015). Adverse effects were evenly distributed in the 2 groups (myocardial infarction occurred in 7 patients who received noninvasive ventilation and 5 patients who received standard medical therapy).

In conclusion, the use of noninvasive ventilation for treating acute respiratory failure secondary to cardiogenic pulmonary edema in the emergency department achieved faster improvements in dyspnea, respiratory rate, and arterial oxygenation without affecting intubation rate or mortality. (The intubation rate decreased in the subgroup of patients with hypercapnia.) Larger multicenter, randomized studies are needed to compare the relative efficacy of continuous positive airway pressure alone versus noninvasive pressure support.

Obstructive Lung Disease

Lung Volume Reduction Surgery Did Not Improve Survival for Patients with Severe Emphysema


The characteristic pathophysiologic feature in patients with emphysema is destruction of lung tissue, leading to loss of lung elasticity and development of hyperinflation. The hyperinflation can be decreased by resecting areas of lung disease. On the basis of this approach, in recent years, surgeons have widely used lung volume reduction surgery as palliative therapy. Because lung volume reduction surgery was being widely used and was expensive, the Centers for Medicare & Medicaid Services (formerly the Health Care Financing Administration) asked the National Institutes of Health to undertake a randomized, controlled trial to determine the actual benefits of the procedure. This paper by the National Emphysema Treatment Trial Research Group describes that trial. Their primary outcome measure was survival and exercise capacity 2 years after
surgery. They also tried to determine which patients would benefit most from the surgery.

The investigators randomly assigned 1218 patients (average age, 67 years) with severe emphysema to receive either surgery or medical therapy. All patients underwent 6 to 10 weeks of pulmonary rehabilitation before randomization, after which the investigators obtained baseline measurements. They performed pulmonary function testing, cycle ergometry, a quality-of-life questionnaire, a dyspnea questionnaire, high-resolution computed tomography, and radionuclide lung scanning and measured distance walked in 6 minutes. The surgery involved resecting 20% to 35% of each lung, targeting the most diseased portions.

After 2 years, mortality in both groups was similar ($P > 0.2$), indicating no benefit from surgery. An interim evaluation of the data identified 140 patients at high risk for death from surgery ($\text{FEV}_1 \leq 20\%$ predicted and homogenous emphysema or diffusing capacity $\leq 20\%$ predicted). Once the investigators identified this high-risk group, they stopped enrolling these participants in the trial. When they excluded this subgroup, the results did not change in favor of surgery: Overall mortality for surgery was 0.09 death/person-year versus 0.10 death/person-year for medical therapy (relative risk, 0.89; $P > 0.2$). Exercise capacity improved by more than 10 watts in 15% of the patients in the surgery group compared with 3% of patients in the medical therapy group ($P < 0.001$). Quality of life was also superior in the patients in the surgery group.

In secondary analyses, the investigators identified 4 subgroups of patients. One pair either had emphysema predominantly in the upper lobe or did not. The other pair had either high or low maximal exercise capacity at baseline (cutoff value was 25 watts for women and 40 watts for men). Among 290 patients with upper lobe predominance and low exercise capacity, mortality was lower in the surgery group (relative risk, 0.47; $P = 0.005$). Among 220 patients with non–upper lobe predominance and high exercise capacity, mortality was higher in the surgical group (relative risk, 2.06; $P = 0.02$). Mortality in the 2 groups was the same for the 2 therapies in the other 2 subgroups: 419 patients with upper lobe disease and high exercise capacity and 149 patients without non–upper lobe disease and low exercise capacity.

In conclusion, lung volume reduction surgery achieved greater improvements in exercise capacity and quality of life than medical therapy but did not affect mortality. Moreover, patients with non–upper lobe emphysema and high exercise capacity at baseline experienced increased mortality after surgery. A specific subgroup of patients did, however, experience increased survival, and on the basis of these findings, Medicare will now reimburse for surgery performed in patients with upper lobe predominance and low baseline level of exercise.

Limitations of this study include the authors’ decision not to correct the $P$ values for multiple comparisons. A second limitation is the use of secondary analyses to discover the only positive outcome. The lack of a biological process that links upper lobe emphysema with low exercise capacity raises further concerns about drawing strong conclusions from this secondary analysis.

**Leg Muscle Fatigue Explained Why Bronchodilation Did Not Improve Exercise Performance in Patients with Chronic Obstructive Pulmonary Disease**


Bronchodilators are a mainstay in the therapy for patients with chronic obstructive pulmonary disease (COPD). But some patients note that bronchodilators do not enhance their ability to exercise. In this study, Saey and colleagues wanted to determine whether leg muscle fatigue during exercise influenced the effect of bronchodilation on exercise endurance time.

They enrolled 18 patients with moderate to severe COPD (average age, 66 years; $\text{FEV}_1$, 38% predicted) who performed exercise on a bicycle to exhaustion. They then measured leg muscle fatigue by 2 methods: maximum voluntary contractions and twitch force during supramaximal magnetic stimulation of the quadriceps. The investigators gave the patients an anticholinergic ipratropium bromide or placebo before the exercise in a randomized, crossover, double-blind design.

Compared with placebo, inhalation of ipratropium bromide (500 μg) tended to increase the time that patients endured cycle exercise at a constant work rate to 440 seconds versus 322 seconds ($P = 0.06$). In the 9 patients who developed quadriceps fatigue (≥15% decrease in quadriceps twitch pressure) during exercise after placebo, ipratropium did not increase endurance time despite an 11% $\text{FEV}_1$ increase. In the 9 patients who did not develop quadriceps fatigue secondary to exercising after placebo, ipratropium increased endurance time to 479 seconds versus 249 seconds ($P < 0.05$). Improvement in endurance time with ipratropium correlated with twitch force of the quadriceps at 10 minutes after exercise preceded by placebo ($r = 0.59; P = 0.01$).

In summary, the development of quadriceps muscle fatigue during exercise seemed to explain why some patients with COPD did not have increased endurance time after ipratropium. About half of patients with COPD will develop leg muscle fatigue during maximal exercise. In this subgroup, bronchodilator therapy will not increase exercise time.
Smoking Impaired the Effectiveness of Oral Glucocorticoid Treatment for Chronic Asthma


More than 20% of patients with asthma smoke cigarettes. Compared with patients who never smoke, these patients have more severe symptoms, worse lung function, more hospitalizations, and higher mortality after a near-fatal heart attack. Chaudhuri and colleagues measured the effect of cigarette smoking on the action of oral glucocorticoids.

This randomized, placebo-controlled, crossover study involved 14 smokers, 10 ex-smokers, and 26 never-smokers. All participants were 40 to 49 years of age and had had asthma for 16 to 25 years. The smokers had a history of 26 pack-years and the ex-smokers had a history of 19 pack-years. All patients had similar levels of airway obstruction (average FEV<sub>1</sub>, 70% predicted). The patients received 40 mg of prednisone or placebo for 2 weeks. The primary end point was the change in FEV<sub>1</sub> before the administration of albuterol. The primary end point was the change in pre-albuterol FEV<sub>1</sub> after active treatment compared with placebo. Secondary end points were change in asthma control score and the morning peak flow. Study adherence was greater than 90%.

Compared with placebo, steroid therapy resulted in no statistically significant improvement in current smokers. Compared with placebo, steroids caused a greater increase in FEV<sub>1</sub> in never-smokers than in smokers (237 vs. 47 mL) and in peak expiratory flow (36.8 vs. 6.5 L/min), as well as a greater decrease (indicating improvement in symptoms) in asthma control score (−0.72 vs. −0.05). Compared with placebo, steroids caused an increase in morning and nighttime peak expiratory flow (29.1 and 52.4 L/min, respectively) in ex-smokers but did not affect FEV<sub>1</sub> or asthma control score.

Oral glucocorticoids are a mainstay in the management of acute severe asthma. This study suggested that smoking impaired the efficacy of oral glucocorticoids in patients with chronic asthma. Since the ex-smokers showed better peak flow after glucocorticoid use, this study also suggested that the effects of smoking may be partially reversible. The findings emphasized the importance of convincing patients with asthma not to smoke.

Preoperative Factors Identified Patients at Risk for Pulmonary Complications after Nonthoracic Surgery


About 5% of patients undergoing nonthoracic surgery develop pulmonary complications. Only 7 rigorous studies have examined the possible reasons for complications, and these studies did not agree with each other. The studies revealed that key elements of the patient history, physical examination, and pulmonary function testing were not predictive. In this study, McAlister and colleagues wanted to determine the accuracy of the preoperative assessment in predicting the risk for pulmonary complications.

The prospective study involved 272 consecutive patients referred for evaluation before nonthoracic surgery. The primary outcomes were pneumonia, atelectasis, or mechanical ventilation before discharge or by day 7 after surgery. An independent investigator, who was blinded to the preoperative data, performed the outcomes assessment. Laboratory testing, such as pulmonary function tests, chest radiographs, and blood gases, was performed at the discretion of the attending physician and was not routinely ordered. History and physical examination were obtained for all patients. Certain laboratory tests were done for most patients; for example, spirometry was done in 145 of 272 patients, and chest radiography was done in 124 patients. In the analysis, McAlister and colleagues set the criterion for a statistically significant P value at 0.005 or less rather than 0.05 or less because there were 10 multiple comparisons.

Postoperative complications occurred in 22 patients (8%): 6 required mechanical ventilation, 9 developed pneumonia, and 7 developed atelectasis. Six variables predicted pulmonary complications: PCO<sub>2</sub> of 45 mm Hg or more (odds ratio, 61.0), forced vital capacity less than 1.5 L (odds ratio, 11.1), maximum laryngeal height (distance from top of thyroid cartilage to suprasternal notch at end expiration) of 4 cm or less (odds ratio, 6.9), forced expiratory time of 9 seconds or more (odds ratio, 5.7), smoking history of 40 pack-years or more (odds ratio, 5.7), and body mass index of 30 kg/m<sup>2</sup> or more (odds ratio, 4.1). On multiple regression analysis, 3 clinical factors independently predicted pulmonary complications: age 65 years or older (odds ratio, 1.8), smoking history of 40 pack-years or more (odds ratio, 1.9), and maximum laryngeal height of 4 cm or less (odds ratio, 2.0). Earlier studies found a history of COPD, exercise for fewer than 2 blocks or 1 flight of stairs, and FEV<sub>1</sub> less than 1 L to be statistically significant, but after correcting for the multiple comparisons, this study found that these were no longer significant.

In summary, pulmonary function testing and clinical assessment identified factors that predicted a greater likelihood of pulmonary complications after nonthoracic surgery. They found that a simple physical sign of airway obstruction and hyperinflation—when the distance between the top of thyroid cartilage to the suprasternal notch (at the end of expiration) was 4 cm or less, about 4 finger breadths—was more accurate than many more sophisticated tests in predicting pulmonary complications after surgery.
Pulmonary Infection

The Clinical Pulmonary Infection Score Had Low Diagnostic Accuracy

Investigators reported in the early 1990s that the clinical pulmonary infection score was a reliable method for diagnosing ventilator-associated pneumonia. As a result, the score became widely used for diagnosing pneumonia without requiring bronchoscopy. The clinical pulmonary infection score is based on the presence of an aggregate of clinical signs on the day that pneumonia is suspected. These variables include temperature, leukocytes, appearance and volume of tracheal secretions, chest radiograph infiltrates, oxygenation (PaO₂/FiO₂ ratio), and culture of tracheal aspirate. The score can range from 0 to 12 points. The investigators in the initial study reported that a score greater than 6 had 93% sensitivity and 100% specificity for the diagnosis of ventilator-associated pneumonia, but no one has ever tested the score in an independent population of patients receiving mechanical ventilation.

