THE B FILES

Case studies of bias in real life epidemiologic studies

Bias File 7. Confounding by indication: a most stubborn bias?

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Department of Epidemiology, Biostatistics & Occupational Health

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca & jay.kaufman@mcgill.ca

McGill

Bias File 7. Confounding by indication: a most stubborn bias?

The story

Confounding is a well known bias in observational (non-randomized) epidemiologic research. Randomized controlled trials overcome this problem to a great extent; the process of randomization ensures that treatment allocation is a process that is not linked to a given patient's profile or prognostic characteristics. It also ensures that participants do not get to choose their treatment. Randomization helps create groups with a balanced covariate profile (i.e. similar prognostic characteristics) at baseline. For these reasons, the effects of drugs and interventions, ideally, should be determined by well designed randomized trials. However, this is not always possible or feasible. Also, long-term effects of therapies cannot easily be studied using randomized trial methodologies.

Pharmacoepidemiology often involves observational studies of the effect of drugs, where the outcomes of individuals on specific drugs are compared to individuals who are not on those drugs (e.g., taking some other type of medication). For example, once a drug is available on the market, post-marketing surveillance is done to evaluate long-term outcomes among patients on that drug. Confounding by indication is a bias frequently encountered in such observational pharmacoepi studies of drug effects. In these studies, participants or their doctors choose whether or not they will take the medications. Allocation of treatment is therefore far from random. Thus, the indication for treatment (i.e. the reason why a specific medication was given to a given individual) will be related to choice of treatment and may also be related to the risk of future health outcomes. The indication for treatment (which is often a mix of several reasons to start or withhold treatment) may be a very strong prognostic indicator. This results in confounding and the resulting imbalance in the underlying risk (prognostic) profile between treated and comparison groups can generate biased results (Signorello et al. 2002). Confounding by indication can often lead to paradoxical results. For example, a drug might look ineffective or even harmful because of poor outcomes among those taking that drug; this, however, could be merely because the drug was given to highly selected individuals who needed the drug because of their poor prognosis. In other words, the sickest patients were given the drug and it is not surprising that their outcomes were poor. On the other hand, suppose that a drug contains a warning that it must never be given to people with hypertension. A crude analysis of the association between the drug and hypertension status in the patient population will appear as though the drug is highly protective against this disease.

The study

The pharmacoepi literature is full of examples of confounding by indication. Signorello et al. (2002) in their nice overview of confounding by indication, provide many good examples. The Physicians' Health Study (N Engl J Med 1989) on aspirin for myocardial infarction is one such example. It was a randomized, double-blind, placebo controlled trial, where one component was to estimate the protective effect of aspirin on cardiovascular disease and mortality.

From the original abstract (N Engl J Med 1989):

The Physicians' Health Study [PHS] is a randomized, double-blind, placebo-controlled trial designed to determine whether low-dose aspirin (325 mg every other day) decreases cardiovascular mortality and whether beta carotene reduces the incidence of cancer. The aspirin component was terminated earlier than scheduled, and the preliminary findings were published. We now present detailed analyses of the cardiovascular component for 22,071 participants, at an average follow-up time of 60.2 months. There

was a 44 percent reduction in the risk of myocardial infarction (relative risk, 0.56; 95 percent confidence interval, 0.45 to 0.70; P less than 0.00001) in the aspirin group (254.8 per 100,000 per year as compared with 439.7 in the placebo group). A slightly increased risk of stroke among those taking aspirin was not statistically significant; this trend was observed primarily in the subgroup with hemorrhagic stroke (relative risk, 2.14; 95 percent confidence interval, 0.96 to 4.77; P = 0.06). No reduction in mortality from all cardiovascular causes was associated with aspirin (relative risk, 0.96; 95 percent confidence interval, 0.60 to 1.54). Further analyses showed that the reduction in the risk of myocardial infarction was apparent only among those who were 50 years of age and older. The benefit was present at all levels of cholesterol, but appeared greatest at low levels. The relative risk of ulcer in the aspirin group was 1.22 (169 in the aspirin group as compared with 138 in the placebo group; 95 percent confidence interval, 0.98 to 1.53; P = 0.08), and the relative risk of requiring a blood transfusion was 1.71. This trial of aspirin for the primary prevention of cardiovascular disease demonstrates a conclusive reduction in the risk of myocardial infarction, but the evidence concerning stroke and total cardiovascular deaths remains inconclusive because of the inadequate numbers of physicians with these end points.

After the trial was stopped early, all participants were then offered the opportunity to take aspirin, and the study population remained under observation. Some participants chose to take aspirin while others did not take it or stopped taking after a while.

A follow up observational study on this cohort was published in 2000 (Cook et al, 2000), where posttrial use of aspirin was assessed at the 7-year follow-up among 18 496 participants with no previous reported CVD. Randomized and posttrial observational results in the PHS were compared, and differences between those self-selecting aspirin and those not were examined.

From the original abstract (Cook et al, 2000):

BACKGROUND: The randomized aspirin component of the Physicians' Health Study (PHS) was terminated early, after 5 years, primarily because of the emergence of a statistically extreme (P<.00001) 44% reduction of first myocardial infarction (MI) among those assigned to aspirin. As a result, there were insufficient numbers of strokes or cardiovascular disease (CVD)-related deaths to evaluate these end points definitively. METHODS: Data on self-selected aspirin use were collected until the beta carotene component ended as scheduled after 12 years. Posttrial use of aspirin was assessed at the 7-year follow-up among 18 496 participants with no previous reported CVD. Randomized and posttrial observational results in the PHS were compared, and differences between those self-selecting aspirin and those not were examined. RESULTS: At 7 years, 59.5% of participants without CVD reported self-selected aspirin use for at least 180 d/y, and 20.8% for 0 to 13 d/y. Use was significantly associated with family history of MI, hypertension, elevated cholesterol levels, body mass index, alcohol consumption, exercise, and use of vitamin E supplements. In multivariate analyses, self-selected aspirin use for at least 180 vs 0 to 13 d/y was associated with lower risk for subsequent MI (relative risk [RR], 0.72; 95% confidence interval [CI], 0.55-0.95), no relation with stroke (RR, 1.02; 95% CI, 0.74-1.39), and significant reductions in CVD-related (RR, 0.65; CI, 0.47-0.89) and total mortality (RR, 0.64; CI, 0.54-0.77). CONCLUSION: These associations between self-selected aspirin use and CVD risk factors increase the likelihood of residual confounding and emphasize the need for large-scale randomized trials, such as the ongoing Women's Health Study, to detect reliably the most plausible small to moderate effects of aspirin in the primary prevention of stroke and CVD-related death.

The bias

As Signorello and colleagues point out:

"although the earlier randomization assured balance between patients assigned and not assigned aspirin during the trial, in the posttrial period, patients on aspirin therapy were not comparable to those who did not take aspirin with respect to the underlying risk of an MI event. The Physician's Health Study investigators compared the characteristics of subjects who reported taking aspirin at least 180 days out of the previous year with those who took less or none at all.13 Subjects who chose to take aspirin for 180 days or more (compared with nonusers) were: 1) slightly heavier, 2) slightly older, 3) about 30% more likely to have a family history of MI, 4) almost 20% more likely to be under treatment of hypertension, 5) almost 50% more likely to be under treatment to lower their cholesterol (and still had higher cholesterol levels), and 6) about 45% more likely to be daily alcohol drinkers. Conversely, they were shown to be more likely to exercise at least once per week and also to take vitamin E supplements. Hence, the baseline profiles between the aspirin and comparison groups differed. Assessing the first MI risk from 7 years to 12 years of follow-up—using self-selected aspirin use— relative risks of 0.83 (95% CI = 0.52–1.31), 0.90 (95% CI = 0.61–1.33), and 0.72 (95% CI = 0.55–0.95) for takers of 14 to 120 aspirin, 121 to 179 aspirin, and 180 aspirin per year, respectively, were observed. These results were adjusted for more than 15 confounding factors, including age, body mass index, smoking, exercise, personal and family history of cardiovascular problems, comorbidities, and other clinical factors. The difference between these results and those of the randomized trial may then be, at least in part, because of uncontrolled confounding by indication." (Signorello et al. 2002)

The lesson

The PHS aspirin study example clearly illustrates the difference between randomized evaluations of drug effects and observational (non-randomized) studies on drug effects. As shown in the graphic below from the *Clinical Epidemiology* text by Grobbee and Hoes (2009), the randomized design effectively breaks any potential correlation between the intervention and reasons to initiate or refrain from a specific intervention (the left arrow in the graphic cannot exist). Using the aspirin study, randomization ensures that aspirin is not selectively offered to, for example, older males who smoke, are overweight, and have family history of cardiovascular problems. Thus, confounding by these factors is unlikely to occur.

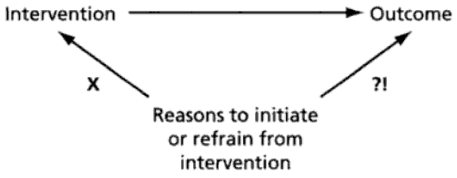

FIGURE 5.7 Major strength of a random allocation of patients to an intervention.

In an observational design, there will always be correlation between intervention and reasons to initiate that intervention (as shown in the schematic below; the left arrow now exists).

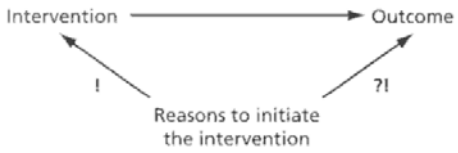

FIGURE 5.6 Reasons underlying the decision to initiate treatment are important potential confounders.

The non-randomized, observational design reflects a real life situation, where a drug will usually be tailored to a patient's profile and likely prognosis. In other words, those who need a drug most are likely to get it, and those who will be harmed by the drug are most likely to not get it. So, it is highly likely that aspirin users will be older, smokers, overweight, have already had cardiovascular events, and/or have comorbid conditions such as diabetes and hypertension. These factors will likely result in confounding by indication because they are also associated with the outcome. In fact, if this confounding is not adequately adjusted, it can produce non-intuitive results - aspirin might appear to have no effect or even appear harmful!

Is there a solution to the problem of confounding by indication? In the PHS follow up study (Cook et al 2000), the authors adjusted for a large number of confounders and still struggled with residual confounding. Conventional multivariable analyses may not entirely address the problem of confounding, especially when confounding is due to unmeasured factors or due to time-varying covariates (covariates that change with time). In a subsequent study (Cook et al, 2002), the same PHS investigators attempted to address this problem with a more sophisticated approach called marginal structural model (MSM). This model, along with related methods such as propensity scores (PS) analysis are newer approaches to handle confounding, when the treatment is not randomly assigned. But these methods still rest on the assumption of having measured all of the important confounders.

The underlying principle in these novel methods is to use the observational study data, but try to simulate the randomized trial design. For example, in the propensity score approach, the first step is to compute each person's probability of being assigned to a particular treatment, given a set of measured covariates. Then, these predicted probabilities (“propensity scores”) are used to reduce bias by generating groups with fairly similar levels of indication for treatment. In the case of aspirin, the first step would be to generate the predicted probability for each person that they be prescribed aspirin, based on measured covariates such as age, sex, weight, previous cardiovascular disease, etc. In other words, what is the profile of someone who is likely to be on aspirin therapy? Then comparison between aspirin and control treatments can be done within individuals with similar levels of the propensity scores. This should help equalize the two groups (in the same way randomization would) and reduce confounding by indication, as long as enough covariates were measured (and measured well).

While methods such as propensity scores and marginal structural models can go beyond conventional methods to adjust for confounding, they may still not be adequate to completely remove the bias. A recent article entitled "A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies" concluded that conventional methods do not control for unmeasured factors, which often remain important when addressing confounding by indication. Methods such as propensity scores can be useful under specific situations, but they may not adequately control confounding by indication in many real-world applications (Bosco et al, 2009).

Sources and suggested readings*

1. Signorello LB, McLaughlin JK, Lipworth L, Friis S, Sørensen HT, Blot WJ. Confounding by indication in epidemiologic studies of commonly used analgesics.Am J Ther. 2002 May-Jun;9(3):199-205.
2. Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. N Engl J Med. 1989 Jul 20;321(3):129-35.
3. Cook NR, Hebert PR, Manson JE, Buring JE, Hennekens CH. Self-selected posttrial aspirin use and subsequent cardiovascular disease and mortality in the physicians' health study. Arch Intern Med. 2000;160(7):921-8.
4. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. Am J Epidemiol. 2002 Jun 1;155(11):1045-53.
5. Grobbee DE, Hoes AW. Clinical epidemiology. Jones & Bartlett, 2009.
6. Bosco J et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol 2009; May 19, 2009.

*From this readings list, the most relevant papers are enclosed.

Acknowledgement

We are grateful to Professor Noel Weiss, University of Washington, Seattle, for his input and contribution to this case study.

The New England
Journal of Medicine

Volume 321 JULY 20, 1989 Number 3

FINAL REPORT ON THE ASPIRIN COMPONENT OF THE ONGOING PHYSICIANS' HEALTH STUDY

STEERING COMMITTEE OF THE PHYSICIANS' HEALTH STUDY RESEARCH GROUP*

Abstract The Physicians' Health Study is a randomized, double-blind, placebo-controlled trial designed to determine whether low-dose aspirin (325 mg every other day) decreases cardiovascular mortality and whether beta carotene reduces the incidence of cancer. The aspirin component was terminated earlier than scheduled, and the preliminary findings were published. We now present detailed analyses of the cardiovascular component for 22,071 participants, at an average follow-up time of 60.2 months.

There was a 44 percent reduction in the risk of myocardial infarction (relative risk, 0.56; 95 percent confidence interval, 0.45 to 0.70; P<0.00001) in the aspirin group (254.8 per 100,000 per year as compared with 439.7 in the placebo group). A slightly increased risk of stroke among those taking aspirin was not statistically significant; this trend was observed primarily in the subgroup with hemorrhagic stroke (relative risk, 2.14; 95 percent confidence interval, 0.96 to 4.77; P = 0.06). No reduction in mortality from all cardiovascular causes was associated with aspirin (relative risk, 0.96; 95 percent confidence interval, 0.60 to 1.54).