In this study, Fartoukh and colleagues wanted to validate the accuracy of the clinical pulmonary infection score. Their prospective study involved 79 episodes of suspected ventilator-associated pneumonia in 68 consecutive patients. The managing physician categorized each patient subjectively according to the likelihood of pneumonia: very low (<20%), low (20% to 40%), moderate (40% to 60%), high (60% to 80%), and very high (80% to 100%). Another physician calculated the clinical pulmonary infection score. The nurse in charge of the patient obtained specimens through endotracheal aspiration and use of a protected telescoping catheter, inserted blindly without an endoscope. Bronchoscopy and bronchoalveolar lavage were performed, and a culture of bronchoalveolar lavage was taken as the reference standard for the pneumonia diagnosis.

Fartoukh and colleagues confirmed the diagnosis of pneumonia in 40 (51%) episodes. The managing physician estimated the likelihood of pneumonia as high or very high in only 20 of these 40 confirmed episodes. Of 39 episodes where pneumonia was not confirmed, the managing physician estimated the likelihood of pneumonia as high or very high in 14 episodes. The managing physician’s estimate had 50% sensitivity and 58% specificity.

Before microbiological testing, the mean clinical pulmonary infection score was 7.5 in the 40 episodes of confirmed pneumonia (score, 6.5) and the 39 episodes without confirmed pneumonia (score, 5.9). A clinical pulmonary infection score greater than 6 had 60% sensitivity and 59% specificity. Whether antibiotics were administered did not influence the diagnostic accuracy of the baseline score. The addition of Gram stain results of directed or blinded, protected telescoping catheter to the clinical pulmonary infection score increased sensitivity to 78% and specificity to 56%. The addition of Gram stain results of bronchoalveolar lavage to the clinical pulmonary infection score achieved 85% sensitivity and 49% specificity. The growth of bacteria in specimens obtained by endotracheal aspiration had 100% sensitivity but only 19% specificity for the diagnosis of pneumonia.

In conclusion, the clinical pulmonary infection score did not reliably diagnose ventilator-associated pneumonia. The ability of an experienced physician to diagnose the presence of pneumonia in a critically ill, mechanically ventilated patient was little better than chance. Attempting to make the clinical diagnosis more rigorous by using quantitative scoring was not much more reliable. To accurately diagnose pneumonia in a critically ill patient, physicians should use bronchoscopy to obtain protected sampling from the distal airways.

Eight and Fifteen Days of Antibiotics Achieved Comparable Results in Treating Ventilator-Associated Pneumonia

For ventilator-associated pneumonia, most experts recommend giving antibiotics for 14 to 21 days, but the recommended length of treatment is arbitrary and may be no more effective than a shorter course. Also, the longer course may foster the emergence of multidrug-resistant bacteria. Unreliable noninvasive techniques for diagnosing ventilator-associated pneumonia have contributed to a poor understanding of pneumonia, as exemplified by the results of the previous study.

Chastre and colleagues compared 8 days of antibiotic therapy with 15 days of antibiotic therapy in patients who have microbiologically proven ventilator-associated pneumonia. The randomized, double-blind trial involved 401 patients from 51 French intensive care units. All patients had been on a ventilator for 48 hours or more and had a new and persistent infiltrate plus 1 of 3 variables: purulent tracheal secretions, a temperature of 38 °C or more, or a leukocyte count greater than 10 × 10⁹ cells/L. All patients received a diagnosis of ventilator-associated pneumonia by quantitative culture results of bronchoscopic specimens. Chastre and colleagues randomly assigned the patients to receive either 8 days (n = 197) or 15 days (n = 204) of antibiotic therapy selected by the treating physician. The primary outcomes, assessed at 28 days, were death from any cause, microbiologically documented recurrence of pulmonary infection, and number of antibiotic-free days.

Mortality was equivalent for the 2 groups (18.8% in the 8-day group vs. 17.2% in the 15-day group). On
follow-up, Chastre and colleagues found gram-negative bacilli in 33% of patients receiving 8 days of therapy and 31% of patients receiving 15 days of therapy.

Recurrent infections were equivalent for the 2 groups (28.9% in the 8-day group vs. 26.0% in the 15-day group). More antibiotic-free days occurred in the 8-day group (13.1 ± 7.4 days vs. 8.7 ± 5.2 days; \( P < 0.001 \)). Patients with pneumonia caused by gram-negative bacilli had a higher recurrence of pulmonary infection when the pneumonia had been treated for 8 versus 15 days (40.6% vs. 25.4%), and recurrence was not associated with an unfavorable outcome. Among patients who developed recurrent infections, multidrug-resistant pathogens emerged less frequently among patients in the 8-day group (42.1% vs. 62.0%; \( P = 0.04 \)). Methicillin-resistant *Staphylococcus aureus* was isolated in 11% of both patient groups. The physicians had prescribed an aminoglycoside or fluoroquinolone plus a \( \beta \)-lactam in 91% of patients in the 8-day group and 92% of patients in the 15-day group and vancomycin in 39% of patients in the 8-day group and 37% of patients in the 15-day group.

In conclusion, 8 days of antibiotic therapy was as effective as 15 days in treating ventilator-associated pneumonia. The findings have implications beyond the treatment of ventilator-associated pneumonia. They raise the question of whether shorter courses of antibiotic therapy may also be effective in treating many other infections.

The Role of Anaerobes for Aspiration Pneumonia in Institutionalized Elderly Patients Needs Reexamination


Aspiration pneumonia has become the second most frequent principal diagnosis for hospitalization of Medicare patients. The listing of aspiration pneumonia as a discharge diagnosis doubled between 1991 and 1998, and hospitalization of elderly people for this condition is expected to increase in the future. In this prospective study, El-Solh and colleagues wanted to define the microbiology of severe aspiration pneumonia in institutionalized elderly patients.

They enrolled 95 patients older than 65 years of age with severe aspiration pneumonia. The patients had been admitted from a long-term care facility to an intensive care unit and required mechanical ventilation. All patients had at least 1 comorbid illness and 87% had 2 or more, including stroke (78%), ischemic heart disease (51%), and COPD (35%). El-Solh and colleagues used the following criteria for aspiration pneumonia: new infiltrate; clinical symptoms or signs (cough or sputum or temperature change plus 2 of the following: pleurisy, dyspnea, delirium, alveolar–arterial oxygen gradient, or change in leukocyte count); and risk factors for oropharyngeal aspiration. They determined the cause of the pneumonia by either bacterial growth in protected bronchoalveolar lavage of 10³ colony-forming units/mL or more; urinary antigen for *Legionella* or pneumococcus; or respiratory assay for influenza antigens.

El-Solh and colleagues obtained quantitative bronchial samples from all 95 patients, from which they identified 67 pathogens. Organisms were aerobic gram-negative enteric bacilli (49%), *Staphylococcus aureus* (12%), and anaerobic bacilli (16%). The most common anaerobes were *Prevotella* and *Fusobacterium*. They recovered aerobic gram-negative bacilli in conjunction with 55% of anaerobic isolates. The amount of dental plaque did not differ between the aerobic and anaerobic groups. Poor functional status, based on a score of 14 to 18 on the Activities of Daily Living test, was the only determinant of anaerobic bacteria.

Fifty-four patients met the criteria for pneumonia. Their treatment consisted of monotherapy for 22 patients and combination therapy for 32. Of the 43 patients with aerobic isolates, 4 received inadequate antibiotics. Of the 11 patients with anaerobic isolates, 7 received inadequate antibiotics. This difference was statistically significant, but 6 of the 7 patients had an effective clinical response. Most of these patients have concurrent aerobic gram-negative infection.

Mortality was 33% for the aerobic group and 36% for the anaerobic group. Hypoalbuminemia and comorbid diseases were independent risk factors for poor outcomes.

In this study, most institutionalized elderly patients presenting with aspiration pneumonia did not have an anaerobic organism. Even when an anaerobic organism was present, most patients responded to antimicrobial therapy that was not directed specifically at anaerobic organisms. On the basis of these findings and given the increasing resistance to antimicrobial agents, we need to reexamine the importance of adding anaerobic coverage for aspiration pneumonia in institutionalized elderly patients.

From Loyola University Medical Center and Hines Veterans Affairs Hospital, Maywood, Illinois.

Grant Support: By a Merit Review grant from the Veterans Affairs Research Service.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Martin J. Tobin, MD, Division of Pulmonary and Critical Care Medicine, Hines Veterans Affairs Hospital, Route 111N, Hines, IL 60141; e-mail, mtobin2@lumc.edu.

Reference

Celiac disease is a common autoimmune disorder that has genetic, environmental, and immunologic components. It is characterized by an immune response to ingested wheat gluten and related proteins of rye and barley that leads to inflammation, villous atrophy, and crypt hyperplasia in the intestine. The disease is closely associated with genes that code for human leukocyte antigens DQ2 and DQ8. Transglutaminase 2 appears to be an important component of the disease, both as a deamidating enzyme that can enhance the immunostimulatory effect of gluten and as a target autoantigen in the immune response. Sensitive and specific serologic tests, including those for anti-transglutaminase antibody, are facilitating fast and noninvasive screening for celiac disease. Thus, they are contributing to a more accurate estimate of the prevalence of the disease and its association with other disorders. Celiac disease is associated with increased rates of anemia, osteoporosis, cancer, neurologic deficits, and additional autoimmune disorders. A gluten-free diet is the mainstay of safe and effective treatment of celiac disease, although its effect on some of the extraintestinal manifestations of the disease remains to be determined.

Clinical Presentation

The clinical presentation of celiac disease varies greatly and ranges from asymptomatic to severe malnutrition. The most common manifestations of celiac disease include abdominal pain, increased frequency of bowel movements, weight loss, bone disease, anemia, and weakness. Celiac disease is sometimes divided into clinical subtypes. The terms symptomatic or classic apply to cases that meet the classic features of celiac disease, which include chronic diarrhea, abdominal distention and pain, weakness, and sometimes malabsorption. In contrast, in the now-common atypical form of the disease, gastrointestinal symptoms may be absent or less pronounced; instead, extraintestinal features, such as anemia, osteoporosis, short stature, infertility, and neurologic problems, are more prominent (11–41) (Table 1). Patients with asymptomatic or silent celiac disease lack classic or atypical symptoms but have villous atrophy that may be discovered during endoscopy or intestinal biopsy for other reasons, or as a result of serologic screening. Because atypical presentations are increasingly found to predominate, celiac disease is now considered to resemble a multisystem disorder rather than a mainly gastrointestinal one (42, 43).

See also:
Print
Glossary ........................................... 290
Web-Only
Conversion of figures and tables into slides
Celiac Disease: Understanding a Complex Autoimmune Disorder

DIAGNOSIS AND MANAGEMENT

Current diagnostic criteria for celiac disease in clinical practice are based on revised guidelines proposed by the European Society for Paediatric Gastroenterology and Nutrition, which have been extrapolated to adults (44). According to these guidelines, celiac disease is present if histologic changes are found on intestinal biopsy while the patient consumes a gluten-containing diet and unequivocal clinical improvement occurs while he or she consumes a gluten-free diet.

Figure 1 shows a possible algorithm for diagnosing celiac disease that is based on the European Society for Paediatric Gastroenterology and Nutrition criteria (44) and on recommendations from the National Institutes of Health Consensus Development Conference on Celiac Disease (45). Patients usually undergo tests for serologic markers once celiac disease is suspected, either because characteristic symptoms are present or because they are in an at-risk group, such as having disorders associated with celiac disease (Table 1) or being a first-degree relative of a person with the disease. Measurement of anti–transglutaminase 2 (or antiendomysial) antibodies of the IgA isotype is more sensitive and specific for celiac disease than is the IgG isotype and is recommended for initial screening. However, IgA deficiency occurs in 1.7% to 2.6% of patients with celiac disease, which is a 10- to 16-fold increase over that in the general population (35). It is therefore helpful to also measure total IgA. If IgA deficiency is found, measurement of IgG class anti–transglutaminase 2 (or antiendomysial) and antigliadin antibodies is recommended.