Further analyses showed that the reduction in the risk of myocardial infarction was apparent only among those who were 50 years of age and older. The benefit was present at all levels of cholesterol, but appeared greatest at low levels. The relative risk of ulcer in the aspirin group was 1.22 (169 in the aspirin group as compared with 138 in the placebo group; 95 percent confidence interval, 0.98 to 1.53; P = 0.08), and the relative risk of requiring a blood transfusion was 1.71.

This trial of aspirin for the primary prevention of cardiovascular disease demonstrates a conclusive reduction in the risk of myocardial infarction, but the evidence concerning stroke and total cardiovascular deaths remains inconclusive because of the inadequate numbers of physicians with these end points. (N Engl J Med 1989; 321: 129-35.)

ALTHOUGH chewing willow bark, which has aspirin-like properties, was prescribed for pain relief by Hippocrates in the fifth century B.C., the possible role of aspirin in reducing the risk of cardiovascular disease has been recognized only very recently. Such a possibility derives from the capacity of aspirin in low doses to inhibit cyclooxygenase-dependent platelet enzymes virtually completely, resulting in the inhibition of aggregability for the life of the platelet.[1] These effects are so profound that higher doses add little benefit but do increase the risk of side effects.[2] Although an early case–control study[3] raised the possibility of a large benefit, most observational studies[4,5] have suggested a cardiovascular benefit of about 20 percent. In such circumstances, the amount of uncontrolled confounding in case–control or cohort studies may be as large as the small-to-moderate effects being sought[6]; consequently, conclusive data can result only from a randomized trial whose sample is sufficiently large.[7,8]

The Physicians' Health Study is a double-blind, placebo-controlled, randomized trial designed to test two primary-prevention hypotheses in a population of healthy male physicians: whether aspirin in low doses (Bufferin, Bristol-Myers Products, 325 mg every other day) reduces mortality from cardiovascular disease, and whether beta carotene (Lurotin, BASF, 50 mg on alternate days) decreases the incidence of cancer. Although the beta carotene component of the trial is continuing at least through 1990, the Data Monitoring Board recommended the early termination of the

From the Channing Laboratory, Departments of Medicine and Preventive Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston. Address reprint requests to Dr. Charles H. Hennekens at 55 Pond Ave., Brookline, MA 02146.

Supported by grants (HL-26490, HL-34595, CA-34944, and CA-40360) from the National Institutes of Health.

*The members of the Steering Committee and of the Physicians' Health Study Research Group are as follows: *Steering Committee:* Charlene Belanger, M.A., Julie E. Buring, Sc.D., Nancy Cook, Sc.D., Kimberley Eberlein, M.P.H., Samuel Z. Goldhaber, M.D., David Gordon, M.A., Charles H. Hennekens, M.D. (chairman), Sherry L. Mayrent, Ph.D., Richard Peto, F.R.S., Bernard Rosner, Ph.D., Meir J. Stampfer, M.D., Fran Stubblefield, B.S., and Walter C. Willett, M.D. The Steering Committee also includes the members of the Data Monitoring Board, as well as Thomas Ryan, M.D., Thomas Blaszkowski, Ph.D., (ex officio), and Andrew Vargosko, Ph.D. (ex officio). *Physicians' Health Study Research Group:* Charlene Belanger, M.A., Eugene Braunwald, M.D., Julie E. Buring, Sc.D., Nancy Cook, Sc.D., Richard Doll, F.R.S., Kimberley Eberlein, M.P.H., Samuel Z. Goldhaber, M.D., David Gordon, M.A., Charles H. Hennekens, M.D. (principal investigator), Edward H. Kass, M.D., Janet Lang, Sc.D., JoAnn Manson, M.D., Micheline Mathews-Roth, M.D., Sherry L. Mayrent, Ph.D., Richard Peto, F.R.S., Bernard Rosner, Ph.D., Suzanne Satterfield, M.D., Frank E. Speizer, M.D., Meir J. Stampfer, M.D., Fran Stubblefield, B.S., James O. Taylor, M.D., and Walter C. Willett, M.D. *End Points Committee:* Harris Funkenstein, M.D., Samuel Z. Goldhaber, M.D., Meir J. Stampfer, M.D., and James O. Taylor, M.D. (chairman). *Data Monitoring Board:* Lawrence S. Cohen, M.D., Theodore Colton, Sc.D., David DeMets, Ph.D. (chairman), Daniel Deykin, M.D., Lawrence Friedman, M.D., and Bernard Rosner, Ph.D. Recently added members are John Cairns, M.D., Peter Greenwald, M.D., and Charles H. Hennekens, M.D. The retired chairman is George B. Hutchison, M.D.

blinded aspirin component of the trial on December 18, 1987. This decision was based on all available evidence, including three major considerations: the presence of a significant (P<0.00001) reduction in the risk of total myocardial infarction among those in the aspirin group; the fact that no effect of aspirin on cardiovascular mortality could be detected in the trial until the year 2000 or later, because of the exceptionally low cardiovascular death rates among the participating physicians; and the fact that aspirin was subsequently prescribed for more than 85 percent of the participants who experienced nonfatal vascular events, which made any finding about cardiovascular mortality particularly difficult to interpret. The trial's preliminary findings were published on January 28, 1988.[9] We report here the results of the final analyses of the cardiovascular component up to January 25, 1988, when the participants were told whether they had been assigned to the aspirin or the placebo group.

Methods

The subjects and methods of the Physicians' Health Study were described in detail in the preliminary report.[9] Briefly, 22,071 physicians were randomly assigned, according to two-by-two factorial design,[10] to one of four treatment groups: aspirin and beta carotene, aspirin and beta carotene placebo, aspirin placebo and beta carotene, or aspirin placebo and beta carotene placebo. Altogether, 11,037 physicians were assigned at random to receive aspirin and 11,034 to receive aspirin placebo. All 22,071 participants randomly assigned to a treatment group have been included in all analyses.

Every six months for the first year and annually thereafter, the participants were sent a supply of monthly calendar packs (provided by Bristol-Myers Products) containing white tablets (aspirin or placebo) for odd-numbered days and red capsules (beta carotene or placebo) for even-numbered days. They were also sent brief questionnaires asking about their compliance with the treatment regimen and the occurrence of any relevant events.

By January 25, 1988, the participants had been followed for an average of 60.2 months (range, 45.8 to 77.0); 99.7 percent were still providing information on morbidity, and the vital status of all 22,071 doctors was known. The reported consumption of aspirin or other platelet-active drugs was 85.71 percent in the aspirin group and 14.23 percent in the placebo group. A total of 1269 physicians (624 taking aspirin and 645 taking aspirin placebo) requested an enteric-coated preparation (supplied by Bristol-Myers Products), and an additional 29 (16 assigned to aspirin and 13 assigned to placebo) specifically requested Ecotrin or its placebo (supplied by SmithKline Beckman).

When a participant reported a relevant outcome event, written consent for the review of his medical records was obtained. The information was requested from hospitals and responsible physicians. Reported diagnoses of cardiovascular disease or deaths were considered confirmed only after the examination of all available information by an End Points Committee of physicians that included two internists, a cardiologist, and a neurologist, all blinded to the assigned treatment. When written consent or the relevant records could not be obtained, a reported event could not be confirmed. Records were available for review for 95.6 percent of the reported myocardial infarctions, 95.2 percent of the strokes, and 94.8 percent of all deaths. All our analyses were based on confirmed events.

The diagnoses of nonfatal myocardial infarction were confirmed with use of the criteria of the World Health Organization.[11] Nonfatal stroke was defined as a typical neurologic deficit that was sudden or rapid in onset, lasted more than 24 hours, and was attributable to a cerebrovascular event. Strokes were further classified according to the severity of the residual impairment at the time of hospital discharge (mild, moderate, or severe) and according to the probable cause (ischemic or hemorrhagic) on the basis of medical records and the judgment of the neurologist. Death due to a cardiovascular cause was documented by convincing evidence of a cardiovascular mechanism from all available sources, including death certificates, hospital records, and — for death outside the hospital — observers' impressions.

For the end points of myocardial infarction and stroke (Tables 1 and 2), only the first event within each category was counted. For cardiovascular mortality (Table 3), all deaths were included in the analyses. Thus, for the 15 subjects who had both a nonfatal myocardial infarction and a nonfatal stroke, both events were counted as end points. For the 23 who had a nonfatal myocardial infarction (or stroke) followed by death from a cardiovascular cause, the nonfatal event was included in our analysis of myocardial infarction (or stroke), and the fatal event was included in cardiovascular deaths. In addition, we performed analyses using as end points only the first cardiovascular event experienced by the participants — myocardial infarction, stroke, or cardiovascular death — and this method yielded virtually identical results. For the combined end point of nonfatal myocardial infarction, nonfatal stroke, and death from a cardiovascular cause, only a participant's first cardiovascular event was counted.

The relative risk was calculated as the number of events per person-year of observation in the aspirin group divided by the corresponding number in the placebo group after adjustment for age and for assignment to beta carotene treatment,[12] even though no significant effect of beta carotene was observed on any of the major cardiovascular end points. Since the trial was so large, the adjusted ratios were, in practice, never materially different from the ratios of the crude numbers affected in each group.[8] For each relative risk, the two-sided P value and the 95 percent confidence interval were calculated.[12] Multiple logistic-regression analysis[12] was used to control simultaneously for the joint effects of any small differences in base-line cardiovascular risk factors, although no individual factor differed significantly between the aspirin and placebo groups.

Results

There were 139 myocardial infarctions among those assigned to aspirin and 239 among those assigned to aspirin placebo (relative risk, 0.56; 95 percent confidence interval, 0.45 to 0.70; P<0.00001) (Table 1). This represents a 44 percent reduction in risk, and the benefits of aspirin were significant for both fatal and nonfatal myocardial infarction. In terms of absolute rates of events, the figures can be extrapolated to 254.8 per 100,000 per year in the aspirin group and 439.7 per 100,000 per year in the placebo group. As for total stroke, there were 119 events in the aspirin group and 98 in the placebo group — an increase in risk that was not statistically significant (relative risk, 1.22; 95 percent confidence interval, 0.93 to 1.60; P = 0.15). Strokes were then subdivided into ischemic and hemorrhagic events and, further, into those resulting in mild, moderate, or severe disability or death (Table 2). In the subgroup with hemorrhagic strokes, aspirin was associated with an increased risk that was of borderline statistical significance (relative risk, 2.14; 95 percent confidence interval, 0.96 to 4.77; P = 0.06). This subgroup included 10 mild hemorrhagic strokes in the aspirin group and 6 in the placebo group (relative risk, 1.67; 95 percent confidence interval, 0.61 to 4.57; P = 0.32), as well as 13 moderate-to-severe or fatal hemorrhagic strokes in the aspirin group and 6 in the placebo group (relative risk, 2.19; 95 percent confidence interval, 0.84 to 5.69; P = 0.11).

With respect to total cardiovascular mortality (Table 3), there were 81 deaths among those assigned to aspirin and 83 among those given placebo (relative risk, 0.96; 95 percent confidence interval, 0.60 to 1.54;

Table 1. Confirmed Cardiovascular End Points in the Aspirin Component of the Physicians' Health Study, According to Treatment Group.*

End Point	Aspirin Group	Placebo Group	Relative Risk	95% Confidence Interval	P Value
Myocardial infarction					
Fatal	10	26	0.34	0.15–0.75	0.007
Nonfatal	129	213	0.59	0.47–0.74	<0.00001
Total	139	239	0.56	0.45–0.70	<0.00001
Person-years of observation	54,560.0	54,355.7	—	—	—
Stroke					
Fatal	9	6	1.51	0.54–4.28	0.43
Nonfatal	110	92	1.20	0.91–1.59	0.20
Total	119	98	1.22	0.93–1.60	0.15
Person-years of observation	54,650.3	54,635.8	—	—	—

*Additional events that could not be confirmed because records were not available included 17 myocardial infarctions (10 in the aspirin group and 7 in the placebo group) and 11 strokes (3 aspirin and 8 placebo).

P = 0.87). In the five categories of death from specific cardiovascular causes, there was only one statistically significant finding: a reduction in the rate of fatal myocardial infarction (10 in the aspirin group and 28 in the placebo group; P = 0.004). For sudden death, there was an apparent increase in risk that did not achieve statistical significance (22 in the aspirin group and 12 in the placebo group; P = 0.09). Combining the category of fatal acute myocardial infarction (ICD [*International Classification of Diseases* number] 410) with the categories of all other fatal ischemic heart disease (ICD 411 to 414) yielded 34 deaths in the aspirin group and 53 in the placebo group (relative risk, 0.60; 95 percent confidence interval, 0.37 to 0.98; P = 0.04). The addition of sudden death (ICD 798) resulted in 56 deaths in the aspirin group and 65 in the placebo group (relative risk, 0.86; 95 percent confidence interval, 0.60 to 1.22; P = 0.41). There was no reduction in the risk of all deaths from noncardiovascular causes (124 in the aspirin group vs. 133 in the placebo group; relative risk, 0.93; 95 percent confidence interval, 0.72 to 1.20; P = 0.59). Thus, there were 205 deaths with a confirmed cause in the aspirin group and 216 in the placebo group (relative risk, 0.95; 95 percent confidence interval, 0.79 to 1.15; P = 0.60).

To help clarify a risk-to-benefit ratio, we considered a combined end point consisting of nonfatal myocardial infarction, nonfatal stroke, and death from a cardiovascular cause. There were 307 important vascular events among those assigned to aspirin and 370 among those assigned to placebo (relative risk, 0.82; 95 percent confidence interval, 0.70 to 0.96; P = 0.01). This represents a statistically significant, 18 percent reduction in important vascular events among those assigned to aspirin.

As expected with a sample of 22,071 participants randomly assigned to treatment groups, there were no differences in the base-line characteristics — age, cigarette smoking, incidence of diabetes mellitus, parental history of myocardial infarction, cholesterol level, systolic blood pressure, diastolic blood pressure, alcohol use, amount of vigorous exercise, and body-mass index. When the possible effects of any differences in the joint distributions of these risk factors were taken into account through logistic regression, the relative risks for each cardiovascular end point were unchanged.

We also examined the possible effects of aspirin in subgroups of physicians with various cardiovascular risk factors. As shown in Table 4, the effects of aspirin on the risk of myocardial infarction were modified by two coronary risk factors — age and blood cholesterol level. The reduction in the risk of myocardial infarction associated with the use of aspirin was apparent only in those 50 years of age or older (P = 0.02). No consistent effect of age on the relation between aspirin and either stroke or cardiovascular mortality was observed. For cholesterol, the beneficial effects of aspirin on myocardial infarction were apparent at all levels but appeared greatest at low levels (P = 0.04). For cigarette smoking, the reduction in the risk of myocardial infarction associated with the use of aspirin was similar among those who had never smoked, past smokers, and current smokers. For stroke, cigarette smoking did not modify the effect of aspirin, but for cardiovascular mortality, it appeared to do so (P = 0.05). However, neither the observed reduction in risk among nonsmokers (P = 0.18) nor the apparently increased risk among current smokers (P = 0.20) was significant. Finally, blood-pressure levels had no consistent effect on the association between aspirin and myocardial infarction, stroke, or mortality from cardiovascular causes.

During the 60.2 months of follow-up, gastrointestinal discomfort was reported by 26.1 percent of the aspirin group and 25.6 percent of the placebo group; 0.5 percent was therefore attributable to the active drug, a nonsignificant excess (P = 0.45) (Table 5). For all gastrointestinal symptoms except ulcer, the corresponding figures were 34.8 percent and 34.2 percent, respectively (P = 0.48, also not significant).

Table 2. Subcategories of Stroke, According to Treatment Group.*

Type of Stroke	Aspirin Group	Placebo Group	Relative Risk	95% Confidence Interval	P Value
Ischemic					
Mild	69	61	1.13	0.80–1.60	0.48
Moderate, severe, or fatal	21	20	1.05	0.57–1.95	0.88
Unknown severity	1	1			
Total	91	82	1.11	0.82–1.50	0.50
Hemorrhagic					
Mild	10	6	1.67	0.61–4.57	0.32
Moderate, severe, or fatal	13	6	2.19	0.84–5.69	0.11
Total	23	12	2.14	0.96–4.77	0.06
Unknown cause					
Mild	2	1	—	—	—
Moderate, severe, or fatal	1	2	—	—	—
Unknown severity	2	1	—	—	—
Total	5	4	—	—	—
Total	119	98	1.22	0.93–1.60	0.15

*Severity was defined as follows: mild, impairment not affecting functioning; moderate, functional impairment; and severe, a major change in way of life or dependence.

Table 3. Confirmed Deaths, According to Treatment Group.

Cause*	Aspirin Group	Placebo Group	Relative Risk	95% Confidence Interval	P Value
Total cardiovascular deaths†	81	83	0.96	0.60–1.54	0.87
Acute myocardial infarction (410)	10	28	0.31	0.14–0.68	0.004
Other ischemic heart disease (411–414)	24	25	0.97	0.60–1.55	0.89
Sudden death (798)	22	12	1.96	0.91–4.22	0.09
Stroke (430, 431, 434, 436)‡	10	7	1.44	0.54–3.88	0.47
Other cardiovascular (402, 421, 424, 425, 428, 429, 437, 440, 441)	15	11	1.38	0.62–3.05	0.43
Total noncardiovascular deaths	124§	133	0.93	0.72–1.20	0.59
Total deaths with confirmed cause	205	216	0.95	0.79–1.15	0.60
Total deaths¶	217	227	0.96	0.80–1.14	0.64
Person-years of observation	54,894.6	54,864.2	—	—	—

*Numbers are code numbers of the *International Classification of Diseases*, ninth revision.

†All fatal cardiovascular events are included, regardless of previous nonfatal events.

‡This category includes ischemic (3 in the aspirin group and 3 in the placebo group), hemorrhagic (7 aspirin and 2 placebo), and unknown cause (0 aspirin and 2 placebo).

§This category includes one death due to gastrointestinal hemorrhage.

¶Additional events that could not be confirmed because records were not available included 23 deaths (12 aspirin and 11 placebo), of which 11 were suspected to be cardiovascular (7 aspirin and 4 placebo) and 12 noncardiovascular (5 aspirin and 7 placebo).

There were 169 participants with ulcer in the aspirin group and 138 in the placebo group (relative risk, 1.22; 95 percent confidence interval, 0.98 to 1.53; P = 0.08). Among those with ulcer, 38 of the participants taking aspirin experienced some hemorrhage, as compared with 22 taking placebo (relative risk, 1.77; 95 percent confidence interval, 1.07 to 2.94; P = 0.04).

With respect to bleeding, 2979 of the aspirin group and 2248 of those taking placebo reported problems such as easy bruising, hematemesis, melena, nonspecific gastrointestinal bleeding, epistaxis, or other bleeding (relative risk, 1.32; 95 percent confidence interval, 1.25 to 1.40; P<0.00001). Furthermore, 48 in the aspirin group and 28 in the placebo group required transfusion (relative risk, 1.71; 95 percent confidence interval, 1.09 to 2.69; P = 0.02). One death from gastrointestinal hemorrhage was reported. This occurred in the aspirin group, and the event was confirmed.

Discussion

The results just described were virtually identical to the findings presented in our preliminary report.[9] Overall, there was a statistically significant, 44 percent reduction in the risk of myocardial infarction that included significant benefits of aspirin for both fatal and nonfatal events. There continued to be an apparent but not significantly increased risk of stroke — primarily in the subgroup of all hemorrhagic strokes — associated with the use of aspirin. In the subgroup of moderate-to-severe or fatal hemorrhagic stroke, the increased risk that we had observed previously was no longer statistically significant (13 events in the aspirin group and 6 in the placebo group; P = 0.11). No reduction in the risk of mortality from all cardiovascular causes was associated with aspirin. Our findings regarding stroke and cardiovascular mortality must be viewed in the context of the fact that the trial had too few events to evaluate either of these end points.

Even in a trial with our large sample, the ability to identify particular subgroups of participants more or less likely to benefit from aspirin is limited. We observed no significant modification of the effects of aspirin on any of the major manifestations of cardiovascular disease across the various risk-factor subgroups classified according to cigarette smoking, the history of diabetes mellitus, parental myocardial infarction, diastolic blood pressure, systolic blood pressure, alcohol use, amount of exercise, and body-mass index. The reduction in the risk of myocardial infarction associated with aspirin was apparent only among those 50 or older. More important, however, was the low absolute risk of myocardial infarction at younger ages, suggesting that any net benefit of aspirin would be greatest among physicians 50 years of age and older. For cholesterol, the beneficial effects of aspirin were apparent at all levels. The observation that the benefits of aspirin for myocardial infarction were greatest at low levels of cholesterol was unexpected. This observation may be correct or may reflect random fluctuations in the data from which it was derived. It is not possible on the basis of the present data to decide between these explanations. For stroke and cardiovascular mortality, there was no apparent modification of the effects of aspirin according to risk factors, with the possible exception of cigarette smoking (in relation only to cardiovascular mortality). The number of these end points was too small for us to detect whether overall results were meaningful; the analysis of these end points according to subgroup was therefore particularly difficult.

The only other randomized trial of the role of aspirin in the primary prevention of cardiovascular disease is a smaller study of British doctors.[13] In contrast to the Physicians' Health Study, the British Doctors' Trial showed no significant differences for fatal or nonfatal myocardial infarction, but the 95 percent confidence intervals were very wide. As in the Physicians' Health Study, there were more strokes among those assigned to aspirin, although the difference was not significant, and there was no significant difference in mortality from all cardiovascular causes. There were differences in the design of the U.S. and British trials, including the doses (325 mg on alternate days vs. 500 mg daily), blinding (double-blind with placebo control vs. single-blind), compliance (85.71 percent in the aspirin group and 85.74 percent in the placebo group after 60.2 months vs. 70 percent in the aspirin group and 98 percent in the group without aspirin after 36 months of a 72-month follow-up), and definition of end points. Perhaps most relevant, however, was the difference in the sample size; 5139 subjects were randomly assigned to aspirin or control in the British

Table 4. Risk of Total Myocardial Infarction Associated with Aspirin Use, According to Level of Coronary Risk Factors.

	Aspirin Group	Placebo Group	Relative Risk	P Value of Trend in Relative Risk
	no. of myocardial infarctions/total no. (%)			
Age (yr)				
40–49	27/4527 (0.6)	24/4524 (0.5)	1.12	
50–59	51/3725 (1.4)	87/3725 (2.3)	0.58	0.02
60–69	39/2045 (1.9)	84/2045 (4.1)	0.46	
70–84	22/740 (3.0)	44/740 (6.0)	0.49	
Cigarette smoking				
Never	55/5431 (1.0)	96/5488 (1.8)	0.58	
Past	63/4373 (1.4)	105/4301 (2.4)	0.59	0.99
Current	21/1213 (1.7)	37/1225 (3.0)	0.57	
Diabetes mellitus				
Yes	11/275 (4.0)	26/258 (10.1)	0.39	0.22
No	128/10,750 (1.2)	213/10,763 (2.0)	0.60	
Parental history of myocardial infarction				
Yes	23/1420 (1.6)	39/1432 (2.7)	0.59	0.97
No	112/9505 (1.2)	192/9481 (2.0)	0.58	
Cholesterol level (mg per 100 ml)*				
<159	2/382 (0.5)	9/406 (2.2)	0.23	
160–209	12/1587 (0.8)	37/1511 (2.5)	0.29	0.04
210–259	26/1435 (1.8)	43/1444 (3.0)	0.61	
≥260	14/582 (2.4)	23/570 (4.0)	0.59	
Diastolic blood pressure (mm Hg)				
≤69	2/583 (0.3)	9/562 (1.6)	0.21	
70–79	24/2999 (0.8)	40/3076 (1.3)	0.61	0.88
80–89	71/5061 (1.4)	128/5083 (2.5)	0.55	
≥90	26/1037 (2.5)	43/970 (4.4)	0.56	
Systolic blood pressure (mm Hg)				
<109	1/330 (0.3)	4/296 (1.4)	0.22	
110–129	40/5072 (0.8)	75/5129 (1.5)	0.52	0.48
130–149	63/3829 (1.7)	115/3861 (3.0)	0.55	
≥150	19/454 (4.2)	26/412 (6.3)	0.65	
Alcohol use				
Daily	26/2718 (1.0)	55/2727 (2.0)	0.45	
Weekly	70/5419 (1.3)	112/5313 (2.1)	0.61	0.26
Rarely	40/2802 (1.4)	65/2897 (2.2)	0.63	
Vigorous exercise at least once a week				
Yes	91/7910 (1.2)	140/7861 (1.8)	0.65	0.21
No	45/2997 (1.5)	92/3060 (3.0)	0.49	
Body-mass index†				
≤23.0126	26/2872 (0.9)	41/2807 (1.5)	0.61	
23.0127–24.4075	32/2700 (1.2)	46/2627 (1.8)	0.68	0.90
24.4076–26.3865	32/2713 (1.2)	75/2823 (2.7)	0.44	
≥26.3866	49/2750 (1.8)	76/2776 (2.7)	0.65	

*To convert cholesterol value to millimoles per liter, multiply by 0.02586.

†Body-mass index is the weight (in kilograms) times the height (in meters) squared.

trial, and 22,071 in the U.S. study. To minimize the effect of this difference in sample size, we undertook an overview of the two trials of primary prevention.[14] Since the U.S. trial was so much larger, this overview showed a significant, 33 percent reduction in nonfatal myocardial infarction associated with aspirin ($P<0.0002$). A repeated analysis, in which data from this final report were used, gave virtually identical results.

The findings of this study of aspirin in the primary prevention of cardiovascular disease should be viewed in the context of all the evidence concerning the possible role of aspirin. In 1985 the Food and Drug Administration approved the labeling of aspirin as an agent to be prescribed for the treatment of patients with a previous myocardial infarction or unstable angina.[15] A recent overview[16] of 25 randomized trials of antiplatelet therapy (aspirin, dipyridamole, or sulfinpyrazone, alone or in combination) in patients with a history of cardiovascular disease demonstrated a 25 percent reduction in the incidence of subsequent important vascular events (nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular disease), a 32 percent decrease in nonfatal myocardial infarction, a 27 percent reduction in nonfatal stroke, and a 15 percent reduction in cardiovascular mortality. All these reductions were statistically significant. Moreover, aspirin was at least as effective as dipyridamole or sulfinpyrazone. Most recently, the Second International Study of Infarct Survival[17] evaluated the role of aspirin in evolving myocardial infarction in a randomized trial with 17,187 participants and demonstrated a 49 percent reduction in nonfatal myocardial infarction, a 46 percent decrease in nonfatal stroke, and a 23 percent reduction in cardiovascular death after five weeks. All these reductions were statistically significant.

Aspirin's benefits in reducing the incidence of subsequent myocardial infarction have been shown conclusively in the survivors of myocardial infarction and stroke and in patients with unstable angina, as well as in those with an evolving heart attack. Our trial demonstrates conclusively a benefit of aspirin in reducing the incidence of first myocardial infarction and thus extends the previous findings to healthy people. Benefits in reducing the incidence of subsequent stroke have been shown conclusively in survivors of myocardial infarction and stroke and in patients with unstable angina, as well as in those with an evolving heart attack. The findings of our trial of primary prevention, although not statistically significant, are compatible with an increase in the number of all strokes among aspirin users. It seems important to distinguish between ischemic stroke, in which one might expect aspirin to be of benefit, and hemorrhagic stroke, in which one might expect an adverse effect. The possibility of an increase in the incidence of hemorrhagic stroke among aspirin users is not unexpected, since any agent that decreases clotting may help prevent ischemic events but increase bleeding. To evaluate this matter further, future trials in which the sample size is adequate to distinguish between ischemic and hemorrhagic stroke are required.

Finally, aspirin's benefits in reducing the incidence of cardiovascular mortality have been shown con-

Table 5. Side Effects According to Treatment Group.