If results of testing for IgA anti–transglutaminase 2 or antiendomysial antibodies is positive, if IgA deficiency is found and results of testing for IgG antibody (anti–transglutaminase 2, antiendomysial, or antigliadin antibodies) is positive, or if results of serologic testing are negative but clinical suspicion is high, intestinal biopsy should be performed (Figure 1). Because the disease may be patchy, as seen on chromoendoscopy and magnification endoscopy (46, 47), an adequate number of tissue samples (4 to 6 pieces) must be obtained (48, 49). Such sampling will further ensure that some sections will be oriented correctly to determine the degree of villous atrophy needed to make the diagnosis, whereas other pieces allow assessment of intraepithelial lymphocytosis, epithelial disarray, and degree of inflammation. Biopsy samples obtained with standard-size forceps from the descending duodenum at the level of the ampulla of Vater are sufficient for diagnosis (50). Interest is increasing in video capsule endoscopy for assessment of small-intestinal diseases, although use of this technique in patients with celiac disease has not been studied.

Characteristic histologic features of celiac disease include varying degrees of villous atrophy, with hyperplasia of the crypts and increased intraepithelial lymphocyte

### Table 1. Disorders Associated with Celiac Disease

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endocrine</strong></td>
<td></td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>11–13</td>
</tr>
<tr>
<td>Autoimmune thyroid disorders</td>
<td>14</td>
</tr>
<tr>
<td>Addison disease</td>
<td>15</td>
</tr>
<tr>
<td>Reproductive disorders</td>
<td>16, 17</td>
</tr>
<tr>
<td>Alopecia areata</td>
<td>18</td>
</tr>
<tr>
<td><strong>Neurologic</strong></td>
<td></td>
</tr>
<tr>
<td>Cerebellar ataxia</td>
<td>19–21</td>
</tr>
<tr>
<td>Neuropathy</td>
<td>22–24</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>25</td>
</tr>
<tr>
<td>Migraine</td>
<td>26</td>
</tr>
<tr>
<td><strong>Cardiac</strong></td>
<td></td>
</tr>
<tr>
<td>Idiopathic dilated cardiomyopathy</td>
<td>27</td>
</tr>
<tr>
<td>Autoimmune myocarditis</td>
<td>28</td>
</tr>
<tr>
<td><strong>Hepatic</strong></td>
<td></td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
<td>29</td>
</tr>
<tr>
<td>Autoimmune hepatitis</td>
<td>30, 31</td>
</tr>
<tr>
<td>Autoimmune cholangitis</td>
<td>32</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td>Anemia</td>
<td>33</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>34</td>
</tr>
<tr>
<td>Selective IgA deficiency</td>
<td>35</td>
</tr>
<tr>
<td>Sjögren syndrome</td>
<td>36</td>
</tr>
<tr>
<td>Juvenile chronic arthritis</td>
<td>37</td>
</tr>
<tr>
<td>Turner syndrome</td>
<td>38</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>39</td>
</tr>
<tr>
<td>Dental enamel defects</td>
<td>40, 41</td>
</tr>
</tbody>
</table>

---

**Glossary**

**Adaptive immune response:** Immune response mediated by B and T cells after exposure to a specific antigen. Involves memory, self/nonself recognition, and specificity.

**Antigen:** Molecule that can bind specific antibodies or lymphocytes of the immune system. An antigen is immunogenic if it can generate an immune response. An autoantigen is a self-antigen.

**Endomysial tissue:** Loose connective tissue around smooth-muscle fibers.

**Epitope:** Discrete site on an antigen, such as a particular amino acid sequence of a protein, that is recognized by an antibody or immune-cell receptor.

**Epitope spreading:** Diversification and spread of the immune response to autoantigens.

**Gluten:** Main storage proteins of wheat, which comprise many different species with similar amino acid sequence and biochemical properties. Glutens are divided into gliadins (the subject of most of the research on the immune response) and glutenins; both are implicated in celiac disease. The term “gluten” is often used generically to also describe similar proteins of rye and barley that are considered toxic in celiac disease.

**Human leukocyte antigens:** Cell-membrane proteins involved in presentation of specific epitopes (such as immunogenic gluten peptides) to T cells of the immune system, promoting T-cell activation.

**Intermolecular help:** Mechanism by which T cells specific for an antigen, such as gluten, help B cells produce antibodies against a self-antigen, such as transglutaminase, provided that complexes are formed between the 2 antigens.

**Innate immune response:** Nonspecific immune response to antigens, including anatomic and physiologic barriers, endocytic and phagocytic activity, and inflammatory secretions.
count. The criteria proposed by Marsh are often used to grade the disease (from 0 to 4) in terms of these features (51). Most symptomatic patients have partial, subtotal, or total villous atrophy, which are Marsh type 3 lesions. Positive identification of these abnormalities leads to a presumptive diagnosis of celiac disease and institution of a gluten-free diet. Clear clinical improvement while the patient is following the diet yields a definitive diagnosis. The serum antibodies generally disappear by 6 to 12 months, although they are not necessarily reliable indicators of the mucosal response (52, 53). When patients do not present with the classic clinical symptoms of celiac disease, a second biopsy that shows histologic improvement confirms the diagnosis. Gluten challenge is not considered necessary for diag-

Table 2. Common Pitfalls in Diagnosis of Celiac Disease

<table>
<thead>
<tr>
<th>Problem</th>
<th>Effect</th>
<th>Possible Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor sensitivity of anti-transglutaminase 2 antibody test</td>
<td>False-negative result for anti-transglutaminase 2 antibody in patients with celiac disease</td>
<td>Test for antiendomysial antibody; proceed with biopsy in case of high clinical suspicion</td>
</tr>
<tr>
<td>Poor specificity of anti-transglutaminase 2 antibody test</td>
<td>False-positive result for anti-transglutaminase 2 antibody in other diseases</td>
<td>Use test with human transglutaminase 2 as antigen; test for antiendomysial antibody</td>
</tr>
<tr>
<td>IgA deficiency</td>
<td>Negative results for IgA anti-transglutaminase 2 and antiendomysial antibodies</td>
<td>Measure total IgA; test for the IgG isotype of anti-transglutaminase 2 (or antiendomysial) and antigliadin antibodies</td>
</tr>
<tr>
<td>Patchiness of villous atrophy</td>
<td>False-negative biopsy result</td>
<td>Obtain adequate number of biopsy samples</td>
</tr>
<tr>
<td>Equivocal biopsy result</td>
<td>Inconclusive diagnosis</td>
<td>Review results with an expert gastrointestinal pathologist; test for HLA-DQ2 and HLA-DQ8 alleles; consider gluten challenge and repeat biopsy</td>
</tr>
<tr>
<td>No initial diagnostic biopsy</td>
<td>Inconclusive diagnosis</td>
<td>Consider gluten challenge; test for HLA-DQ2 and HLA-DQ8 alleles</td>
</tr>
<tr>
<td>Patient using immunosuppressive therapy</td>
<td>False-negative result on serologic testing</td>
<td>Consider biopsy if suspicion of celiac disease is high</td>
</tr>
</tbody>
</table>
Celiac Disease: Understanding a Complex Autoimmune Disorder

nosis, except in patients for whom no initial diagnostic biopsy was done or results of biopsy are unclear or uncharacteristic of celiac disease. In such cases, biopsy is repeated after clinical relapse subsequent to gluten challenge, or after 3 to 6 months if gluten challenge does not lead to symptoms (44). Patients should be told that they may have a severe reaction to the gluten challenge.

Of note, diagnosis of celiac disease based solely on serologic markers is not yet accepted, and identification of the characteristic mucosal abnormalities on intestinal biopsy is required. However, intestinal biopsy can also yield false-negative results, either because the intestinal damage is patchy or because mucosal changes are not detectable on light microscopy (2, 54). If results of biopsy are negative but serologic tests are positive and celiac disease is strongly suspected, the results of the biopsy should be reviewed with an expert gastrointestinal pathologist before additional biopsy is considered. In addition, if histologic examination yields equivocal results, it is useful to proceed with HLA typing. Although about 30% of the general population has the HLA-DQ2 or HLA-DQ8 markers, nearly all patients with celiac disease have them (55). Therefore, a negative result for both markers has an excellent negative predictive value for the disease (56). Table 2 summarizes issues that clinicians often face in diagnosing celiac disease and ways to manage them.

The mainstay of treatment of celiac disease is strict lifelong adherence to a gluten-free diet, in which the patient avoids food products containing wheat, rye, or barley. Even though various studies have found oat to be generally well tolerated (57), some patients appear to be sensitive to it, and the presence of oat-specific intestinal T cells has been demonstrated in persons with celiac disease (58). More important, concern about contamination from the above-mentioned cereals in commercial preparations of oat has led to reluctance in recommending it (57, 59). Commonly substituted grains in the gluten-free diet include rice, corn, quinoa, and buckwheat. Although use of a gluten-free diet safely and effectively manages celiac disease, adherence is not a trivial task in an age in which wheat flour is nearly ubiquitous in foods. Patients whose disease does not respond to dietary therapy should undergo a systematic evaluation (60, 61). The 2 most important points to clarify are whether the patient actually has the disease and whether the patient is truly consuming a gluten-free diet. Evaluation requires review of the original biopsy slides and assessment by an expert dietician. Associated conditions that must be ruled out include pancreatic insufficiency, lymphocytic colitis, bacterial overgrowth, and true refractory sprue with a clonal T-cell population (62–64).

**Antibodies as Diagnostic Markers**

Celiac disease is associated with circulating antibodies against gliadin and endomysial tissue. These markers have proven to be highly valuable in the diagnosis and management of celiac disease. Antiendomysial antibody has higher sensitivity and specificity than does antigliadin antibody and is regarded as a superior serologic marker for celiac disease. Although the antireticulin antibody test was widely used formerly and is still part of some antibody panels, it has inferior sensitivity and for the most part has been replaced by the antiendomysial antibody test. In 1997, the transglutaminase 2 enzyme was established to be the autoantigen for antiendomysial antibody (65). Evidence indicates that the associated antireticulin and antijejunal antibodies are also directed at the same antigen (66, 67).

The antiendomysial antibody test is an immunofluorescence staining procedure performed by examining the binding of antibodies in patient serum to endomysial tissue from human umbilical cord or monkey esophagus. Antibodies that bind to the endomysial tissue are detected by using a microscope after they are tagged with fluorescent anti-IgA or anti-IgG antibodies. Results are qualitative or semi-quantitative.

In contrast, the anti-transglutaminase antibody test is an enzyme-linked immunosorbent assay, which is less operator-dependent and more quantitative than the immunofluorescence technique. In this method, guinea pig or human transglutaminase 2 is coated onto plastic wells, and patient serum is brought into contact with the wells. Captured serum anti–transglutaminase 2 antibodies are detected by addition of an enzyme-linked antibody against the bound IgG or IgA anti–transglutaminase 2 antibody, followed by addition of a substrate that reacts with the enzyme to produce color and measurement of the generated color by using a spectrophotometer. Use of purified or recombinant human transglutaminase 2 improves performance compared with guinea pig transglutaminase, especially with regard to specificity (68–71). Radiolabeled precipitation assay has been reported to also perform well in the detection of anti–transglutaminase 2 antibodies (72). However, this test is less widely available than enzyme-linked immunosorbent assay.

**Sensitivity and Specificity of Antibody Markers**

The sensitivity and specificity of serologic markers for celiac disease vary considerably among studies because of such factors as choice of gold standard, patient selection bias, population differences, and methodologic variability. A systematic and rigorous review of the literature on the sensitivity and specificity of serologic markers celiac disease was recently published as part of an evidence report on celiac disease by the Agency for Healthcare Research and Quality (56). The investigators used strict criteria to exclude studies with methodologic flaws and specifically included only studies that used biopsy as the gold standard diagnostic test and described the biopsy criteria. Despite wide heterogeneity among the evaluated studies, results indicate that IgA anti–transglutaminase 2 antibody and IgA antijejunal antibody have a sensitivity greater than 90% and a specificity greater than 95%. In contrast, IgA antigliadin antibody has a sensitivity of about 80% and specificity of 80% to 90%. The study also reports that IgG...
class anti–transglutaminase 2 and antiendomysial antibodies have specificities greater than 95% but poor sensitivities (around 40%). The IgG antigliadin antibody has sensitivities and specificities of around 80%. When these figures are considered, one can conclude that with the availability of tests for IgA anti–transglutaminase 2 and antiendomysial antibodies, other tests are of limited value. However, testing for IgG anti–transglutaminase 2 (or antiendomysial) and antigliadin antibodies is useful for diagnosing celiac disease in IgA-deficient persons.