Side Effect*	Aspirin Group	Placebo Group	P Value
	number/percent		
Gastrointestinal symptoms (except ulcer)	3843 (34.8)	3779 (34.2)	0.48
Discomfort (535)	2882 (26.1)	2823 (25.6)	0.45
Other noninfectious disorders of the digestive tract (536, 537.8, 537.9)	345 (3.1)	288 (2.6)	0.02
Miscellaneous symptoms of the digestive tract (533.123, 787, 789.0)	2384 (21.6)	2405 (21.8)	0.75
Upper gastrointestinal ulcers	169 (1.5)	138 (1.3)	0.08
Esophageal ulcer (530.2)	11 (0.1)	6 (0.05)	0.23
Gastric ulcer (531)	25 (0.2)	15 (0.1)	0.11
Duodenal ulcer (532)	46 (0.4)	27 (0.2)	0.03
Peptic ulcer (533)	156 (1.4)	129 (1.2)	0.11
Gastrojejunal (534)	3 (0.03)	4 (0.04)	0.70
Bleeding problems	2979 (27.0)	2248 (20.4)	<0.0001
Easy bruising (459)	1587 (14.4)	1027 (9.3)	<0.0001
Hematemesis (578.0)	38 (0.3)	28 (0.3)	0.22
Melena (578.1)	364 (3.3)	246 (2.2)	<0.00001
Nonspecific gastrointestinal bleeding (578.9)	440 (4.0)	422 (3.8)	0.55
Epistaxis (784.7)	862 (7.8)	640 (5.8)	<0.0001
Other bleeding† (599.7, 958.2)	724 (6.6)	596 (5.4)	0.0004

*Numbers in parentheses are code numbers of the *International Classification of Diseases*, ninth revision.

†Twenty-nine percent were related to shaving or brushing the teeth (32 percent aspirin and 27 percent placebo), and 72 percent were hematuria (70 percent aspirin and 75 percent placebo).

clusively in the survivors of myocardial infarction and stroke and in patients with unstable angina, as well as in those with an evolving heart attack. In our trial of primary prevention, however, no reduction in the risk of mortality from all cardiovascular causes was associated with aspirin. Although it may be tempting to speculate about possible explanations for this finding, the primary consideration must be that the cardiovascular mortality rate among the physicians in this trial was only 15 percent of that expected for a general population of white men with the same age distribution over a similar period. This reduction was even greater than our conservative estimates of the "healthy volunteer"[6] effect had suggested. The chief consequence of this extremely low cardiovascular mortality rate was to render it impossible to detect reliably whether aspirin had any effect on total mortality from cardiovascular causes until at least the year 2000. In addition, over 85 percent of the participants who had a nonfatal myocardial infarction were being treated with aspirin, making any finding about cardiovascular mortality particularly difficult to interpret.

The side effects of aspirin are clearly related to the size of the dose. The Aspirin Myocardial Infarction Study[18] tested a dose of 1000 mg a day, and gastrointestinal symptoms were reported by 23.7 percent of those assigned to active treatment and 14.9 percent of those assigned to placebo. In the recently completed United Kingdom Transient Ischemic Attack trial,[2] those receiving 300 mg daily experienced 30 percent fewer side effects than those receiving 1200 mg — a significant reduction ($P<0.0001$). In the Physicians' Health Study, in which the dose was 325 mg on alternate days, the rate of gastrointestinal discomfort attributable to aspirin was relatively low (0.5 percent). As expected, increased risks of upper gastrointestinal ulcers and bleeding problems were attributable to aspirin in this trial. Of greater clinical relevance, however, were their relatively low rates of occurrence, due in part to the low dose and the frequency of administration, and also to our eligibility criteria and the prerandomization run-in phase of the study, which excluded participants unable to tolerate the drug.[9] For all these reasons, the frequency and severity of gastrointestinal discomfort, ulcers, and bleeding complications attributable to aspirin were far lower than those reported in previous trials.

There is some suggestion that low doses of aspirin are at least as effective in reducing the risk of cardiovascular disease as high doses. An overview of the trials of secondary prevention indicated that patients receiving 300 mg a day gained the same beneficial effect as those receiving 1500 mg.[16] The benefits of aspirin observed in both the Second International Study of Infarct Survival[17] and the Physicians' Health Study resulted from low-dose therapy — 160 mg daily and 325 mg every other day, respectively. A dose of 325 mg of buffered or enteric-coated aspirin every other day completely inhibits platelet aggregation, with no recovery even 48 hours after the last dose.[19] Pharmacologic studies have suggested that the dose of aspirin necessary to inhibit platelet aggregation may be lower than 80 mg.[20] However, in the acute stage of a cardiovascular event (unstable angina, transient ischemic attack, or myocardial infarction), a dose of about 160 mg to give a more rapid antithrombotic effect might be prudent.[17]

Since both trials of aspirin in primary prevention were confined to men, there is no direct evidence on the role of aspirin in the primary prevention of cardiovascular disease in women. Both the overview[15] and the Second International Study of Infarct Survival[17] demonstrated aspirin's significant benefits for secondary prevention of the various manifestations of cardiovascular disease in women. One can consider generalizing the results of our primary-prevention trial to women, but any such clinical judgment must take into account their absolute risks of various cardiovascular events as well as the possibility that aspirin in low doses may have a different pharmacologic effect in women.[21] The only way to evaluate this question directly is through a randomized trial of healthy women that has a sufficiently large sample to detect the most plausible small-to-moderate effects.

Since the publication of our preliminary report, 74 percent of the participating physicians who were assigned to placebo have requested that their calendar packs be changed to aspirin. For the general public, however, the decision whether to take aspirin to reduce the risks of cardiovascular disease should be an individual judgment made only with the advice of a physician or other health care provider. In the making

of this decision, the cardiovascular risk profile of the patient and the risks and benefits of the drug should be weighed.

The dedicated and conscientious collaboration of our study's 22,071 physicians has provided important information and has made it possible to conduct the trial at a small fraction of the usual cost. With the continued commitment of the participating physicians, we shall collect more observational data on stroke and cardiovascular mortality. In addition, the ongoing randomized component of the trial should provide important information on the possible role of beta carotene in the prevention of cancer.

We are indebted to the staff of the Physicians' Health Study, particularly Mary Breen, Joanne Crowley, Medora Hearn, Michael Jonas, Jean MacFayden, Patrick McCormick, Geneva McNair, Ayn Stamper, Jemma Pierre, Leslie Power, Miriam Schvartz, M.D., and Phyllis Johnson Wojciechowski; and to Dominick Mellace of Bristol-Myers Products for his logistic support.

References

1. Moncada S, Vane JR. Arachidonic acid metabolites and the interactions between platelets and blood-vessel walls. N Engl J Med 1979; 300:1142-7.
2. UK-TIA Study Group. United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: interim results. Br Med J 1988; 296:316-20.
3. Jick H, Miettinen OS. Regular aspirin use and myocardial infarction. Br Med J 1976; 1:1057.
4. Hennekens CH, Karlson LK, Rosner B. A case-control study of regular aspirin use and coronary deaths. Circulation 1978; 58:35-8.
5. Hammond EC, Garfinkel L. Aspirin and coronary heart disease: findings of a prospective study. Br Med J 1975; 2:269-71.
6. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown, 1987.
7. Friedewald WT. Aspirin and coronary deaths. Circulation 1978; 58:39-40.
8. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976; 34:585-612.
9. The Steering Committee of the Physicians' Health Study Research Group. Preliminary report: findings from the aspirin component of the ongoing Physicians' Health Study. N Engl J Med 1988; 318:262-4.
10. Stampfer MJ, Buring JE, Willett W, Rosner B, Eberlein K, Hennekens CH. The 2×2 factorial design: its application to a randomized trial of aspirin and carotene in U.S. physicians. Stat Med 1985; 4:111-6.
11. World Health Organization. Ischaemic heart disease registers: report of the Fifth Working Group, including a second revision of the operating protocol: Copenhagen, 26–29 April 1971. Copenhagen, Denmark: Regional Office for Europe, World Health Organization, 1971.
12. Kleinbaum D, Kupper L, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, Calif.: Lifetime Learning Publications, 1982:336.
13. Peto R, Gray R, Collins R, et al. Randomised trial of prophylactic daily aspirin in British male doctors. Br Med J 1988; 296:13-6.
14. Hennekens CH, Peto R, Hutchison GB, Doll R. An overview of the British and American aspirin studies. N Engl J Med 1988; 318:923-4.
15. Aspirin for heart patients. FDA Drug Bull 1985; 15:34-6.
16. Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged antiplatelet treatment. Br Med J 1988; 296:320-31.
17. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988; 2:349-60.
18. Aspirin Myocardial Infarction Study Research Group. A randomized, controlled trial of aspirin in persons recovered from myocardial infarction. JAMA 1980; 243:661-9.
19. Stampfer MJ, Jakubowski JA, Deykin D, Schafer AI, Willett WC, Hennekens CH. Effect of alternate-day regular and enteric-coated aspirin on platelet aggregation, bleeding time, and thromboxane A_2 levels in bleeding-time blood. Am J Med 1986; 81:400-4.
20. Patrono C, Ciabattoni G, Patrignani P, et al. Clinical pharmacology of platelet cyclooxygenase inhibition. Circulation 1985; 72:1177-84.
21. Aspirin for TIA's. FDA Drug Bull 1980; 10:2.

Self-Selected Posttrial Aspirin Use and Subsequent Cardiovascular Disease and Mortality in the Physicians' Health Study

Nancy R. Cook, ScD; Patricia R. Hebert, PhD; JoAnn E. Manson, MD; Julie E. Buring, ScD; Charles H. Hennekens, MD

Background: The randomized aspirin component of the Physicians' Health Study (PHS) was terminated early, after 5 years, primarily because of the emergence of a statistically extreme ($P<.00001$) 44% reduction of first myocardial infarction (MI) among those assigned to aspirin. As a result, there were insufficient numbers of strokes or cardiovascular disease (CVD)–related deaths to evaluate these end points definitively.

Methods: Data on self-selected aspirin use were collected until the beta carotene component ended as scheduled after 12 years. Posttrial use of aspirin was assessed at the 7-year follow-up among 18 496 participants with no previous reported CVD. Randomized and posttrial observational results in the PHS were compared, and differences between those self-selecting aspirin and those not were examined.

Results: At 7 years, 59.5% of participants without CVD reported self-selected aspirin use for at least 180 d/y, and 20.8% for 0 to 13 d/y. Use was significantly associated with family history of MI, hypertension, elevated cholesterol levels, body mass index, alcohol consumption, exercise, and use of vitamin E supplements. In multivariate analyses, self-selected aspirin use for at least 180 vs 0 to 13 d/y was associated with lower risk for subsequent MI (relative risk [RR], 0.72; 95% confidence interval [CI], 0.55-0.95), no relation with stroke (RR, 1.02; 95% CI, 0.74-1.39), and significant reductions in CVD-related (RR, 0.65; CI, 0.47-0.89) and total mortality (RR, 0.64; CI, 0.54-0.77).

Conclusion: These associations between self-selected aspirin use and CVD risk factors increase the likelihood of residual confounding and emphasize the need for large-scale randomized trials, such as the ongoing Women's Health Study, to detect reliably the most plausible small to moderate effects of aspirin in the primary prevention of stroke and CVD-related death.

Arch Intern Med. 2000;160:921-928

From the Division of Preventive Medicine, Department of Medicine (Drs Cook, Manson, and Buring), and the Department of Ambulatory Care and Prevention (Dr Buring), Brigham and Women's Hospital, Harvard Medical School, and the Department of Epidemiology, Harvard School of Public Health (Drs Manson and Buring), Boston, Mass; and the Departments of Preventive Medicine and Medicine, Vanderbilt University School of Medicine, Nashville, Tenn (Dr Hebert). Dr Hennekens is now Visiting Professor of Medicine, and Epidemiology and Public Health at the University of Miami School of Medicine, Miami, Fla.

LOW-DOSE ASPIRIN has demonstrated clear net benefits in randomized trials of secondary prevention of cardiovascular disease (CVD), including a wide range of previous occlusive diseases as well as acute evolving myocardial infarction (MI).[1] Evidence from primary prevention trials supports a clear reduction in first MI among men, but the balance of benefits vs risks for stroke and CVD-related mortality has not yet been evaluated definitively. Data from observational studies on these questions of small to moderate effects have the benefits of larger numbers of end points and longer duration, but also have the inherent limitation that their results may result at least in part from confounding by unmeasured, unmeasurable, or unknown risk factors. The Physicians' Health Study (PHS),[2] a randomized trial of low-dose aspirin in the primary prevention of CVD, offered a unique opportunity to compare the randomized and posttrial observational results and to examine whether and the ways in which those who self-selected to use aspirin after the end of the randomized aspirin component differed from those who did not.

The early impact on first MI found in the PHS would be expected to translate into a reduced risk for CVD-related mortality,[3] but this has yet to be demonstrated in primary prevention trials, including the British Doctors' Trial[4] and the Thrombosis Prevention Trial,[5] primarily because of inadequate numbers of deaths. The impact of aspirin on stroke also remains unclear. Overviews of secondary prevention trials indicate that aspirin produces clear and consistent reductions in stroke,[6] but in primary prevention there have been insufficient numbers of strokes to evaluate this end point definitively. In the PHS, a possible but nonsignificant ob-

SUBJECTS AND METHODS

STUDY POPULATION

The PHS was a randomized, double-blind, placebo-controlled 2 × 2 factorial trial of aspirin and beta carotene in 22 071 US male physicians. The methods and results have been described in detail previously.[2,13] Beginning in 1982, participants, aged 40 to 84 years, with no history of MI, stroke, transient cerebral ischemia, or cancer (excluding nonmelanoma skin cancer), were randomized to one of the 4 treatment arms. Participants received monthly calendar packs containing 325 mg of aspirin (Bufferin; provided by Bristol-Myers, Princeton, NJ) with placebo on alternate days, 50 mg of beta carotene (Lurotin; provided by BASF Corporation, Ludwigshafen, Germany) with placebo on alternate days, both active drugs, or both placebos. Eligible participants had no contraindications to aspirin use and were not regularly taking aspirin, other platelet-active medications, or supplements of beta carotene (vitamin A). Written informed consent was obtained from all study participants, and the research protocol was reviewed and approved by the institutional review board at the Brigham and Women's Hospital, Boston, Mass.

On January 25, 1988, the aspirin component of the trial was terminated early, after an average of 60.2 months of follow-up.[2] The reported consumption of aspirin or other platelet-active drugs was 85.7% in the aspirin group and 14.2% in the placebo group at this time. After this date, the beta carotene component of the trial continued uninterrupted, but all physicians were asked whether they preferred active aspirin or placebo to be included in their calendar packs.

The beta carotene component of the trial continued until its scheduled end on December 31, 1995.[13] At this time, follow-up information had accrued for an average of 12 years. At the end of 11 years of follow-up (the last year completed by all participants), 99.7% of participants were providing morbidity information, and mortality information was complete for all but 1 of the 22 071 participants. At that time, 78.7% of participants were still taking beta carotene or its placebo.

END POINT DEFINITIONS

Throughout the entire follow-up period, written informed consent was requested to review the participant's medical records when a relevant end point was reported. When necessary, details were requested from hospitals and treating physicians. Reports of CVD or cause of death were considered confirmed or refuted only after the examination of all available information by an end-points committee consisting of 2 internists, a cardiologist, and a neurologist, all of whom were unaware of treatment assignments and aspirin exposure. When written consent or relevant records could not be obtained, a reported event was not classified as confirmed but remained unrefuted.

Diagnoses of nonfatal MI were confirmed using World Health Organization criteria.[14] Nonfatal stroke was defined as a focal neurologic defect, sudden or rapid in onset, that lasted more than 24 hours and was attributed to a cerebrovascular event. Death due to a cardiovascular cause was confirmed by convincing evidence from available sources, including death certificates, hospital records, and (for death outside the hospital) observers' accounts.

served increase in strokes resulted primarily from an apparent excess of a small number of hemorrhagic strokes in the aspirin group (23 vs 12; 95% confidence interval [CI], 0.96-4.77).[2]

Observational studies of these questions have, not surprisingly, been inconsistent.[7-12] In all observational studies, confounding by indication, such as a possible increased (or decreased) self-selection of aspirin use among those at higher (or lesser) risk for CVD, is difficult, if not impossible, to quantify. Our report evaluates randomized and observational posttrial data from the PHS and examines the relation of self-selection of aspirin use to risk factors for CVD as well as to subsequent morbidity and mortality after the early termination of the aspirin component of the PHS.

RESULTS

RANDOMIZED RESULTS

During the 5-year period of the randomized aspirin component of the PHS, there were 378 MIs, 217 strokes (173 ischemic and 35 hemorrhagic), 164 CVD-related deaths, and 444 total deaths (**Table 1**), as reported previously.[2] At this time, there was a highly significant ($P<.00001$) 44% reduction in first MI among those assigned to active aspirin. There was also a possible but nonsignificant 22% increase in total stroke, which was largely confined to a possible but nonsignificant increase in hemorrhagic stroke. The death rate due to CVD was much lower than originally anticipated,[3] so the 95% confidence interval for this end point was wide. Using a combined end point of MI, stroke, or CVD-related death, there was a significant 18% reduction due to aspirin use (relative risk [RR], 0.82; 95% CI, 0.70-0.96).[2] The total number of deaths was similar in both groups.

POSTTRIAL ASPIRIN USE

After the randomized trial period, participants were asked whether they wanted the white pill in their calendar packs to contain active aspirin or placebo. At the 7-year follow-up, nearly 86.6% requested active aspirin (98.6% of those in the randomized aspirin group, and 74.6% in the placebo group). Actual use as assessed by study questionnaire, however, was lower. Of the 18 496 participants with no previous reported CVD, 11 010 (59.5%) reported taking aspirin at least 180 days during the past year; 2136 (11.6%), 121 to 179 days; 1501 (8.1%), 14 to 120 days; and 3849 (20.8%), 0 to 13 days. Use was higher among those who had been randomized to active aspirin, with 65.5% reporting use of at least 180 days, compared with 53.5% in the placebo group. These proportions remained relatively stable through the remainder of follow-up. This com-

DEFINITION OF ASPIRIN USE

By the 7-year follow-up questionnaire, all participants had entered the posttrial period (past January 25, 1988), and self-selected aspirin use at this time was the primary exposure in these analyses. On this and subsequent questionnaires, participants were asked, during the past 12 months, how many days they had taken the white pills from their calendar packs, with possible response categories of 0, 1 to 13, 14 to 30, 31 to 60, 61 to 90, 91 to 120, 121 to 180, and more than 180 days. Participants were also asked on how many days they had taken additional aspirin or medication containing aspirin, using the same response categories. Total aspirin use was estimated from the white pill count and reported outside use. Responses were collapsed into the following 4 categories for analysis: 0 to 13, 14 to 120, 121 to 179, and at least 180 days of self-selected aspirin use in the past year.

STATISTICAL METHODS

Analyses of randomized aspirin use have been published previously and are based on person-years from time of randomization to the time of first CVD event or termination of the randomized aspirin component.[2] Analyses of self-selected aspirin use at the 7-year follow-up considered events occurring during the observational posttrial period from 7 years through database closure on October 24, 1995, an approximate 5-year follow-up period. All analyses of the posttrial period excluded participants with any previous report of CVD, including MI or other ischemic heart disease, stroke, transient ischemic attack or other cerebrovascular disease; atrial fibrillation; coronary artery bypass graft; percutaneous transluminal coronary angioplasty; carotid artery surgery; or angina. A total of 18 496 participants remained available for analysis. Analyses of cancer or total mortality also excluded those with a diagnosis of cancer before the 7-year follow-up (n = 595).

Known cardiovascular risk factors were considered as correlates of self-selected aspirin use. These included age, smoking, body mass index (BMI; calculated as weight in kilograms divided by the square of height in meters), hypertension, cholesterol level and treatment, diabetes, previous pulmonary embolism or deep vein thrombosis, intermittent claudication, migraine, family history of MI, use of warfarin sodium (Coumadin) or heparin, exercise, use of alcohol, and use of supplements of vitamin E. All of these variables were assessed at the 7-year follow-up, except for BMI and family history of MI (assessed at baseline), smoking (assessed at 5 years), and exercise (assessed at 3 years). Crude means and proportions of these variables were computed for the 4 categories of self-assessed aspirin use. Multivariate odds ratios were assessed using logistic regression with the 4-category aspirin variable using the CATMOD procedure of SAS software (SAS Institute, Cary, NC), comparing each aspirin group to the 0- to 13-d/y category. To test for trend across these aspirin categories, cumulative logistic regression with the LOGISTIC procedure of SAS software was used.

The association between self-selected aspirin use and subsequent CVD or mortality in the posttrial period was assessed using Cox regression models. Trend in risk across the 4 categories was tested with an ordinal variable. All analyses presented also control for the cardiovascular risk factors described above, as well as for randomized aspirin assignment. Interactions of self-selected use with randomized assignment and age were also examined.

pares with use in the randomized period of 65.3% taking aspirin at least 180 days at the 5-year follow-up in the active aspirin group (86.0% taking it ≥90 days) and 6.8% in the placebo group taking aspirin at least 180 days.

We compared characteristics according to category of self-selected posttrial aspirin use at the 7-year follow-up, excluding those in whom CVD had developed before 7 years (**Table 2**). All reported odds ratios and *P* values control for all other characteristics considered. Several statistically significant differences between groups emerged. First, posttrial aspirin use differed by previous randomized assignment. Of those who chose to take aspirin at least 180 d/y, 55.1% had been randomized to active aspirin, compared with 37.8% among those with no or little use of aspirin. There were slight differences in age and BMI at baseline, with frequent users at 7 years being slightly older and heavier. Larger differences were apparent in the proportions with family history of MI, with frequent users of aspirin more likely to have such a history. There was significantly less use of warfarin or heparin, although 8 frequent users of aspirin also reported use of these anticoagulants. There were no significant differences in previous reports of diabetes or migraine, although those with migraine reported slightly less subsequent use of aspirin. There was also no difference in previous report of headache (data not shown).

Table 1. Results From the 5-Year Randomized Aspirin Component of the Physicians' Health Study Among 22 071 US Male Physicians*

	Treatment, No. of Events			
	Active Aspirin	Placebo	RR (95% CI)†	*P*
Myocardial infarction	139	239	0.56 (0.45-0.70)	<.00001
Stroke				
Total	119	98	1.22 (0.93-1.60)	.15
Ischemic	91	82	1.11 (0.82-1.50)	.50
Hemorrhagic	23	12	2.14 (0.96-4.77)	.06
Cardiovascular mortality	81	83	0.96 (0.60-1.54)	.87
Total mortality	217	227	0.96 (0.80-1.14)	.64

**From Steering Committee of the Physicians' Health Study Research Group.*[2]

†Comparison of randomized aspirin vs aspirin placebo.

There were significant differences in treatment of hypertension and high cholesterol level, as well as differences in cholesterol levels, among the self-selected aspirin groups. Treatment for hypertension was more prevalent among those who chose to use aspirin at least 180 d/y. After controlling for this variable, blood pressure level itself did not predict posttrial use (data not shown).

Table 2. Correlates of Self-selected Aspirin Use at 7-Year Follow-up Among Subjects With No Major Previous Cardiovascular Disease*

	Aspirin Use, d/y†				
	0-13 (n = 3849)	14-120 (n = 1501)	121-179 (n = 2136)	≥180 (n = 11 010)	Trend
Randomized aspirin, %	37.8	39.0	53.8	55.1	. . .‡
OR	1.00	1.07	1.95	2.04	1.74
P	. . .‡	.31	<.001	<.001	<.001
Randomized beta carotene, %	50.0	50.0	49.8	50.1	. . .
OR	1.00	1.00	0.98	1.00	1.00
P	. . .	.98	.76	.97	.94
Mean age, y	58.9	58.1	58.3	59.1	. . .
OR (per 10 y)	1.00	0.95	0.93	1.04	1.01
P	. . .	.16	.03	.06	.001
Mean baseline BMI, kg/m^2	24.8	24.9	24.8	24.9	. . .
OR	1.00	1.00	1.01	1.02	1.01
P	. . .	.80	.22	.01	.006
Family history of MI, %	10.9	11.7	13.0	13.7	. . .
OR	1.00	1.07	1.18	1.27	1.20
P	. . .	.49	.04	<.001	<.001
PE-DVT, %	1.7	1.4	1.1	1.1	. . .
OR	1.00	1.02	0.89	0.80	0.82
P	. . .	.95	.64	.20	.14
Intermittent claudication, %	1.3	1.1	0.8	1.1	. . .
OR	1.00	0.99	0.75	0.84	0.89
P	. . .	.98	.32	.32	.39
Use of warfarin sodium or heparin, %	0.8	0.3	0.1	0.1	. . .
OR	1.00	0.42	0.13	0.10	0.14
P	. . .	.08	.006	<.001	<.001
Diabetes, %	3.6	3.2	2.7	3.6	. . .
OR	1.00	0.99	0.81	0.97	1.00
P	. . .	.96	.19	.79	.98
Migraine, %	9.9	10.9	8.9	8.8	. . .
OR	1.00	1.12	0.90	0.91	0.92
P	. . .	.26	.27	.17	.08
Treatment for hypertension, %	13.7	10.9	13.3	16.3	. . .
OR	1.00	0.79	0.98	1.17	1.20
P	. . .	.01	.80	.006	<.001
Cholesterol level					
treatment, %	3.4	2.5	4.8	5.5	. . .
OR	1.00	0.68	1.35	1.49	1.45
P	. . .	.05	.03	<.001	<.001
Mean level, untreated, mmol/L (mg/dL)	5.2 (200.9)	5.2 (201.5)	5.3 (203.3)	5.3 (204.1)	. . .
OR (per 10 mg/dL)	1.00	1.00	1.02	1.02	1.02
P	. . .	.68	.06	<.001	<.001
Smoking					
Current, %	6.8	8.1	6.4	7.0	. . .
OR	1.00	1.40	1.02	1.10	1.02
P	. . .	.01	.87	.25	.69
Past, %	41.3	44.7	44.1	43.7	. . .
OR	1.00	1.18	1.10	1.06	1.01
P	. . .	.01	.09	.17	.83
Alcohol use					
Daily, %	16.3	16.1	16.8	18.5	. . .
OR	1.00	1.28	1.38	1.44	1.31
P	. . .	.02	<.001	<.001	<.001
Weekly, %	43.8	50.8	49.1	48.9	. . .
OR	1.00	1.49	1.46	1.45	1.27
P	. . .	<.001	<.001	<.001	<.001
Monthly, %	13.9	13.7	15.1	12.9	. . .
OR	1.00	1.28	1.42	1.22	1.12
P	. . .	.02	<.001	.002	.03
Exercise (≥1 time per week), %	53.3	58.1	56.7	57.8	. . .
OR	1.00	1.18	1.11	1.19	1.13
P	. . .	.01	.07	<.001	<.001
Current vitamin E use, %	4.3	5.2	3.1	5.2	. . .
OR	1.00	1.26	0.73	1.21	1.21
P	. . .	.11	.04	.04	.007

**OR indicates odds ratio; BMI, body mass index; MI, myocardial infarction; PE-DVT, pulmonary embolism and/or deep vein thrombosis; TIA, transient ischemic attack; CABG, coronary artery bypass graft; and PTCA, percutaneous transluminal coronary angioplasty. Crude means and proportions are shown for each aspirin category. Odds ratios and* P *values are from logistic model predicting aspirin use and compare each category to the 1-13–d/y group, adjusting for all other variables listed. Multivariate trend tests are from a cumulative logistic model.*

†Excludes those with reported MI, stroke, TIA, atrial fibrillation, CABG, PTCA, carotid artery surgery, or angina before the 7-year follow-up.

‡Reference category.