Of note, no statistically significant difference was found between the human anti–transglutaminase 2 antibody and antiendomysial antibody tests (56). Therefore, either one can be considered in the initial panel of serologic tests, but the other test may be added if results of the first test are discordant with the intestinal biopsy report. At the same time, the prevalence and serum levels of anti–transglutaminase 2 and antiendomysial antibodies correlate with the degree of villous atrophy. Various studies indicate that the sensitivity of testing for anti–transglutaminase 2 and antiendomysial antibodies is decreased in villous atrophy of mild histologic grade (56, 73, 74). Therefore, serologic tests may not detect partial villous atrophy.

The Mucosal Lesion

The key triggers in celiac disease are specific immunogenic peptides of dietary gluten proteins in wheat and similar proteins in rye and barley. These peptides, which are resistant to digestion by gastric and pancreatic enzymes, find their way into the lamina propria, presumably after changes in intercellular tight junctions and increased intestinal permeability (such as may occur after gastrointestinal infection) (2, 75). The subsequent infiltration by CD4+ T cells into the lamina propria and by mainly CD8+ and CD4− CD8− T cells into the epithelium is a hallmark of active celiac disease. The function of HLA-DQ2– and
HLA-DQ8–restricted CD4+ T cells of the lamina propria in the immune response has been well studied. In genetically predisposed persons who express the HLA-DQ2 and HLA-DQ8 molecules, antigen-presenting cells process the intruding glutamine- and proline-rich gluten peptides and present them to gluten-specific CD4+ T cells. One such peptide is a 33–amino acid sequence that is resistant to digestive enzymes and is a potent activator of specific T-cell lines from patients with celiac disease (76). Recognition of HLA-bound gluten peptides by T cells leads to their activation and release of various cytokines. Some of these cytokines (released by T112 cells) drive the activation and clonal expansion of B cells that produce antibodies. Other cytokines (released by T111 cells) promote various inflammatory mechanisms, including secretion of matrix metalloproteinases by fibroblasts and inflammatory cells that can degrade the mucosal matrix and produce the intestinal lesion (77, 78). Detailed knowledge of the actual mechanism that produces the lesion is, however, limited (Figure 2).

Considerably less information is available on the activation and mode of action of intraepithelial T cells. They are known to interact with stress proteins expressed by epithelial cells and exhibit cytolytic activity that leads to destruction of the epithelium in celiac disease (79–81). Unlike the antigen-specific activation of CD4+ T cells that involves the adaptive immune response, activation of intraepithelial lymphocytes appears to be additionally mediated by the innate immune system (81–84). In particular, expression of the interleukin-15 cytokine after the innate immune response to intruding gluten peptides appears to play a central role in driving various processes that lead to intraepithelial lymphocyte–mediated destruction of epithelial cells and mucosal damage (81–83).

Transglutaminase 2 may play an important role in the immune response. In normal tissue, it catalyzes the cross-linking of specific glutamine residues to primary amines, leading to formation of isopeptide bonds within or between proteins (85–87). The cross-linking activity of transglutaminase 2 is involved in various functions, such as wound healing, formation of cell envelopes in apoptosis, and stabilization of the extracellular matrix (88). Its expression is therefore increased during tissue injury and is especially elevated in intestinal biopsy samples from patients with celiac disease (89, 90). In addition to having cross-linking activity, transglutaminase 2 can deamidate glutamine residues (91, 92). Glutamine-rich gluten peptides, such as the aforementioned 33–amino acid sequence, are therefore excellent substrates for transglutaminase 2 (76). The resulting deamidated and thus negatively charged peptides have much higher affinity for the HLA-DQ2 and HLA-DQ8 molecules that are involved in presenting them to T cells (93). This transglutaminase 2–driven modification is believed to be a key step in the immune response in celiac disease (944).

It is not yet clear how the ensuing immune reaction also targets the transglutaminase 2 molecule itself. Considering that transglutaminase 2 can form covalent complexes with gliadin, a possible hypothesis is that the anti–transglutaminase 2 immune response is generated by epitope spreading through intermolecular help, where gluten acts as a carrier protein for transglutaminase 2. Accordingly, gluten-specific T cells are proposed to help transglutaminase 2–specific B cells that produce anti–transglutaminase 2 antibodies, given that transglutaminase 2–gluten complexes are formed in vivo (93) (Figure 2). This presumed gluten-specific T-cell–driven mechanism of intermolecular help would result in an anti–transglutaminase 2 immune response in the absence of transglutaminase 2–specific T lymphocytes. The strict dependence of anti–transglutaminase 2 antibodies on gluten intake in patients seems to support this mechanism (95).

The role of autoantibodies in disease pathogenesis is controversial and varies from one disease to another. Similarly, the contribution of anti–transglutaminase 2 antibodies to the observed mucosal lesion in celiac disease is not clear. Previous findings suggest that transglutaminase 2 is required for activation of transforming growth factor-β (96), which is involved in differentiation of epithelial cells (97, 98). Therefore, local production of anti–transglutaminase 2 autoantibodies that have been shown to interfere with transglutaminase 2 bioactivity (99) might have a deleterious effect on cell differentiation, contributing to the mucosal transformation observed in celiac disease. Paradoxically, if the antibodies play an inhibitory role, they might also block the proposed role of transglutaminase 2 in driving the immune response through deamidation and cross-linking. Clearly, celiac disease is a complex disorder that results from an intricate interplay among various immunologic, genetic, and environmental factors, many aspects of which remain to be elucidated.

**Dermatitis Herpetiformis**

Gluten sensitivity is sometimes expressed in the form of dermatitis herpetiformis, a pruritic, chronic skin disease characterized by symmetrical papulovesicular lesions and presence of granular deposits of IgA in the dermal papillae. This condition affects about 10% to 20% of patients with celiac disease (100, 101). A gluten-free diet is the treatment of choice, although it may be combined with drug therapy, usually with dapsone, to effectively and quickly resolve the itching and rash. Inflammatory small-bowel changes identical to those in celiac disease accompany the skin lesions in dermatitis herpetiformis, even in the absence of gastrointestinal symptoms. The serologic antibody profile is also similar to that for celiac disease: Antigliadin as well as anti–transglutaminase 2 antibodies are present, although at lower levels, possibly reflecting a milder enteropathy (102). One study has shown the presence of antibodies exclusively against transglutaminase 3 (also known as epidermal transglutaminase), a cytosolic enzyme involved in cell envelope formation during keratinocyte differentiation (88). Of note, that study also showed that transglutaminase 3, but not transglutaminase 2, is found in complex with the IgA...
DISORDERS ASSOCIATED WITH CELIAC DISEASE

Several studies have demonstrated a close association between celiac disease and other disorders (Table 1). Celiac disease is increasingly being diagnosed in patients with predominantly extraintestinal manifestations. It is therefore important that the clinician considers the possibility of celiac disease when encountering these disorders. Symptoms that are suggestive of celiac disease should be recognized and followed by serologic testing. Some of the associated disorders, including osteoporosis, anemia, short stature, and certain reproductive problems, are generally secondary to celiac disease–related malabsorption and resolve with use of a gluten-free diet. Other major groups of associated conditions include certain endocrine disorders, cancer, and neurologic problems. In these cases, the relationship between diet and disease is more complex.

Endocrine Disorders

Celiac disease is associated with some immune-mediated endocrine disorders, most commonly type 1 diabetes and thyroid disease. Each of these conditions affects 5% to 10% of patients with celiac disease (13) (Table 1). The effect of adherence to a gluten-free diet on the metabolic control of diabetes or management of thyroid disease is limited at best (13, 104), and additional studies are clearly needed to reach firm conclusions.

Cancer

The incidence of certain types of cancer is increased among patients with celiac disease (56, 105, 106). These include non-Hodgkin lymphoma at any site, enteropathy-associated T-cell lymphoma (a rare high-grade T-cell non-Hodgkin lymphoma of the small intestine), small-intestinal adenocarcinoma, and esophageal and oropharyngeal squamous carcinoma (2). Strict adherence to a gluten-free diet seems to protect against developing some cancers (56, 105).

Neurologic Disorders

Among the most common neurologic problems associated with celiac disease are peripheral neuropathy, cerebellar ataxia, epilepsy, and migraine. In a recent study of 26 patients with celiac disease, 31% had abnormalities on neurophysiologic studies, compared with 4% of controls with reflux disease (23). Nutritional factors have been suspected in the associated neurologic deficits but are rarely found (107–109). Some reports show certain neurologic symptoms to respond to gluten-free diet, but others indicate no effect (23, 24, 109–112).

Research on the underlying mechanisms for the relationship between celiac disease and other disorders is still at a preliminary stage, even though some of the associations have been known for many years. It is now evident that the link results in part from common genetic background, most importantly the HLA region of chromosome 6 (13, 113–115). In addition to genetic predisposition, immunologic factors probably play a role. One way that this may occur is by antibody or T-cell cross-reactivity, a mechanism that is suspected of triggering the immune response in some autoimmune diseases (116–120). Alternatively, it may result from involvement of additional autoantigens through epitope spreading. Finally, the autoimmune response specific to celiac disease may be directly responsible for some of the extraintestinal manifestations. For example, considering that transglutaminase 2 plays a critical role in release of insulin from pancreatic islet cells (121, 122), an immune response against transglutaminase 2 may be involved at some point in the associated type 1 diabetes.

CONCLUSIONS

Celiac disease is a multisystem autoimmune disorder that is currently believed to affect about 1% of the general population. Although the clinical classification and diagnosis of the disease are based on gastrointestinal manifestations, patients are increasingly identified after the extraintestinal complications of the disease are detected. The clinician should therefore not only consider celiac disease in patients who are experiencing the classic gastrointestinal symptoms but also in those who have disorders whose prevalence is high among patients with celiac disease. Use of serologic markers has revolutionized the screening and diagnosis of celiac disease. Current evidence indicates that IgA anti–transglutaminase 2 and IgA anti-endomysial antibodies have good sensitivity and specificity and are superior to other markers for celiac disease. Nevertheless, confirmation of characteristic mucosal damage on intestinal biopsy remains the gold standard for diagnosis.

Substantial progress in the understanding of celiac disease has been made in the past decade. Both the adaptive and innate arms of the immune system are involved in the response to gluten and the subsequent action of lamina propria and intraepithelial lymphocytes in driving the autoimmune response that eventually leads to mucosal damage. Expression of HLA-DQ2 and HLA-DQ8 molecules is an essential genetic component of the disease, being necessary for the immune reaction against gluten. Furthermore, apart from becoming a target antigen of the immune response, transglutaminase 2 enzyme appears to be involved in modifying and enhancing the immunostimulatory effect of gluten peptides. However, many important questions remain, especially with regard to additional molecular and genetic factors that drive the immune response against gluten, the mechanism of involvement of the transglutaminase 2 autoantigen in the immune response, the underlying factors that affect the association of celiac disease with other disorders, and the role of a gluten-free diet in treating the extraintestinal complications of celiac disease. A better understanding of the underlying mechanism of pathogenesis...
of celiac disease and associated disorders will help in devising new strategies for diagnosis and treatment of the disease, including prevention of its long-term complications, and serve as a model for investigation of other autoimmune disorders.

From Cornell University and Columbia University, New York, New York.

Requests for Single Reprints: Armin Alaeddini, PhD, Department of Neurology and Neuroscience, LC-807, Cornell University, 1300 York Avenue, New York, NY 10021; e-mail, ara2004@med.cornell.edu.

Potential Financial Conflicts of Interest: Consultancies and Honoraria: P.H.R. Green (Prometheus Laboratory).

Current author addresses are available at www.annals.org.