Table 3. Posttrial Self-selected Asprin Use at 7 Years and Subsequent CVD and Mortality in the Physicians' Health Study*

	Aspirin Use, d/y†				
	0-13	14-120	121-179	≥180	*P* for Trend
Myocardial infarction					
No. of events	76	24	39	172	
RR (95% CI)	1.00	0.83 (0.52-1.31)	0.90 (0.61-1.33)	0.72 (0.55-0.95)	.02
Stroke					
Total					
No. of events	55	21	30	160	
RR (95% CI)	1.00	1.05 (0.63-1.74)	1.11 (0.71-1.74)	1.02 (0.74-1.39)	.96
Ischemic					
No. of events	42	13	20	110	
RR (95% CI)	1.00	0.85 (0.46-1.59)	0.98 (0.57-1.68)	0.92 (0.64-1.33)	.73
Hemorrhagic					
No. of events	6	2	1	25	
RR (95% CI)	1.00	0.89 (0.18-4.42)	0.28 (0.03-2.37)	1.29 (0.52-3.20)	.44
CVD-related death					
No. of events	60	10	20	115	
RR (95% CI)	1.00	0.45 (0.23-0.89)	0.63 (0.38-1.05)	0.65 (0.47-0.89)	.03
Death due to acute MI					
No. of events	10	4	3	27	
RR (95% CI)	1.00	1.02 (0.32-3.26)	0.53 (0.14-1.93)	0.86 (0.41-1.79)	.67
Cerebrovascular death					
No. of events	13	2	2	20	
RR (95% CI)	1.00	0.44 (0.10-1.99)	0.29 (0.06-1.30)	0.51 (0.25-1.05)	.10
Cancer death†					
No. of events	62	32	24	149	
RR (95% CI)	1.00	1.33 (0.87-2.04)	0.72 (0.45-1.15)	0.78 (0.58-1.05)	.03
Non-CVD and noncancer death					
No. of events	67	24	21	93	
RR (95% CI)	1.00	1.02 (0.63-1.62)	0.68 (0.42-1.12)	0.53 (0.39-0.74)	<.001
Total death‡					
No. of events	182	62	60	348	
RR (95% CI)	1.00	0.91 (0.68-1.22)	0.64 (0.47-0.85)	0.64 (0.54-0.77)	≤.001

**Analyses are adjusted for all cardiovascular risk factors listed in Table 2. End points are accrued from the 7-year follow-up through October 1995. RR indicates relative risk; CI, confidence interval; CVD, cardiovascular disease; and MI, myocardial infarction.*

†Excludes those with reported MI, stroke, transient ischemic attack, atrial fibrillation, coronary artery bypass graft, percutaneous transluminal coronary angioplasty, carotid artery surgery, or angina before the 7-year follow-up.

‡Excluding those with previous cancer before year 7. May not equal the sum of CVD-related, cancer, and non-CVD, noncancer deaths since some patients with cancer had CVD or non-CVD, noncancer death.

The relationship of aspirin use with cholesterol was more complex. Treatment for elevated cholesterol levels was more prevalent among those choosing aspirin. However, besides treatment, level of cholesterol also predicted aspirin use, and there was a significant interaction of cholesterol level and treatment on the choice of aspirin. Among those not receiving medication for lowering of cholesterol levels, there was a positive relation of cholesterol level with aspirin use. Among those using such medication, those with lower cholesterol levels were more likely to choose aspirin. Thus, those with cholesterol levels under greater control were more likely to use aspirin.

Although there were no significant differences in smoking status among the groups, there were strong differences in alcohol use and exercise, even after adjusting for the other cardiovascular risk factors. Frequent users of aspirin tended to report more frequent use of alcohol (at least up to daily use) and exercise, particularly when compared with those who did not use aspirin. Frequent aspirin users were also more likely to take supplements of vitamin E.

POSTTRIAL CVD

We assessed the relationship of self-selected posttrial aspirin use at 7 years with subsequent CVD and mortality in the period from 7 to 12 years of follow-up among those with no major CVD before this time (**Table 3**). All analyses were adjusted for all of the risk factors considered in Table 2 as well as randomized aspirin assignment. During this 5-year posttrial follow-up period, there were 311 unrefuted reports of MI, 266 strokes (including 185 ischemic and 34 hemorrhagic), 205 CVD-related deaths, and 782 total deaths, 652 of these among persons with no previous cancer. For MI, there was a statistically significant 28% lower rate of events in the frequent users (≥180 d/y) compared with the nonusers (0-13 d/y) (RR, 0.72; 95% CI, 0.55-0.95). The test for trend across all 4 categories was also significant ($P = .02$). There were no apparent effects of self-selected aspirin use on stroke, including ischemic or hemorrhagic stroke.

For CVD-related mortality, we observed a statistically significant 35% lower rate among frequent users of aspirin compared with nonusers (RR, 0.65; 95% CI,

Table 4. Self-selected Aspirin Use at 7 Years and Subsequent CVD and Mortality in the Physicians' Health Study by Randomized Aspirin Assignment*

	Effect of Aspirin Use, ≥180 vs 0-13 d/y†		Interaction, *P*
	Active Aspirin Group‡	Placebo Group	
Myocardial infarction			
RR (95% CI)	0.84 (0.55-1.27)	0.62 (0.42-0.90)	.28
P	.41	.01	
Stroke			
Total			
RR (95% CI)	1.13 (0.68-1.89)	0.93 (0.62-1.38)	.54
P	.63	.70	
Ischemic			
RR (95% CI)	1.26 (0.66-2.41)	0.77 (0.49-1.21)	.21
P	.48	.25	
Hemorrhagic			
RR (95% CI)	1.85 (0.42-8.03)	0.89 (0.27-2.99)	.45
P	.41	.86	
CVD-related death			
RR (95% CI)	0.90 (0.55-1.49)	0.49 (0.32-0.76)	.07
P	.69	.002	
Death due to acute MI			
RR (95% CI)	1.94 (0.45-8.43)	0.50 (0.19-1.32)	.13
P	.38	.16	
Cerebrovascular death			
RR (95% CI)	1.34 (0.39-4.62)	0.20 (0.06-0.66)	.03
P	.65	.008	
Cancer death‡			
RR (95% CI)	0.81 (0.50-1.31)	0.78 (0.53-1.15)	.90
P	.39	.21	
Non-CVD and noncancer death			
RR (95% CI)	0.57 (0.35-0.95)	0.53 (0.35-0.81)	.81
P	.03	.003	
Total death§			
RR (95% CI)	0.75 (0.56-1.00)	0.58 (0.46-74)	.18
P	.05	<.001	

*Analyses are adjusted for all cardiovascular risk factors listed in Table 2. Abbreviations are explained in the first footnote to Table 3.
†Excludes those with reported MI, stroke, transient ischemic attack, atrial fibrillation, coronary artery bypass graft, percutaneous transluminal coronary angioplasty, carotid artery surgery, or angina before the 7-year follow-up.
‡Randomized treatment assignment during the trial period.
§Excluding those with previous cancer.

0.47-0.89). Rates were also lower in the 2 intermediate groups, and the test for trend was significant ($P = .03$). Although the numbers of events were small, the difference was more apparent for deaths due to cerebrovascular disease than for those due to acute MI. There was also a marginally significant 22% lower rate of death due to total cancer in the frequent user vs nonuser group (RR, 0.78; 95% CI, 0.58-1.05), after excluding those with reported cancer before the 7-year follow-up. The test for trend across the 4 groups was statistically significant ($P = .03$), with a lower rate in the 121- to 179-d/y aspirin group as well. The rate of death due to non-CVD and noncancer causes was also lower in the frequent user group (RR, 0.53; 95% CI, 0.39-0.74). These differences were reflected in the rates of total mortality, which was 36% lower in the frequent user group than in the nonuser group, even after excluding those with previous cancer (RR, 0.64; 95% CI, 0.54-0.77). The trend across the 4 groups was highly significant ($P<.001$), with the 121- to 179-d/y group also experiencing lower mortality.

Randomized aspirin assignment was not a significant predictor of any of these outcomes in the posttrial period after controlling for observational aspirin use. No statistically significant interactions of randomized aspirin assignment and posttrial self-selected use were found for MI or stroke (**Table 4**), although the risk reductions for MI and CVD-related mortality were stronger among those who had been randomized to placebo. The interaction was significant and in the same direction for cerebrovascular-related deaths. For cancer-related deaths, other (non-CVD and noncancer) deaths, and total deaths among those with no previous cancer, the RR estimates were comparable among those who had been randomized to aspirin and to placebo. In separate analyses, no significant interactions of age and posttrial aspirin use were found (data not shown).

COMMENT

In the randomized trial, aspirin decreased the risk for a first MI by 44% (95% CI, 30%-55%). There were, however, insufficient numbers of events to reliably determine whether there were reductions for stroke or CVD-related mortality. In the posttrial period, those who self-selected

for aspirin use of at least 180 d/y had rates of MI that were 28% lower than those who took aspirin 0 to 13 d/y, a difference that was statistically significant and consistent with the 44% reduction seen in the randomized trial. There was no significant decrease or increase in total or ischemic stroke in the randomized or posttrial periods. Although based on very small numbers, the randomized results suggested a possible increase in hemorrhagic stroke with aspirin use that was not apparent in the observational data. In addition, in the posttrial period, there was a significant reduction in CVD-related mortality with self-selected aspirin use, including mortality due to cerebrovascular causes, cancer, and noncancer and non-CVD causes and, as a result, in total mortality, findings that were not seen during the randomized period.

DATA FROM from randomized trials show a clear impact of aspirin on CVD, although data on mortality are limited. In primary prevention, an overview of the PHS and British Doctors' Trial found a 33% reduction in nonfatal MI, but no clear effect on CVD-related death or nonfatal stroke.[15] In addition, there was no significant effect on ischemic stroke, but a possible 1.9-fold increase in risk for hemorrhagic stroke, based on small numbers. Secondary prevention trials, as well as trials of suspected evolving MI,[16] have shown a clear benefit of aspirin use on stroke as well as on MI and death due to vascular causes. In an overview of 142 trials of antiplatelet therapy in high-risk patients who had survived a previous occlusive event, nonfatal MI and nonfatal stroke were reduced by one third, and death due to vascular causes was reduced by one sixth.[6] Aspirin alone was associated with a highly significant 25% reduction in vascular events. A recent meta-analysis of primary and secondary prevention trials with stroke subtype information found significant reductions in MI, stroke, and CVD-related mortality, with an increase in risk for hemorrhagic stroke.[17] The totality of evidence from randomized trials thus suggests a definite benefit of aspirin on MI, stroke, and CVD-related death in secondary prevention and in treatment of a suspected evolving MI. The evidence concerning primary prevention indicates a clear protective effect for MI, at least in men, but insufficient data are available for stroke or CVD-related death, and a possible small increased risk for hemorrhagic stroke remains plausible.

With respect to observational studies of aspirin, the earliest show a reduction in risk for first MI among men and women[7] but no significant benefit concerning deaths due to coronary heart disease.[8] A study among the elderly found a nonsignificant reduction in MI among men, but possible increases in ischemic heart disease among women.[9] A more recent prospective study among middle-aged women in the Nurses' Health Study found a significant reduction in first MI among those taking aspirin 1 to 6 times a week compared with those taking none[10] and nonsignificant reductions in CVD-related deaths and important vascular events. No effects on stroke were seen. In addition, women who self-selected aspirin 7 or more times a week had no decreases in MI, stroke, or death. In the Cardiovascular Health Study, examining short-term predictors of stroke among the elderly, an increase was seen among men and women self-selecting aspirin that appeared to be stronger in the subgroup with no previous CVD.[11] In a more recent analysis after 4.2 years of follow-up in the same population, women, but not men, who used aspirin frequently experienced higher rates of ischemic stroke, and in both sexes combined there was an increase in hemorrhagic stroke.[12] These observational findings need to be interpreted with caution, however, because questions remain concerning the reasons for aspirin use and doses used, and because the observed increased risk for ischemic stroke is not consistent with the totality of evidence from randomized trials.[18]

In the analysis of posttrial self-selected aspirin use in the PHS, different dose-response effects were seen for various outcomes, although the numbers in the middle 2 categories of aspirin use were small. The observed reduction in MI was restricted to those taking aspirin at least 180 d/y. These results are consistent with an earlier analysis in the PHS that showed the strongest benefit on MI among those with the highest levels of adherence.[19] The confidence intervals for the 2 middle posttrial aspirin groups were wide, however, and could not exclude even a large reduction. For mortality, significant risk reductions were also observed in lower categories of use, especially for CVD mortality. These reductions in mortality, however, especially for non–CVD-related mortality, are not consistent with trial results.

The differences seen according to frequency of aspirin use as well as apparent discrepancies among the trial and observational results, particularly for mortality, may be due to residual confounding. Several risk factors were highly predictive of self-selected aspirin use in these data, including age, BMI, family history of MI, treatment of hypertension, and elevated cholesterol level. Smoking was not predictive of aspirin use, but other lifestyle factors were associated with more frequent self-selection of aspirin, including exercising at least once per week, more frequent alcohol consumption, and use of vitamin E supplements. Risk factors for CVD as well as an interest in prevention, as assessed through measures such as exercise, antioxidant use, and lipid-lowering medications, are thus associated with the frequency of aspirin use among these physicians.

Correlates of self-selected aspirin use seen in this study were similar to those seen among women in the Nurses' Health Study.[10] Predictors in common included the traditional CVD risk factors of hypertension, high cholesterol levels, higher body weight, and family history of MI. In the Nurses' Health Study, however, aspirin users were more likely to smoke and exercised less than nonusers, although they also reported higher alcohol consumption. Both study populations differ in that less than 10% of the nurses taking aspirin were doing so for primary prevention of CVD. Although not undergoing assessment in the PHS, the latter proportion is likely higher in the population of physicians enrolled in a trial in which aspirin clearly reduced the risk for first MI. The prevalence of self-selected aspirin use was also much higher in the PHS, with 60% vs 22% in the Nurses' Health Study reporting use at least once every other day.

Thus, despite multivariate adjustment for a large number of risk factors, uncontrolled confounding by indication and unmeasured health behavior is likely to remain in these observational data. The unexpected reductions seen in non–CVD-related mortality, in particular, are unsupported in trial data, and may be due to such residual confounding. In primary prevention, the benefit of aspirin on MI is clear, but effects on stroke and vascular death remain unclear because of inadequate numbers of these events in randomized trials. Thus, the most reliable evidence on the balance of risks and benefits of aspirin will accrue from large-scale and long-term randomized trials, most notably the ongoing Women's Health Study, testing the effect of low-dose aspirin and vitamin E on CVD and cancer among 39 876 female health professionals.[20]

Accepted for publication June 29, 1999.

This work was supported by grants CA-34944, CA-40360, HL-26490, HL-34595, and HL-58476 from the National Institutes of Health, Bethesda, Md.

The authors are indebted to Martin Van Denburgh for his expert programming assistance.

Corresponding author: Nancy R. Cook, ScD, Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Ave E, Boston, MA 02215-1204 (e-mail: ncook@rics.bwh.harvard.edu).

REFERENCES

1. Hennekens CH, Buring JE, Sandercock P, Collins R, Peto R. Aspirin and other antiplatelet agents in the secondary and primary prevention of cardiovascular disease. *Circulation.* 1989;80:749-756.
2. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1989;321:129-135.
3. Data Monitoring Board of the Physicians' Health Study. Issues in the early termination of the aspirin component of the Physicians' Health Study. *Ann Epidemiol.* 1991;1:395-405.
4. Peto R, Gray R, Collins R, et al. A randomised trial of the effects of prophylactic daily aspirin in British male doctors. *BMJ.* 1988;296:320-331.
5. The Medical Research Council's General Practice Research Framework. Thrombosis prevention trial: randomised trial of low-intensity oral anticoagulation with warfarin and low-dose aspirin in the primary prevention of ischaemic heart disease in men at increased risk. *Lancet.* 1998;351:233-241.
6. Antiplatelet Trialists' Collaboration. Collaborative overview of randomized trials of antiplatelet therapy, I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ.* 1994; 308:81-106.
7. Boston Collaborative Drug Surveillance Group. Regular aspirin intake and acute myocardial infarction. *BMJ.* 1974;1:440-443.
8. Hammond EC, Garfinkel L. Aspirin and coronary heart disease: findings of a prospective study. *BMJ.* 1975;2:269-271.
9. Paganini-Hill A, Chao A, Ross RK, Henderson BE. Aspirin use and chronic diseases: a cohort of the elderly. *BMJ.* 1989;299:1247-1250.
10. Manson JE, Stampfer MJ, Colditz GA, et al. A prospective study of aspirin use and primary prevention of cardiovascular disease in women. *JAMA.* 1991;266: 521-527.
11. Manolio TA, Kronmal RA, Burke GL, O'Leary DH, Price TR, for the CHS Collaborative Research Group. Short-term predictors of incident stroke in older adults: the Cardiovascular Health Study. *Stroke.* 1996;27:1479-1486.
12. Kronmal RA, Hart RG, Manolio TA, Talbert RL, Beauchamp NJ, Newman A. Aspirin use and incident stroke in the Cardiovascular Health Study. *Stroke.* 1998; 29:887-894.
13. Hennekens CH, Buring JE, Manson JE, et al. Lack of effect of long-term supplementation with beta-carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med.* 1996;334:1145-1149.
14. World Health Organization. *Ischaemic Heart Disease Registers: Report of the Fifth Working Group (Including a Second Revision of the Operating Protocol): Copenhagen, 26-29 April 1971.* Copenhagen, Denmark: Regional Office for Europe, World Health Organization; 1971.
15. Hennekens CH, Peto R, Hutchison GB, Doll R. An overview of the British and American aspirin studies. *N Engl J Med.* 1988;318:923-924.
16. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet.* 1988;2:349-360.
17. He J, Whelton PK, Vu B, Klag MJ. Aspirin and risk of hemorrhagic stroke: a meta-analysis of randomized controlled trials. *JAMA.* 1998;280:1930-1935.
18. Buring JE, Bogousslavsky J, Dyken M. Aspirin and stroke [editorial]. *Stroke.* 1998; 29:885-886.
19. Glynn RJ, Busing JE, Manson JE, LaMotte F, Hennekens CH. Adherence to aspirin in the prevention of myocardial infarction. *Arch Intern Med.* 1994;154:2649-2657.
20. Women's Health Study Research Group. The Women's Health Study: rationale and background. *J Myocard Ischemia.* 1992;4:30-40.

American Journal of Therapeutics 9, 199–205 (2002)

Confounding by Indication in Epidemiologic Studies of Commonly Used Analgesics

Lisa B. Signorello,[1,2]* Joseph K. McLaughlin,[1,2] Loren Lipworth,[1,2] Søren Friis,[3] Henrik Toft Sørensen,[2,4] and William J. Blot[1,2]

Confounding by indication is a bias frequently encountered in observational epidemiologic studies of drug effects. Because the allocation of treatment in observational studies is not randomized and the indication for treatment may be related to the risk of future health outcomes, the resulting imbalance in the underlying risk profile between treated and comparison groups can generate biased results. Confounding by indication is often present in studies of drugs that are not widely prescribed, because the indications for their use are narrow and not likely to be present in comparison groups; however, this bias is also observed in the study of widely used over-the-counter and prescription drugs, as exemplified by studies of analgesics. In this article we review examples from the published literature to demonstrate how confounding by indication can affect the findings of pharmacoepidemiologic studies relating analgesic use to various health outcomes.

Keywords: non-narcotic analgesics, confounding factors, epidemiology, acetaminophen, aspirin.

INTRODUCTION

In epidemiology, when one seeks to estimate the effect of an exposure on an outcome (for example, a disease or other medical event), a confounder is an extraneous factor that, if not properly accounted for, can bias or distort the association observed between the exposure and outcome. In order for a factor to be considered a confounder, it must satisfy specific criteria. A confounding factor must be: a risk factor for the outcome, independent of the exposure of interest; associated with the exposure in the source population from which the cases arise; and not be an intermediate step in the causal pathway between the exposure and the outcome.[1] In the presence of confounding, if one calculates a crude estimate of the effect of the exposure on the outcome, one would partially see the effect of the exposure and partially the effect of the confounder. This bias prevents investigators from validly assessing a cause-and-effect relationship. Numerous sources exist that provide greater detail about the general issue of confounding.[1,2]

Confounding by indication is the term given to a specific type of confounding often encountered in observational epidemiologic studies of drugs or other clinical therapies. Unless the allocation of treatment is dictated by a randomization process (ie, in the context of a randomized intervention trial), there will always be an indication for treatment — that is, a reason why some patients receive a particular treatment and others do not. The indication for treatment is rarely easy to characterize, because it is typically a combination of several factors that contribute to the physician's decision-making process about whether to assign a particular treatment or, in the case of over-the-counter (OTC) medications, to the individual's decision to self-medicate with a particular drug. Factors that make up the indication (or, just as importantly, the contraindication) for treatment can include the stage or severity of the disease, family history of disease, symptoms, concomitant medical conditions, concurrent therapies,

[1]*International Epidemiology Institute, Rockville, Maryland, USA;*[2]*Department of Medicine, Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA;*[3]*Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark;*[4]*Department of Clinical Epidemiology, Aarhus University Hospital and Aalborg Hospital, Aarhus, Denmark.*

**Address for correspondence: International Epidemiology Institute, 1455 Research Blvd., Suite 550, Rockville, MD 20850, USA. E-mail: lisa@iei.ws*

DOI: 10.1097/01.MJT.0000015792.80073.AE

past response to other treatments, the physician's personal estimation of the prognosis of the patient, the patient's individual preferences, and the age or sex of the patient.[3,4]

Unfortunately, the indication for treatment is usually not present in comparison groups of nontreated individuals, and usually predictive—if only indirectly—of future health outcomes. This creates a situation whereby the indication itself becomes a confounder (thus the term "confounding by indication"), and the underlying risk profile or baseline prognosis of treated individuals is likely to differ from nontreated individuals, rendering the two groups noncomparable.

It is often true that, among patients with the same disease, those with a worse prognosis are more heavily represented in the treated group. In addition, among patients who have the same disease but undergo different treatments, certain treatments are often reserved for the most ill.[5] For example, in a hypothetical population of hypertensive patients, assume that 90% of those treated with a particular drug—but only 15% of those untreated—have an advanced or severe form of the disease. In a study of the possible effect of the drug on mortality, comparing the drug-treated patients to a randomly sampled control group from the general population is hazardous, because the inevitably higher mortality rate among those taking the drug may be incorrectly attributed to the drug and not to the higher baseline risk of death among the drug users, because of their underlying hypertension. Even comparing the treated hypertensive patients to the nontreated hypertensive patients would be inadequate, because—given the disparity in the severity of disease between the two patient groups—mortality would be higher, independent of treatment, for the treated than the untreated group. In crude (or incompletely adjusted) analyses, one would see the effect of poor prognosis mixed with any treatment effect of the drug, potentially masking any beneficial effect that the drug provides.[3,6,7]

Confounding by indication is often present in studies of drugs that are not widely prescribed, because the indications for their use are narrow and not likely to be present in comparison groups; however, this bias is also frequently noted in the study of widely used prescription and OTC drugs, as exemplified by studies of analgesics. In this article we present examples in which confounding by indication may have played a role in the findings of epidemiologic studies examining the relationship between analgesics and various health outcomes. The examples selected in this article have important public health implications, because the exposures studied are exceptionally common (ie, aspirin and acetaminophen) and the health outcomes studied impart a significant burden in most populations (ie, cardiovascular disease, renal failure, asthma, and upper gastrointestinal disorders).

ASPIRIN AND MYOCARDIAL INFARCTION

The Physicians' Health Study is among a number of randomized trials that have provided strong evidence to show that low-dose aspirin is a significant factor in the primary and secondary prevention of adverse cardiovascular outcomes.[8–12] The Physicians' Health Study was a randomized, double-blind, placebo-controlled trial that had as one of its objectives to estimate the protective effect of aspirin on cardiovascular disease and mortality. This component of the study (randomization to aspirin intake every other day) was halted after an average of 5 years of follow-up when a clear and significant 44% reduction in the risk of first myocardial infarction (MI) was detected.[9] All participants were then offered the opportunity to take aspirin after the trial stopped, and the study population remained under observation.

After the termination of aspirin randomization, in the posttrial period (at 7 years follow-up), Physicians' Health Study participants completed a follow-up questionnaire that asked about use of aspirin. Twelve years into the trial, the investigators performed a follow-up to their original analysis, affording an opportunity to directly compare the earlier randomized results with new, essentially observational results (resulting from self-selected aspirin use).[13] This provided a unique opportunity to directly estimate the magnitude of possible confounding by indication that would be present in an observational study of aspirin intake and MI.

The term *confounding by indication* is appropriate because the indication for taking aspirin (as a result of medical advice or personal choice) was independently related to the risk of MI. Although the earlier randomization assured balance between patients assigned and not assigned aspirin during the trial, in the posttrial period, patients on aspirin therapy were not comparable to those who did not take aspirin with respect to the underlying risk of an MI event. The Physician's Health Study investigators compared the characteristics of subjects who reported taking aspirin at least 180 days out of the previous year with those who took less or none at all.[13] Subjects who chose to take aspirin for 180 days or more (compared with nonusers) were: 1) slightly heavier, 2) slightly older, 3) about 30% more likely to have a family history of MI, 4) almost 20%

more likely to be under treatment of hypertension, 5) almost 50% more likely to be under treatment to lower their cholesterol (and still had higher cholesterol levels), and 6) about 45% more likely to be daily alcohol drinkers. Conversely, they were shown to be more likely to exercise at least once per week and also to take vitamin E supplements. Hence, the baseline profiles between the aspirin and comparison groups differed. Assessing the first MI risk from 7 years to 12 years of follow-up—using self-selected aspirin use—relative risks of 0.83 (95% CI = 0.52–1.31), 0.90 (95% CI = 0.61–1.33), and 0.72 (95% CI = 0.55–0.95) for takers of 14 to 120 aspirin, 121 to 179 aspirin, and ≥180 aspirin per year, respectively, were observed.[13] These results were adjusted for more than 15 confounding factors, including age, body mass index, smoking, exercise, personal and family history of cardiovascular problems, comorbidities, and other clinical factors. The difference between these results and those of the randomized trial may then be, at least in part, because of uncontrolled confounding by indication.

Although in this particular exercise the difference between the randomized and the observational results was not extreme, under another set of circumstances—especially in a more heterogeneous population or in a study where complete information on confounders was not collected—it could have been, and the authors clearly demonstrated the potential for these confounding factors to mask the protective effect of aspirin.

ANALGESICS AND RENAL FAILURE

Conclusive evidence has linked renal failure to heavy use of phenacetin-containing analgesic preparations[14]; however, in the time since phenacetin was removed from the market, several epidemiologic studies have also suggested that other analgesics, namely aspirin and acetaminophen (paracetamol, a major metabolite of phenacetin[15]), could be associated with the risk of renal failure, and a number of review papers have been published on this topic.[16–20] Most studies have used patients who are enrolled late in the natural history of chronic renal failure or patients who have reached end-stage renal disease.[16] This is problematic because, at the onset of renal failure, doctors sometimes recommend a switch from using aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) to using acetaminophen when an analgesic or antipyretic drug is needed, because of acetaminophen's relative safety with regard to renal function.[21,22]

In a case-control study in the United States, an increased risk of end-stage renal disease was associated with acetaminophen, but a decreased risk was associated with aspirin,[23] possibly reflecting this switching from the one analgesic to the other. In a recent case-control study that we performed in Sweden,[24] we found a positive association between both acetaminophen and aspirin and the risk of chronic renal failure. Despite the fact that the cases were identified in the early stages of renal failure, all had pre-existing renal or systemic disease that was likely to be symptomatic in the years preceding their enrollment in the study. Symptoms of the underlying conditions predisposing to renal failure could have triggered analgesic consumption among these patients, possibly accounting for some measure of the observed association. This variant of confounding by indication is often called *protopathic bias*.[25] Protopathic bias occurs when drugs are prescribed to treat symptoms that are actually early manifestations of the outcome of interest (or manifestations of precursors to the outcome of interest).[3,25,26] Technically, this type of bias is not true confounding, but a reversal of cause and effect, in which the outcome precedes the exposure of interest.[27] Because analgesics are used to treat aches and pains, which are frequently the harbingers of more serious disease, epidemiologic studies relating analgesic use to disease outcomes are especially vulnerable to protopathic bias.