References


Marsh MN. Gastrointest Endosc. 1990;65:909-11. [PMID: 2205160]


64. Abdulkarim AS, Burgart LJ, See J, Murray JA. Evidence that intestinal intraepithelial lymphocytes are activated by an ancient enteropathy-related autoantigen. Gastroenterology. 1997;112:1518-20. [PMID: 9178673]


Celiac Disease: Understanding a Complex Autoimmune Disorder

1351-7. [PMID: 8623906]
Current Author Addresses: Dr. Alaedini: Department of Neurology and Neuroscience, LC-807, Cornell University, 1300 York Avenue, New York, NY 10021.

Dr. Green: Department of Medicine, Room 645, Columbia University, 161 Fort Washington Avenue, New York, NY 10032.
Smoking Kills: Experimental Proof from the Lung Health Study

In 1938, Raymond Pearl reported in Science that tobacco smoking shortened the life span (1). In a study of determinants of longevity in East Baltimore families, he developed life tables for smokers and nonsmokers, using data from 6813 men, most of whom were smokers. Pearl showed that life span after age 30 years for white men was reduced by about 10 years in “heavy smokers” compared with nonsmokers (Figure). Longevity was also lower for “moderate smokers.” In retrospect, this powerful observation received surprisingly little attention, even though an effect of this magnitude on total mortality, a crude but integrating measure of population health, must have come from strong increments in risk for death from specific diseases. George Seldes, an investigative reporter who championed tobacco control, attributed the limited media coverage of the finding to the influence of the tobacco industry (2).

Further studies of smoking and overall mortality were not reported until the 1950s, when early findings of cohort studies initiated to prospectively investigate the risks of lung cancer and other diseases associated with smoking were published. By 1964, the first U.S. Surgeon General’s Report on smoking and health (3) had reviewed the results of 7 cohort studies, including the study of British physicians for which 50-year findings were recently reported (4). All of the studies showed increased risk for dying in smokers compared with nonsmokers, and some studies showed that the risk increased with the number of cigarettes smoked. The 1964 report carefully discussed the possibility that either biased selection of participants or failure to control for the effects of confounding factors accounted for the negative association of smoking with longevity. The report evaluated these possibilities qualitatively and quantitatively and set them aside. The Surgeon General’s Advisory Committee concluded that “cigarette smoking contributes substantially to mortality from certain specific diseases and to the overall death rate.”

Even after publication of this report, however, the tobacco industry and its consultants continued to argue that smoking had not been established as a “cause” of lung cancer and other diseases (5). These critics characterized epidemiologic evidence as insufficient to infer causality and as subject to selection bias and confounding; they insisted that experimental evidence was the accepted standard for establishing causation. In principle, a randomized clinical trial of active smoking would exclude the possibility of selection bias and confounding, since the smoking status of individuals would depend on random assignment rather than on personal characteristics that could determine the risk for disease and death. A randomized trial of smoking is neither feasible nor ethical. An alternative strategy is to study the effect of stopping smoking by conducting a randomized trial of smoking cessation. Such trials have involved random assignment of smokers to an intensive intervention group or to a control group of “usual” management. Lower mortality rates in the intervention group would provide strong and incontrovertible evidence that smoking caused increased mortality, since removing the causal factor at a greater rate from 1 of 2 comparable groups resulted in lower mortality. This result would also prove that smoking cessation decreases mortality rates. Among the trials of smoking cessation, few have had enough participants or long enough follow-up to accurately measure the risks for disease and death.

In this issue, Anthonisen and colleagues (6) describe their findings from more than 14 years of follow-up of participants in the Lung Health Study, which was of sufficient size and duration to test whether smoking cessation reduces mortality. This randomized trial tested the hypothesis that smoking cessation and inhaled bronchodilator therapy improved the early natural history of chronic obstructive pulmonary disease (7). After 5 years of follow-up, the rate of smoking cessation was higher for those randomly assigned to the special intervention group (21.7%) than for those assigned to the usual care group (5.4%). With continued follow-up to 14.5 years, mortality rates from all causes, lung cancer, and cardiovascular disease were higher in the usual care group. The characteristics of the 2 groups were similar at baseline (7), and the findings support the inference that reduction of smoking caused a reduction in mortality. In fact, the reductions were substantial for causes of death that have long been linked to smoking: lung cancer, respiratory disease, and coronary heart disease.

The Lung Health Study is not the first trial to measure the effect of smoking cessation on mortality. The Multiple Risk Factor Intervention Trial (MRFIT) was a randomized trial of risk factor reduction strategies in 12,866 middle-aged men at high risk for cardiovascular disease (8). The intervention arm included counseling for smoking cessation, and more men in the intervention group stopped smoking compared with the usual care group. However, the intervention did not reduce either coronary heart disease deaths or total deaths. The authors attributed this “negative” finding to an unexpectedly high level of risk factor reduction in the usual care group, which reduced the effect of the MRFIT interventions. The R.J. Reynolds Tobacco Company used this “negative” finding in a 1985 advertisement but agreed to desist after the Federal Trade Commission cited it for making false and misleading statements (9).

Beyond the findings of the clinical trials, observational studies provide massive and incontrovertible evidence that smoking is linked to death from all causes and from specific causes (10–13). In general, mortality risks of smoking increase with numbers of cigarettes smoked and decrease...
progressively with increasing number of years after stopping smoking (10, 12). Cessation has benefits at all ages (12). The risks for cardiovascular diseases drop more quickly after cessation than do risks for cancer, but nonetheless lung cancer mortality rates were lower in the Lung Health Study’s intervention group with 14.5 years of follow-up. In many countries, declining mortality rates from coronary heart disease and from lung cancer in men reflect trends of smoking cessation, providing population-level evidence of its benefit.

Given the extensive evidence available and the long-accepted conclusion that smoking causes increased mortality rates, why are the new findings from the Lung Health Study noteworthy? From a historical perspective, they provide further “proof,” based on experiment rather than observation, that smoking is causally responsible for the increased risk for death in smokers. No one can make a serious claim to the contrary in light of this randomized trial evidence. The Lung Health Study findings also offer a reminder that smoking kills middle-aged people; persons who die in middle age lose over 23 years of life on average (14). In fact, if we are to begin to control the rising numbers of smoking-related deaths soon, we must increase rates of smoking cessation now, since we won’t see the benefits for decades. The new results from the Lung Health Study confirm again that smoking cessation prolongs life.

These findings have clinical as well as public health implications. For clinicians, they offer a reminder of the benefits of smoking cessation. The clinical toolbox includes few interventions that are certain to have immediate benefits for life span. Interventions by clinicians do increase the rates of successful quitting, and clinicians should follow recommended guidelines by obtaining a smoking history from all patients and assisting smokers in quitting (15). Mentioning the powerful new findings from the Lung Health Study may motivate some smokers to stop.

Jonathan M. Samet, MD, MS
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD 21205

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Jonathan M. Samet, MD, MS, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Room W6041, Baltimore, MD 21205; e-mail, jsamet@jhsphs.edu.


References
2. Seldes G. The suppressed story of tobacco. IN FACT. 1942;VI:10.
13. U.S. Department of Health and Human Services. The Health Effects of

Figure. The survivorship lines of life tables for white males falling into three categories relative to the usage of tobacco.


© 2005 American College of Physicians
“Practice Makes Perfect” . . . Or Does It?

The current public focus on health care quality is mobilizing payers and regulatory bodies alike to measure quality of care and to link quality with physician reimbursement through “pay for performance” (1). For the medical profession to address the public’s concern and improve its care of patients, it must understand the determinants of quality so that physicians can model and emulate predictors of good quality and recognize and remedy predictors of poor quality.

To address one potential predictor of quality, in this issue Choudhry and colleagues (2) systematically reviewed data relating experience and age to physician performance. Surprisingly, 70% of studies demonstrated a negative association between length of time in practice and several measures of good physician performance. Although the magnitude of the association varied, this general relationship held for medical knowledge, adherence to nationally accepted guidelines and standards, and patient outcomes. The negative association between experience and measures of quality initially surprised us, but its consistency across specialties (including internal medicine), across measures of performance, and across many studies spanning several decades sends a wake-up call to the medical profession. Choudhry and colleagues acknowledge that attributes not evaluated in the studies, such as humanism and judgment, might improve with experience. However, even if these traits improve with experience, the most caring and wise physicians, regardless of their age, would be even better with current information about accepted clinical practice and systems of care that support such practices.

The profession cannot ignore this striking finding and its implications: Practice does not make perfect, but it must be accompanied by ongoing active effort to maintain competence and quality of care. The Physician Charter on Medical Professionalism reminds physicians that this commitment to strive for excellence is a critical responsibility underlying the medical profession’s contract with society (3). However, Choudhry and colleagues’ study informs us that we cannot maintain competence passively through accumulating clinical experience (4). New educational programs for physicians must incorporate principles of adult learning and link education and clinical care within the time constraints and productivity pressures of practice.

To address one potential predictor of quality, in this issue Choudhry and colleagues (2) systematically reviewed data relating experience and age to physician performance. Surprisingly, 70% of studies demonstrated a negative association between length of time in practice and several measures of good physician performance. Although the magnitude of the association varied, this general relationship held for medical knowledge, adherence to nationally accepted guidelines and standards, and patient outcomes. The negative association between experience and measures of quality initially surprised us, but its consistency across specialties (including internal medicine), across measures of performance, and across many studies spanning several decades sends a wake-up call to the medical profession. Choudhry and colleagues acknowledge that attributes not evaluated in the studies, such as humanism and judgment, might improve with experience. However, even if these traits improve with experience, the most caring and wise physicians, regardless of their age, would be even better with current information about accepted clinical practice and systems of care that support such practices.

The profession cannot ignore this striking finding and its implications: Practice does not make perfect, but it must be accompanied by ongoing active effort to maintain competence and quality of care. The Physician Charter on Medical Professionalism reminds physicians that this commitment to strive for excellence is a critical responsibility underlying the medical profession’s contract with society (3). However, Choudhry and colleagues’ study informs us that we cannot maintain competence passively through accumulating clinical experience (4). New educational programs for physicians must incorporate principles of adult learning and link education and clinical care within the time constraints and productivity pressures of practice.

Fourth, internal medicine should be a leader in advancing the idea that educating physicians to improve quality of care is a multidimensional challenge (5, 6). Delivering information to physicians is only one component of the educational process; assisting physicians with translating that knowledge into high-quality care is even more important. Professional organizations can provide this “translational education” by showcasing best-practice models for delivery of care and by simultaneously providing resources, tools, and guidance that help physicians adapt such models to their specific practices.

Finally, all physicians, not just those with time-limited certification, must embrace the concepts behind maintenance of certification, which provides an opportunity to prevent the outcomes demonstrated in Choudhry and colleagues’ study. The American Board of Medical Specialties’ framework for Maintenance of Certification (MOC) abandons one-time certification for life, replacing it with ongoing measurement of performance. This model of professional development provides a way to identify gaps between current and ideal practice, which is the first step toward acquiring needed new knowledge, skills, or processes of care (7, 8). This model promotes the habit of continuing improvement not only for individual physicians but also for the systems and practice settings in which they work. Maintenance of certification is not an end in itself but a way to achieve the goal of better care and better outcomes for patients.

An important innovation in the MOC framework is self-evaluation of actual performance in practice, which incorporates components of measuring and improving quality of care. The American Board of Internal Medicine’s (ABIM’s) Pract-
Practice Improvement Modules have been designed to link practice self-evaluation with relevant education and information for practice redesign. However, the Practice Improvement Modules are not the only method for practice assessment, and the ABIM program will recognize several approaches to provide evidence of participation in practice performance measurement and improvement.

As we move forward with a shared goal to improve patient care, certifying organizations and specialty societies must work together to help physicians address the challenge identified by Choudhry and colleagues—how to maintain the highest quality of care throughout a career that spans several decades. Recent collaborative efforts between the ABIM and the American College of Physicians relating to maintenance of certification and quality improvement have demonstrated that this approach is both feasible and effective. Using a set of principles that have been mutually agreed upon, our organizations are committed to linking quality of care, evaluation of performance, and lifelong education in real-world practice. Choudhry and colleagues' article is a reminder that being a good physician takes continuous effort and that practice without reflection and striving for continued improvement is a formula for mediocrity. Although Choudhry and colleagues' article challenges the individual physician, internists have their 2 professional organizations working together to help them integrate lifelong learning, practice assessment and improvement, and change in the practice environment. The goal is continuous professional development and public accountability for the highest-quality patient care.

Steven E. Weinberger, MD
American College of Physicians
Philadelphia, PA 19106

F. Daniel Duffy, MD
Christine K. Cassel, MD
American Board of Internal Medicine
Philadelphia, PA 19106

Acknowledgments: The authors thank John Tooker, MD, MBA, and Eric B. Larson, MD, MPH, for their helpful suggestions.

Potential Financial Conflicts of Interest: None disclosed.


Current author addresses are available at www.annals.org.


References

© 2005 American College of Physicians
Current Author Addresses: Dr. Weinberger: American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106-1572.

Drs. Duffy and Cassel: American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, PA 19106-3699.
She Is a Beautiful Lady

Donna, growing frustrated at not having the mere privacy of spending even a few waking moments in Donna’s absence, remained at home. Dean began to feel quite insecure after still possessive of him, both determined that they would memory and activity levels. He remained pleasant, Donna care staff, not surprisingly, relayed to me Dean’s worsening agreed for home-based primary care services. The home might wonder aloud if the Mayo Clinic perhaps could help myself. With a nonchalant and determined voice, Donna wished I had Dean’s incredible capacity for detachment such a detached stance. There were moments when I often it would baffle me as to how he could watch the whole process of these discussions centering on him with such a detached stance. There were moments when I wished I had Dean’s incredible capacity for detachment myself. With a nonchalant and determined voice, Donna might wonder aloud if the Mayo Clinic perhaps could help Dean’s memory.

With understandable reluctance, Donna eventually agreed for home-based primary care services. The home care staff, not surprisingly, relayed to me Dean’s worsening memory and activity levels. He remained pleasant, Donna still possessive of him, both determined that they would remain at home. Dean began to feel quite insecure after spending even a few waking moments in Donna’s absence. Donna grew frustrated at not having the mere privacy of her bathroom. At such times, Dean would scream for her like a toddler searching for his parent. From a charming, well-sculpted lady who almost always prompted a second gaze from passersby, Donna had become the personification of an elderly, fatigued, and burdened lady with a single mission in life, taking care of Dean.

But Donna never reported any of this during their clinic visits. She had heard enough of friendly reprimands about caregiver burnout, support groups, and specially designed programs for such patients and their families. Somewhat miraculously, however, she permitted us to offer Dean respite care provision in our nursing home care unit for 2 weeks. Dean, with his memory impairment, had no perceivable problems. When Donna visited him 3 days after his admission, she was distraught at his response to her presence. She seemed quite disappointed that he did not miss her or panic in her absence. Worse still, he seemed only vaguely familiar with her and would not call her by her name. Donna’s emotions flooded. There was no stopping her. She complained that he had grown worse with the hospital stay and immediately took him home.

Soon Dean lost the memory of his relationship with Donna, which in turn deflated her passion and purpose in life. Dean began to miss his follow-up appointments. The family reported that Donna had an unexpected fall at home. She was hospitalized, and a few days later, she passed away.

Their daughter told me this: “Dad had no reaction and seemed not to acknowledge Mom’s death. We explained it to him several times. My 10-year-old seemed to know much more of what had happened than my Dad. Mom was embalmed and looked so very beautiful. We draped her in a gorgeous olive green satin dress with sequins and rich embroidery. She looked so stunning. We took Dad to her casket. He appeared perplexed. He was vaguely familiar with her face, but could not recognize his life partner of 61 years when she was still and motionless, her presence. She seemed quite disappointed that he did not miss her or panic in her absence. Worse still, he seemed only vaguely familiar with her and would not call her by her name. Donna’s emotions flooded. There was no stopping her. She complained that he had grown worse with the hospital stay and immediately took him home.

Sara Sasha Battar, MD
VA Central Iowa Health Care System
Knoxville, IA 50138

Requests for Single Reprints: Sara Sasha Battar, MD, VA Central Iowa Health Care System, Room 111 Building 3, 1515 West Pleasant, Knoxville, IA 50138; e-mail, sara.battar@med.va.gov.

COMMENTS AND RESPONSES

Cardiac Resynchronization Therapy in Heart Failure

TO THE EDITOR: McAlister and colleagues (1), in their systematic review, concluded that cardiac resynchronization therapy (CRT) reduces all-cause mortality and heart failure hospitalizations when added to medical therapy in selected patients with heart failure. Among some U.S. cardiologists, there appears to be a perception that patients with heart failure should receive a device with defibrillator ability and biventricular pacing function. It is likely that the report by McAlister and colleagues will further contribute to this perception and possibly influence practice trends.

McAlister and colleagues used a comprehensive search strategy and intensive extraction and evaluation of data from pertinent studies. However, their analysis could not overcome certain methodologic deficiencies (for example, concealed randomization, lack of blinding or partial blinding, and intention-to-treat analyses) that affected several of the studies. Such deficiencies can lead to preferential management or treatment, resulting in better or improved outcomes that favor the intervention of interest (in this case, CRT). In addition, the crossover design of several studies further restricted the already limited data available for evaluation of mortality rates and hospitalizations.

Among the parallel studies noted in McAlister and colleagues’ Figures 2 and 3—the Multicenter InSync Randomized Clinical Evaluation (MIRACLE) (2), the Multicenter InSync Randomized Clinical Evaluation ICD [implantable cardioverter defibrillator] (MIRACLE-ICD) (3), and the Comparison of Medical Therapy, Pacing, and Defibrillation in Chronic Heart Failure (COMPANION) trial (4)—only MIRACLE (2) and MIRACLE-ICD (3) had high-quality methods. These 2 trials revealed small, statistically insignificant differences in mortality rates between patients randomly assigned to CRT and those assigned to inactive-device (no cardiac resynchronization) therapy. Of interest, subgroup analyses in the COMPANION trial (4) suggested a beneficial effect of CRT—defibrillator therapy on death rates in patients taking an angiotensin-converting enzyme inhibitor, a β-blocker, or spironolactone. This beneficial effect was not seen in patients who were not taking these drugs.

In summary, McAlister and colleagues’ superb effort might be undermined by the weak quality of several studies used in the various analyses. McAlister and colleagues’ conclusions regarding the use of CRT in patients with symptomatic heart failure should be assessed and confirmed in large trials without methodologic deficiencies.

Enrique V. Carbajal, MD
Grace W. Huang, DO
Billy Hu, MD
VA Central California Health Care System
Fresno, CA 93703

References

TO THE EDITOR: McAlister and colleagues (1) concluded that in selected patients with heart failure, CRT improves functional and hemodynamic status, reduces heart failure hospitalizations, and reduces all-cause mortality. Such conclusions, however, are unsafe and are based on incorrect presentation of the existing scientific data.

A major problem with McAlister and colleagues’ analysis is that it does not deal appropriately with potential differences in the efficacy of CRT delivered with or without an ICD. Of particular concern is the presentation of the all-cause mortality data from the COMPANION trial, which fails to preserve randomization. This ignores a basic tenet of meta-analysis (2, 3). The authors incorrectly split the control arm results in 2 and compared half with the results from CRT alone and the remaining half with results from the CRT-ICD arm. In doing so, they created decreases in all-cause mortality for both active treatment groups. However, in an appropriate and unconfounded analysis that compared only the CRT-alone group (n = 617) with the control group (n = 308) and preserved randomization, we found a nonsignificant reduction in mortality attributable to CRT (relative risk, 0.85 [95% CI, 0.67 to 1.09]) (Figure 1). Indeed, it is the inappropriate inclusion of the data for the CRT-ICD group that led to the appearance of an overall statistically significant pooled reduction in all-cause mortality. Removal of these confounded data leads to a nonsignificant pooled decrease in the relative risk for all-cause mortality (relative risk, 0.84 [CI, 0.69 to 1.03]). Clearly, there are still important unanswered questions about the impact of CRT and the additional impact of CRT-ICD on mortality.

Second, McAlister and colleagues’ analysis of heart failure hospitalizations was inconclusive. Their pooled data provided an estimate of 0.68 (CI, 0.41 to 1.12) for the relative risk for hospitalization. This result was based on only 316 events from proof-of-concept studies designed to assess whether CRT could improve symptoms, ventricular function, and exercise capacity. Although this result was nonsignificant, the authors then performed a subgroup analysis, on which they based their conclusion that CRT reduces heart failure hospitalizations in selected patients. The authors failed to include results from the large COMPANION trial, presumably because heart failure hospitalization was reported as part of a composite outcome combined with death from heart failure (2). It is unfortunate that the authors were unable to include results from this large study.

Finally, the authors acknowledged that the trials in their analysis probably overestimate the potential benefits of CRT. In all but 1 trial, only patients who underwent successful device implantation and survived a run-in period were eligible for randomization. Moreover, 2 patients died within hours of crossing over to CRT in the Multisite Stimulation in Cardiomyopathies (MUSTIC) trial (4). For all of these reasons, the encouraging trends observed in meta-analyses presented so far are not secure evidence of benefit in morbidity or mortality. The double-blind MIRACLE and MIRACLE-ICD trials do suggest that CRT improves symptoms in about 35% and 15% of
patients, respectively, compared with controls (5, 6). However, further manipulation of pharmacologic therapy might have achieved similar results. At least 2 other large controlled trials, involving more than 1200 patients, have yet to be reported (7, 8). Although McAlister and colleagues’ meta-analysis aimed to address an important question, the efficacy and safety of CRT, it failed to provide substantive evidence of this therapy’s effect on mortality and hospitalization due to heart failure.

Melanie Calvert, PhD
Nick Freemantle, PhD
University of Birmingham Medical School
Edgbaston B15 2TT, England, United Kingdom

John G.F. Cleland, MD
University of Hull
Kingston upon Hull HU16 5JQ, England, United Kingdom

Note: Drs. Freemantle and Cleland are steering committee members for the CArdiac REsynchronisation in Heart Failure (CARE-HF) trial. Drs. Freemantle and Calvert are responsible for the analysis of the CARE-HF trial.

References


IN RESPONSE: We agree with Drs. Carbajal, Huang, and Hu that several trials included in our systematic review had potential methodologic deficiencies. Although we had space to highlight only 1 of these deficiencies in our published manuscript (randomization after implantation of the CRT device in all but 1 trial), we refer interested readers to our full Agency for Healthcare Research and Quality report, available at www.ahrq.gov/clinic/tp/resyntp.htm, for a full description of the methods and the quality assessment tables for all of the studies included in both our efficacy and safety analyses.

We refute the allegation of Drs. Calvert, Freemantle, and Cleland that our analysis of the COMPANION trial data (1) “ignores a basic tenet of meta-analysis.” As outlined in our manuscript, we did conduct intention-to-treat analyses for all end points, and halving the control group data when incorporating a 3-arm trial into a meta-analysis is one of the techniques endorsed by the Cochrane Collaboration, as outlined at www.cochrane-net.org/openlearning/HTML/modA2-5.htm. Indeed, we would like to point out that, despite Calvert and colleagues’ assertion that we created incorrect decreases for the CRT-alone arm of the COMPANION trial, careful perusal of our Figure 2 reveals that the relative risk we reported for mortality in that arm (0.84 [CI, 0.61 to 1.14]) is not statistically significant and is in fact very similar to the relative risk of 0.85 (CI, 0.67 to 1.09) that they found in their analysis.

On the other hand, Dr. Calvert and colleagues state that the heart failure hospitalization data are inconclusive. Certainly, this is true of the pooled estimate that we reported and that they reiterate in their letter. However, in reexamining our data, we discovered an error in Figure 3 in our published manuscript. We mistyped the number of “no deaths” in the CRT group for the Guidant CONTAK CD CRT-D System Trial (5) when entering the data into our models. A corrected Figure is included here (Figure). The corrected pooled relative risk for heart failure hospitalizations is thus 0.67 (CI, 0.48 to 0.92), a result that is statistically and clinically significant and does not include the COMPANION trial results (the data for this end point were not available at the time of our request to the COMPANION investigators).

While we disagree with the specific points that Dr. Calvert and colleagues raise in the first half of their letter, we do not disagree that the trials in our analysis may have overestimated the benefits of CRT. Indeed, we made this very point in our manuscript. We also agree with these CARE-HF investigators that our meta-analysis (which is based on only 429 deaths and 251 heart failure hospital-
izations) should not be used as a reason to prematurely terminate ongoing large trials in this area (such as CARE-HF and RAFT [Resynchronization/Defibrillation for Advanced Heart Failure Trial]). In closing, we would like to reiterate 3 key points from our original study that we fear some readers may have missed: 1) CRT is not a panacea; 2) CRT is efficacious in selected patients with advanced systolic heart failure and evidence of electromechanical dysynchrony; and 3) attempts to define which patients are most likely to benefit from CRT should remain a research priority.

Finlay A. McAlister, MD, MSc
University of Alberta Hospital
Edmonton, Alberta T6G 2R7, Canada

Natasha Wiebe, Mmath
University of Alberta Hospital
Edmonton, Alberta T6G 2J3, Canada

William Abraham, MD
Ohio State University
Columbus, OH 43210

References

Relative Cost-Effectiveness of Different Tests for Chlamydia trachomatis

TO THE EDITOR: I read Hu and colleagues’ cost-effectiveness analysis for Chlamydia trachomatis screening (1) with great interest. One topic that was not explicitly covered but could easily be addressed by the same decision model is the relative cost-effectiveness of different laboratory tests for C. trachomatis detection. Nucleic acid amplification methods, which Hu and colleagues used in their analysis, are substantially more sensitive than nonamplified DNA methods, which in turn are much more sensitive than the older enzyme immunoassay methods (2, 3). From the individual patient perspective, then, the amplified tests are clearly optimal. However, many laboratories continue to offer these other less sensitive tests (4), a phenomenon that appears to be driven by cost. If it could be demonstrated that amplified DNA tests were cost-effective (or, even better, cost-saving) versus nonamplified tests from a health system perspective, then perhaps payers could be convinced to cover the modestly increased cost of amplified testing for C. trachomatis. This would benefit both individual patients and the public health.

Brian Jackson, MD
University of Utah and ARUP Laboratories
Salt Lake City, UT 84108

Potential Financial Conflicts of Interest: Dr. Jackson is employed by ARUP Laboratories, which performs testing for Chlamydia trachomatis.

References

IN RESPONSE: We appreciate Dr. Jackson’s interest and comment. We believe that our model has the potential to help inform many questions relating to chlamydia screening and chose to start by exploring the impact of different approaches to routine screening for young women from a long-term, societal perspective. Our analysis was intended to inform broad recommendations for national screening guidelines, with particular emphasis on optimal target age range and frequency. At the same time, a wide variety of diagnostic tests are available for Chlamydia trachomatis detection, including cell culture, antigen-detection tests, nucleic acid hybridization tests, and, most recently, nucleic acid amplification tests. Compared with nonamplified tests, nucleic acid amplification tests have been demonstrated to have superior sensitivity and greater acceptability among adolescents and young adults (1–3), although at a higher cost. As pointed out in

www.annals.org
Stamm’s commentary (4), many public health–based screening programs have limited resources and consequently are able to offer screening to less than half of the target population. This fact illustrates an important distinction between the cost-effectiveness (that is, “value for money”) of an available technology from a societal perspective and the affordability of that technology from the perspective of one particular payer (for example, a public health clinic). An analysis that comparatively evaluates a wide array of available screening tests and considers a shorter time horizon, while explicitly taking into account the available budget, would be useful for regional and local decision making. Such analyses are complex: To accurately reflect the tradeoffs associated with different tests, data on the likelihood of adherence to different tests and the correlation between adherence and preferences would be required. We agree that this type of analysis is of high priority.

Sue J. Goldie, MD, MPH
Delphine Hu, MD, MPH
Harvard School of Public Health
Boston, MA 02115

Edward W. Hook III, MD
University of Alabama at Birmingham
Birmingham, AL 35294

Potential Financial Conflicts of Interest: Dr. Hook has received research support from Abbott Laboratories; Roche Molecular Systems, Inc.; Gen-Probe; and Becton, Dickinson, & Co. and honoraria from Abbott Laboratories and Gen-Probe.

References

Treating Controls in Unblinded Trials

TO THE EDITOR: In the study by Yardley and colleagues (1) on vestibular rehabilitation for patients with chronic dizziness, it is not clear what scientific information was gained by crossing controls over to vestibular training at 3 months. Efficacy could not be measured, since the intervention group was not crossed over to the control condition (nor could it be, given an inability to wash out the initial vestibular training). Maintenance of effect at 3 months was already measurable in the intervention group, so a similar 3-month period for controls would not add new information. And since vestibular training was not the standard of care, a “staggered-start” trial to measure adverse effects during a 3-month control window was not ethically mandated.

One nonscientific reason for offering controls vestibular training may have been to reduce the likelihood that they would drop out of the trial because of disappointment at not receiving the training. Because the trial was unblinded, controls would have been aware of their control status. These patients, who suffered from chronic dizziness, may have entered the trial in hopes of receiving the vestibular training. Offering them such training after 3 months may have served as an inducement to them to remain in the trial as controls.

There are, however, 2 concerns about offering controls delayed intervention in an unblinded trial without valid scientific reasons for doing so. First, such an offer might be considered potentially coercive. But more important, it portrays the study intervention as “treatment” rather than as something being investigated under equipoise concerning its efficacy (2, 3). An honest null hypothesis at this trial’s inception must have allowed for the possibility that vestibular training would have proven ineffective at 3 months compared with the control condition. Given this possibility, what is the justification for providing this “treatment” to controls at 3 months as part of the trial?

Paul S. Heckerling, MD
University of Illinois at Chicago
Chicago, IL 60612

References

IN RESPONSE: Dr. Heckerling argues that by offering vestibular rehabilitation to patients in the control group after 3 months, we implied that vestibular rehabilitation is an effective treatment, thereby violating the important principle of equipoise. However, in a pragmatic (1) or phase III trial such as ours, it is assumed that the treatment has already been shown to be beneficial under ideal conditions (normally, this will have been established in efficacy or phase II trials). Consequently, equipoise is maintained instead with regard to the research question: Is the treatment more effective than usual care in typical clinical practice? Moreover, it is accepted that clinician and patient attitudes toward the treatment will affect outcomes. It was important for this reason (as well as to permit fully informed consent) that we explained to potential participants that there was some existing evidence of efficacy but no previous demonstration of effectiveness, particularly when vestibular rehabilitation was delivered by practice nurses in a primary care setting. We therefore also drew attention in our paper to the likelihood that positive motivation and placebo effects may have contributed to outcomes, although the pattern of findings was indicative of specific effects on the balance system rather than simply a generalized improvement in subjective well-being. However, if controls had stayed in the study simply because they were “coerced” by our offer of a treatment of unknown effectiveness after 3 months, then substantial dropout might have been expected at follow-up in both control and intervention groups. On the contrary, dropout at 3 months was as low in the treatment group as in the control group and remained similarly low in both groups at 6 months.

Regarding Dr. Heckerling’s subsidiary point, we concur that
Debilitating Muscle Cramps after Teriparatide Therapy

TO THE EDITOR: Background: Teriparatide, a recombinant human 1–34 amino acid sequence of parathyroid hormone, was recently approved in the United States for the treatment of men and postmenopausal women at high risk for osteoporotic fracture and in Europe for the treatment of postmenopausal women with osteoporosis. When given by once-daily injection, teriparatide increases bone mass by stimulating formation of new bone, resulting in the restoration of bone architecture (1). The adverse reactions that have been described are relatively mild. However, we observed a previously unreported reaction to teriparatide.

Case Report: Teriparatide therapy was started in a 75-year-old woman with progressive osteoporosis (compression fracture at T8 and T-scores ranging from −3.0 to −4.2) who had been previously treated with antiarthritic therapy. The patient received careful instruction on proper use of the medication and on its adverse effects. At initial use, within 5 minutes of self-administration, the patient developed a severe, debilitating lower back spasm and discomfort. Lower back pain lasted approximately 60 minutes, during which time the patient was seen in a local emergency department and required narcotics for pain relief. The patient was subsequently seen in our endocrinology outpatient clinic, where medical personnel administered teriparatide. Findings on physical examination were unremarkable. Under observation and monitoring, the patient developed recurrent lower back pain 6 minutes after administration of 20 μg of teriparatide. Back examination revealed tense and tender paraspinal muscles in the lumbar region. Findings on the remainder of the musculoskeletal examination were unremarkable. Cramping and lower back discomfort lasted approximately 30 minutes, and intravenous narcotics were again required for pain relief. With rest and analgesic therapy, the patient was able to walk 1 hour after the incident.

Conclusion: Although leg cramps have been reported with the use of teriparatide, to our knowledge this is the first description of severe muscle spasm occurring within minutes of the injection of this hormone. No precipitating factors were seen in our patient. However, one of the inactive ingredients of teriparatide is metacresol, a preservative that has been linked to local and systemic reactions to insulin injections (2). Whether metacresol played a role in our patient’s adverse reaction requires further investigation.

Parul J. Kakharia, MD
George Washington University Hospital
Washington, DC 20037

David J. Nashel, MD
Eric S. Nylen, MD
Veterans Affairs Medical Center
Washington, DC 20472

References

Statins and Nasal Polyps

TO THE EDITOR: Background: Chemotaxis of eosinophils is the crucial event in the development of nasal polyposis related to chronic hyperplastic eosinophilic sinusitis (1, 2). Several statins, including atorvastatin, inhibit T-helper 1 differentiation and induce T-helper 2 polarization and production of T-helper 2 cytokines, such as interleukin (IL)-5, that promote the activation and chemotaxis of eosinophils.

Methods and Findings: A 57-year-old woman with nonallergic rhinosinusitis (low serum levels of total and specific IgE, negative results on skin prick tests, and neutrophils and Staphylococcus aureus in nasal lavage), asthma, and no aspirin sensitivity presented in May 2002 for recent onset of rhinosinusitis associated with asthma attacks. She had no systemic disease. After a 2-week course of amoxicillin–clavulanate and long-term treatment with intranasal and inhaled corticosteroids plus inhaled long-acting β2-agonist, the patient’s condition nearly normalized.

Six months later, the patient returned because of severe persistent nasal obstruction with extensive bilateral polypsis, mild airway obstruction, and increased levels of exhaled nitric oxide despite long-term treatment with intranasal and inhaled corticosteroids. Eosinophilia was noted in nasal scrapings and in peripheral blood (969 eosinophils/mL). The FEV1, exhaled nitric oxide, and nasal cytologic characteristics throughout follow-up. Computed tomography showed extensive opacification of all paranasal sinuses; right maxillary and frontal sinuses were completely stuffed by polypoid tissue.

An otolaryngologist surgically removed the polyps, and histologic examination showed relevant edema and inflammatory infiltrate, with predominant eosinophils. Nasal polyps returned within 1 month after surgery and were associated with persistent eosinophilia. The patient reported taking atorvastatin, which her practitioner had prescribed in October 2002 for hyperlipemia. Three weeks after atorvastatin withdrawal, nasal symptoms had dramatically improved, nasal polyps had disappeared, and only a few neutrophils and eosino-
phils were found in nasal scrapings. Suspecting that atorvastatin was responsible for the development of eosinophilic polyposis, we asked the patient to resume atorvastatin therapy. Nasal symptoms and polyps recurred shortly thereafter, together with nasal eosinophilia. The patient improved again soon after drug withdrawal. Of note, the patient returned 6 months later for recurrence of polyps: Her general practitioner had prescribed simvastatin. We have seen 2 other patients, one 68 years of age and one 74 years of age, who had nonatopic chronic obstructive pulmonary disease and reported having severe nasal obstruction since the start of statin therapy (6 months and 6 years, respectively). Both patients were former smokers. In both cases, nasal lavage showed eosinophilia and nasal biopsies confirmed the presence of eosinophilic nasal polyposis. Polyps resolved in both patients after statin withdrawal.

Discussion: To our knowledge, this is the first report of nasal polyposis associated with statin treatment. Growing evidence suggests that statins, besides their lipid-lowering effect, also have anti-inflammatory and immunomodulatory properties (3–5). These properties involve inhibition of T-helper 1 differentiation and T-helper 2 polarization, with increased production of T-helper 2 cytokines (IL-4, IL-5, and IL-10) that promote activation and chemotaxis of eosinophils. The latter mechanism is the crucial event in the development of chronic hyperplastic eosinophilic sinusitis (1, 2).

Conclusion: Statins may lead to development of eosinophilic polypoid rhinosinusitis by switching inflammation from the T-helper 1 phenotype to the T-helper 2 phenotype.

Caterina Bucca, MD
Andrea Marsico, MD
Erika Panaro, MD
University of Turin
10128 Turin, Italy

Patrizia Bigo, MD
Luisa Brusino, MD
St. Giovanni Battista Hospital
10126 Turin, Italy

References

Correction: Cardiac Resynchronization in Patients with Symptomatic Heart Failure

In a review on cardiac resynchronization in patients with symptomatic heart failure (1), Figure 3 contained errors. The relative risk for heart failure hospitalizations in the Guidant CONTAK CD CRT-D System Trial should have been reported as 0.82 (95% CI, 0.53 to 1.26), and the pooled relative risk should have been reported as 0.67 (CI, 0.48 to 0.92).

Reference
BOOK NOTES
Structured abstracts of information on newly published books, computer programs, selected Web sites, and other material are provided in this portion of Medical Writings. Order phone numbers and Web sites can be used to place orders directly with publishers.

Medical Malpractice: A Physician’s Sourcebook

Field of medicine: All fields.

Audience: Practitioners desiring to minimize the risk for becoming a malpractice defendant or testifying as an expert witness.

Purpose: To provide an overview of malpractice law and its implications for selected medical specialties.

Content: The 17 chapters are grouped into 4 parts: insurance, law, clinical implications, and reform prospects.

Highlights: All clinicians will benefit from understanding part I (insurance underwriting) and part II (theory of malpractice law, discovery process, and being an expert witness). Part III has chapters specific to telemedicine, emergency medicine, anesthesiology, obstetrics, and plastic surgery. Part IV argues for legislative reform of malpractice.

Limitations: Part IV’s proposals appear logical, fair, rational, valid, and efficient for patient and practitioner. However, they are unlikely to be enacted unless linked to economic concessions to the trial bar. In addition, the text omits the practical vicissitudes of conducting a malpractice defense (for example, structuring settlements that avoid National Practitioner Databank reporting, a growing practice of institutions that admit liability in return for dismissal of all individual defendants). Finally, part I would benefit from some mention of the paradoxes of the underwriting cycle.

Related reading: Curiously, no existing book clearly summarizes malpractice for physician readers. Practitioners will find the many legal treatises on malpractice litigation too dense for ready comprehension. More useful are the publications reporting the Harvard Malpractice Study that are cited in the notes to chapter 16 in Medical Malpractice. They prove the randomness of both litigation and compensation.

Reviewer: David Hsia, JD, MD, MPH, U.S. Agency for Healthcare Research and Quality, Rockville, Maryland.

Disclaimer: This review does not represent the views of any U.S. government agency.

Contemporary Catholic Health Care Ethics

Field of medicine: Ethics, philosophy, religion, and Christian bioethics.

Audience: Students of bioethics in Catholic colleges, medical and administrative staff in Catholic hospitals, chaplains, and anyone interested in the intersection of medicine and religion.

Purpose: A primer of medical ethics with pertinent Catholic teachings.

Content: The book is divided into 3 parts: “Theological Basis,” “Method,” and “Application.” The first maintains that Catholic concepts of human dignity, holism, sanctity of human life, and quality of life should contextualize bioethical decisions. Part 2 discusses deontological and utilitarian moral theories and the “principle of double effect,” axioms favored by the Church for deciding on actions that have both good and bad consequences. It also explains shifts away from previously held positions in Catholic moral theory that have occurred since the Second Vatican Council. Part 3 mostly covers contentious issues surrounding terminal care—competency, advanced directives, euthanasia, withdrawing care, medical futility, and the practical administration of hospital ethics committees. It also has chapters on genetic engineering, embryonic stem-cell research, pain management, and health resource allocations.

Highlights: “Catholic medical ethicists,” the author writes, “are united in our insistence that God’s will is to be taken seriously and that there are indeed objective moral norms.” Yet even within this common moral framework, he underscores a diversity of bioethical thought among official and unofficial Catholic moral philosophers. The author, a recognized authority, describes himself as a “ consequentialist,” and in many instances, notably contraception, he disagrees with doctrine and with other Catholic philosophers.

Limitations: A chapter on abortion is notably absent. Also missing is a systematic discussion how basic Christian virtues (faith; hope; charity; self-sacrifice; love of neighbor; and concern for the poor, the oppressed, and strangers) should inform physician and patient. The book has no examples of medical cases, so it is hard to imagine how a beginning student will understand how bioethical principles are considered and applied in reality.

Related readings: Lammers and Verhey’s On Moral Medicine: Theological Perspectives in Medical Ethics, 2nd edition (WB Erdmans, 1998), is a rich anthology of selections from Christian writings on health, illness, and medicine that gives an overview of religious thought. Pellegrino and Thomasma’s The Christian Virtues in Medical Practice (Georgetown Univ Pr, 1996) explores “the kind of person a Christian physician ought to be.”

Reviewer: David L. Freeman, MD, Caritas Carney Hospital, Boston, Massachusetts.
The Effects of a Smoking Cessation Program on Long-Term Survival in Smokers with Mild Lung Disease

What is the problem and what is known about it so far?
Smoking cigarettes increases a person’s risk for many health problems, including cancer, heart disease, and lung disease. Several studies have shown many benefits for smokers who quit. However, no studies have proven that smoking cessation programs lead to better long-term survival. Chronic obstructive pulmonary disease (COPD), a disease of the air sacs and air passages of the lungs, is one of the many health problems to which smoking contributes. People with COPD are short of breath and sometimes cough and wheeze. Damage to the lungs and symptoms slowly worsen over time, especially if people do not stop smoking. Smoking cessation is particularly important for people with COPD who want to slow the worsening of their disease.

Why did the researchers do this particular study?
To see if a smoking cessation program improved long-term survival in people with mild COPD.

Who was studied?
5887 middle-aged people with early COPD (they had abnormal results on lung tests but did not consider themselves to be sick). The people were participating in a large study called the Lung Health Study, which examined whether a smoking cessation program and treatment with a drug called ipratropium could prevent worsening of disease in people with mild COPD.

How was the study done?
The researchers assigned participants to attend a special smoking cessation program or to receive usual care. The smoking cessation program included 12 two-hour group sessions over a 10-week period. The sessions used behavior modification and nicotine gum. The researchers followed study participants for up to 14.5 years and collected information about smoking status and death from heart disease, lung disease, cancer, or any other cause.

What did the researchers find?
After 5 years, 21.7% of the people in the special program had completely stopped smoking compared with only 5.4% of people in the usual care group. However, despite the relatively small proportion of participants who stopped smoking, the death rate in the group that received the smoking cessation program was lower, by about 15%, than the rate in the group that received usual care. Not surprisingly, the benefit with respect to survival was largest among participants who actually stopped smoking: Their death rates were 46% lower than those of the people who continued to smoke.

What were the limitations of the study?
This study involved only people with early signs of COPD on lung tests. It is unknown whether the survival benefits of a smoking cessation program would also exist among people without such signs.

What are the implications of the study?
Intensive smoking cessation programs can improve long-term survival, even when successful in only a minority of patients.
Is It Cost-Effective To Treat High-Risk Cardiac Patients with Clopidogrel plus Aspirin as Opposed to Aspirin Alone?

What is the problem and what is known about it so far?
Heart attacks and strokes are often caused by blood clots that clog arteries. Blood is made up of a liquid (serum) in which many different particles (formed elements) are suspended. Formed elements include such things as red cells, white cells, and platelets. Clot formation depends largely on activities of the platelets. Platelets can respond to external signals from other parts of the body by changing their surface characteristics, such as their ability to stick to each other, forming a clot. Doctors have found that they can prevent platelets from becoming sticky by giving certain types of medication, including aspirin and clopidogrel. Researchers have recently shown that when both of these drugs are given together, they are more effective than aspirin alone in preventing heart attack, stroke, or cardiovascular death in high-risk patients. The trouble is that clopidogrel is very expensive and it is not certain that the cost of the combination can be justified.

Why did the researchers do this particular study?
To find out if it was cost-effective to treat high-risk cardiac patients with clopidogrel plus aspirin as opposed to aspirin alone.

Who was studied?
Instead of studying actual patients, the researchers used computers to see what would have happened to a simulated (imaginary) group of patients who had certain characteristics in which they were interested.

How was the study done?
The researchers calculated the costs and clinical results of treating a 64-year-old patient with clopidogrel plus aspirin for 1 year followed by aspirin alone or with aspirin alone over the patient’s lifetime. They then calculated what the results would have been for older or younger patients. Cost calculations included medication, the cost of caring for heart attacks or strokes, and the cost of treating any potential complications of taking the medications (such as severe bleeding).

What did the researchers find?
On average, patients treated with clopidogrel plus aspirin would have lived about 1 month longer than those treated with aspirin alone, but some would have developed bleeding complications that would have worsened the quality of their lives. Treating all eligible patients with clopidogrel plus aspirin would have cost $15 400 per extra year (adjusted for quality of life) compared with aspirin alone. Each additional month of therapy added slightly to life expectancy and cost slightly more. Adding 1 year of combined therapy to 250 000 high-risk Americans would have led to a total gain of 25 500 quality-adjusted years of life at a cost to society of $392 million.

What are the limitations of the study?
The analysis is only as strong as the studies on which the risk estimates were based.

What are the implications of the study?
Combined therapy with clopidogrel and aspirin, as opposed to aspirin alone, increases average life expectancy in high-risk patients at a cost comparable to that of other effective treatments that are generally judged acceptable by society.
Is an Older, More Experienced Doctor a Better Doctor?

What is the problem and what is known about it so far?
Quality health care is sometimes defined as care that delivers the best possible results by using the right decisions made at the right time in the right way for the right patients. Making the right decisions requires familiarity with the latest medical knowledge and use of good clinical judgment. It is not known whether experience in practice influences this knowledge and judgment. Younger doctors may be more knowledgeable about the latest medical advances but may have less developed clinical judgment because of their relative lack of experience. Conversely, older doctors who have practiced medicine for many years may have more developed clinical judgment but may know less about the latest advances in medicine. It is also not known whether these factors affect patient outcomes.

Why did the authors do this review?
To determine whether experience in practice is associated with the quality of health care delivered by doctors. The authors reviewed research that used different measures of experience and different measures of health care quality and pooled the results.

How did the authors do this review?
They searched MEDLINE, a large online collection of scientific articles, and identified research articles published between 1966 and 2004 that studied the association between experience in practice and health care quality.

What did the authors find?
59 studies that addressed the association between experience and health care quality. Most studies found that health care quality decreased as doctor experience or age increased. Doctors who were older or who had been in practice longer seemed to be less current with recent medical advances and also followed standards of care less closely. Several studies also suggested that patients of older doctors experienced worse outcomes.

What are the implications of the review?
Existing research suggests that doctors who are older or who have been in practice longer may provide lower-quality health care. However, these findings do not apply to all older doctors and must be confirmed by more specific research. This review should provoke careful study of the relationship of physician experience and the quality of care.