ACETAMINOPHEN AND CANCER MORTALITY

We recently conducted a study to investigate the association between acetaminophen use and mortality from cancer and other diseases. Using a population-based prescription database and national mortality files in Denmark, we compared the mortality rates of nearly 50,000 adults prescribed acetaminophen to those observed in the general population.[28] Overall, acetaminophen users exhibited a uniform two- to three-fold excess of mortality for all site-specific cancers (excesses of mortality for all other causes of death were also observed, although the increases were generally not as great as for cancer). A systematic pattern emerged, whereby statistically significant four- to eight-fold increases in cancer mortality were observed during the first year after the acetaminophen prescription, but all relative risks were sharply attenuated, and some became close to unity or nonsignificant beginning in the second or third year of follow-up. There was also no indication that cancer mortality increased with increasing numbers of acetaminophen prescriptions. The nonspecific results implicating acetaminophen as a risk factor for every type of cancer death, the mortality excess decreasing or disappearing with increasing follow-up, and the lack of dose-response all argue against a causal association. The most plausible explanation is that individuals who are suffering from cancer are prescribed acetaminophen for pain relief during time periods 1 or more years before death.

Hence, the relationship observed between acetaminophen use and cancer mortality is most likely an artifact generated by protopathic bias in which, in this exaggerated case, the drug is prescribed after overt manifestation of the disease rather than precursor symptoms of the disease are discovered, making the bias much easier to identify.

ACETAMINOPHEN AND ASTHMA

Recent studies have suggested an association between acetaminophen use and asthma. An international ecologic analysis found that sales of this drug were positively correlated with asthma prevalence at the national level.[29] At the individual level, the same authors also reported a significant association between acetaminophen use and asthma[30]; however, this case-control study used prevalent asthma cases and evaluated analgesic usage after the onset of the disease.[30] Thus, it is likely that the indication for taking acetaminophen could be related to the disease itself. Asthmatics are often recommended to take acetaminophen for analgesic purposes because of the severe adverse reactions that aspirin and other NSAIDs can induce in many asthmatics.[31,32] It is also possible that asthmatics may have a higher prevalence of comorbidities than healthy controls and may seek to relieve symptoms associated with asthma or headaches induced by β_2-agonist treatment. Thus, the causal nature of the reported association between acetaminophen and asthma is questionable and requires a more appropriate study design to assess.

ACETAMINOPHEN AND UPPER GASTROINTESTINAL DAMAGE

The consumption of aspirin and NSAIDs long has been implicated, convincingly, as an etiologic factor in upper gastrointestinal (GI) disorders such as ulcers and bleeding.[33,34] For decades, patients prone to gastric damage, or who have a history of gastric or duodenal ulcer or upper GI bleeding in particular, are often counseled to replace their aspirin and NSAID use with acetaminophen.[33,35] This pattern of prescribing is well-established in clinical practice and has also been documented in epidemiologic studies. For example, McIntosh et al[36] directly addressed this issue within their case-control study of analgesics and peptic ulcer. These authors separated new and recurrent cases of peptic ulcer, and showed that the acetaminophen use of new and recurrent cases was significantly different; more than twice as many recurrent cases than new cases used acetaminophen daily. Among the cases with recurrent ulcer, 62% of those taking acetaminophen daily reported that they had received medical advice to use acetaminophen instead of aspirin or other antiarthritic drugs because of their history of ulcer. Overall, despite the obvious association between acetaminophen intake and recurrent ulcer, these investigators noted no association between daily use of acetaminophen and risk of a first diagnosis of peptic ulcer (relative risk = 1.2, 95% CI = [0.5–2.6] for an average daily dose of 1050 mg among the cases, for more than 4 consecutive weeks in the past 6 months).[36] This study demonstrates that failure to distinguish between new and chronic cases could create an artifactual association between acetaminophen and ulcer. A previous diagnosis of ulcer has clear implications for subsequent exposure to different analgesics, in particular the avoidance of medications considered to be ulcerogenic.[36] Thus, in this case, the indication for taking acetaminophen is likely to be linked to one's underlying risk of future adverse gastric outcomes, and any study that attempts to establish an association between acetaminophen and these outcomes will need to overcome this confounding by indication.

One such example is a recent case-control study of upper GI complications in relation to prescribed acetaminophen and NSAID use.[37] Although several previous epidemiologic studies have found no association between acetaminophen intake and adverse GI effects,[33,36,38–40] and clinical evaluations suggest only minor amounts of gastric mucosal damage in response to acetaminophen use,[41,42] Garcia Rodriguez and Hernandez-Diaz[37] found that current users of prescribed acetaminophen (at doses greater than 2 g per day) were about three and a half times more likely to suffer from upper GI complications than nonusers. The authors used a general practitioners' research database to identify their study subjects. Detailed information on the length of time that patients were covered by the computerized database (ie, how much of their medical history would be available to the investigators) was not provided, and it is likely that complete or long-term medical histories were not available for many subjects. In that case, predisposing medical events or diagnoses (for example, incidences of dyspepsia, gastroesophageal reflux disease, or ulcer, which would make up part of the indication or contraindication for drug treatment) would be missed. These types of conditions would clearly be risk factors for future upper GI disorders, as the authors demonstrate with their study data, in which prior GI conditions (independent of any analgesic treatment) strongly increased the risk of future upper GI complications.[37] Without complete data on these aspects of the subjects' medical history, the investigators would be unable to successfully use stratification or statistical adjustment to fully control for the confounding influence of these factors. When

the authors adjusted for the data that they did have (including the available information on antecedents of upper GI disorders), the relative risks they calculated fell dramatically, indicating strong confounding bias. Since the information on predisposing conditions is likely to be incomplete, residual confounding by indication is nearly unavoidable, and the relative risk estimates are therefore potentially biased.

As mentioned previously, patients known or suspected to be at high risk of GI complications may preferentially be prescribed acetaminophen rather than NSAIDs. This is especially true when high doses are required, because high-dose NSAIDs are known to be associated with very large increases in gastric complication risk.[34] Consequently, even the dose-response relationship observed by Garcia Rodriguez and Hernandez-Diaz[37] could be attributable to confounding by indication. Physicians would be especially careful not to prescribe high doses of NSAIDs to those at greatest risk of GI bleeding, and patients who are switched to increasingly higher doses of acetaminophen may be the groups at increasingly higher risk of adverse outcomes.

Another potential explanation for an association between acetaminophen and upper GI disorders is protopathic bias, when the GI disorder precedes and triggers the use of acetaminophen. It has been shown that analgesic use may be prompted by the symptoms associated with GI discomfort.[43] In their case-control study, Langman et al[43] found a modest positive association with acetaminophen use in the time period immediately preceding hospital admission for upper GI bleeding, but no association with longer-term regular use of acetaminophen in the 3 months preceding admission. Taking into account the reported reasons for taking acetaminophen, the association between acetaminophen and GI bleeding was not detected when the reason for use was non–GI-related (ie, for headache, colds, influenza), but was observed only when the reason was indigestion.

CONCLUSIONS

Use of OTC and prescription analgesics is widespread, and new products are frequently introduced. Hypotheses will arise continually regarding the association between these drugs and various health outcomes. Thus, investigators are increasingly faced with the challenge of addressing confounding by indication through design or analytic methods, and readers of the literature are increasingly faced with the challenge of critically interpreting the results of these studies. In studies of commonly used analgesics, these challenges may be greater than those encountered when studying other drugs. Unlike more specialized prescription drugs, which may have well-defined indications for use, analgesics are used universally for a broad and often nonspecific array of indications. Therefore, the task of identifying the constellation of factors that make up the indication for their use—factors that need to be accounted for in analysis—can be more difficult, if not impossible. When studying adverse effects of prescription drugs, data on the indication for prescribing may be accessible in the patients' medical records, or in computerized databases, which is usually not the case for OTC analgesics. Thus, investigators more often need to rely on collecting data on important confounders via self-report, which may lead to higher degrees of information bias. Regular or heavy users of analgesics also often have a high level of comorbidity[44] that can easily obscure the relationship between analgesics and future health outcomes, if not properly accounted for. Finally, as mentioned previously, because pain is a common symptom that precedes the clinical diagnosis of countless health outcomes, protopathic bias can spuriously generate associations between analgesics and the disorders that triggered their use, reversing the true sequence of cause and effect.

Although new analytic techniques that seek to counter the effects of confounding by indication (including case-crossover and case-time control designs[45] and propensity score methods[46,47]) are at the disposal of investigators, this bias is still very difficult to avoid in observational studies. A critical awareness of this methodologic problem is therefore warranted when evaluating pharmacoepidemiologic studies regarding analgesic use. This is particularly true for clinicians, who often base decisions in medical practice on the available published evidence, and are often called upon to interpret publicized study findings for their patients.

REFERENCES

1. Rothman KJ: Objectives of Epidemiologic Study Design. In: *Modern Epidemiology*. Little, Brown and Company, Boston, 1986, pp77–98.
2. Walker AM: Confounding. In: *Observation and Inference: An Introduction to the Methods of Epidemiology*. Epidemiology Resources, Inc., Newton Lower Falls, 1991, pp119–128.
3. Signorello LB, McLaughlin JK, Lipworth L, et al: Confounding by indication: a conceptual and illustrative review. (Unpublished data.)
4. Walker AM, Stampfer MJ: Observational studies of drug safety. Lancet 1996;348:489.

5. Walker AM: Confounding by indication. Epidemiology 1996;7:335–336.
6. Miettinen OS: The need for randomization in the study of intended effects. Stat Med 1983;2:267–271.
7. Strom BL, Melmon KL: The use of pharmacoepidemiology to study beneficial drug effects. In: *Pharmacoepidemiology*, 3rd ed. (Ed. Strom BL). John Wiley & Sons, Chichester, 2000, pp553–572.
8. Hennekens CH, Buring JE, Sandercock P, et al: Aspirin and other antiplatelet agents in the secondary and primary prevention of cardiovascular disease. Circulation 1989;80:749–756.
9. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. N Engl J Med 1989;321:129–135.
10. Hennekens CH, Peto R, Hutchison GB, et al: An overview of the British and American aspirin studies. N Engl J Med 1988;318:923–924.
11. Antiplatelet Trialists' Collaboration: Collaborative overview of randomized trials of antiplatelet therapy, I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ 1994;308:81–106.
12. Antithrombotic Trialists' Collaboration: Collaborative meta-analysis of randomized trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ 2002;324:71–86.
13. Cook NR, Hebert PR, Manson JE, et al: Self-selected posttrial aspirin use and subsequent cardiovascular disease and mortality in the Physicians' Health Study. Arch Int Med 2000;160:921–928.
14. Elseviers MM, De Broe ME: The implication of analgesics in human kidney disease. In: *Analgesic and NSAID-Induced Kidney Disease*. (Ed. Stewart JH). Oxford University Press, New York, 1993, pp32–47.
15. Hinson JA: Reactive metabolites of phenacetin and acetaminophen: A review. Environ Health Perspect 1983;50: 37–49.
16. McLaughlin JK, Lipworth L, Chow WH, et al: Analgesic use and chronic renal failure: A critical review of the epidemiologic literature. Kidney Int 1998;54:679–686.
17. Delzell E, Shapiro S: A review of epidemiologic studies of nonnarcotic analgesics and chronic renal disease. Medicine 1998;77:102–121.
18. Feinstein AR, Heinemann LA, Curhan GC, et al: Relationship between nonphenacetin combined analgesics and nephropathy: a review. Kidney Int 2000;58:2259–2264.
19. Barrett BJ. Acetaminophen and adverse chronic renal outcomes: an appraisal of the epidemiologic evidence. Am J Kidney Dis 1996;28(Suppl):S14–S19.
20. Klag MJ, Whelton PK, Perneger TV: Analgesics and chronic renal disease. Curr Opin Nephrol Hypertens 1996;5:263–241.
21. Morrison G: Kidney. In: *Current Medical Diagnosis and Treatment*. (Eds. Tierney LM Jr, McPhee SJ, Papadakis MA). Appleton & Lange, Norwalk, 1995, pp775.
22. Lakkis FG, Martinez-Maldonado M: Conservative management of chronic renal failure and the uremic syndrome. In: *The Principles and Practice of Nephrology*. (Eds. Jacobson HR, Striker GE, Klahr S). Mosby, New York, 1995, pp616.
23. Perneger TV, Whelton PK, Klag MJ: Risk of kidney failure associated with the use of acetaminophen, aspirin, and nonsteroidal anti-inflammatory drugs. N Engl J Med 1994;331:1675–1679.
24. Fored CM, Ejerblad E, Lindblad P, et al: Acetaminophen, aspirin, and chronic renal failure. N Engl J Med 2001; 345:1801–1808.
25. Salas M, Hofman A, Stricker BHC: Confounding by indication: An example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981–983.
26. Feinstein AR: *Clinical Epidemiology: The Architecture of Clinical Research*. W.B. Saunders, Philadelphia, 1985.
27. Shapiro S: Confounding by indication? Epidemiology 1997;8:110–111.
28. Lipworth L, Friis S, Mellemkjaer L, et al: A population-based cohort study of mortality among adults prescribed paracetamol in Denmark. (Unpublished data.)
29. Newson RB, Shaheen SO, Chinn S, et al: Paracetamol sales and atopic disease in children and adults: an ecological analysis. Eur Respir J 2000;16:817–823.
30. Shaheen SO, Sterne JA, Songhurst CE, et al: Frequent paracetamol use and asthma in adults. Thorax 2000;55: 266–270.
31. National Asthma Education Program Expert Panel. Guidelines for the Diagnosis and Management of Asthma. National Asthma Education Program – Expert Panel Report. Bethesda, MD: National Institutes of Health and US Department of Health and Human Services, August 1991. Publication No. 91–3042, pp133.
32. Dukes MNG (ed): *Meyler's Side Effects of Drugs*. Elsevier, Amsterdam, 1996, pp193.
33. Ivey KJ: Gastrointestinal effects of antipyretic analgesics. Am J Med 1983;75:53–64.
34. Bidaut-Russell M, Gabriel SE: Adverse gastrointestinal effects of NSAIDs: consequences and costs. Best Pract Res Clin Gastroenterol 2001;15:739–753.
35. Elliott DP: Preventing upper gastrointestinal bleeding in patients receiving nonsteroidal antiinflammatory drugs. DICP 1990;24:954–958.
36. McIntosh JH, Fung CS, Berry G, et al: Smoking, nonsteroidal anti-inflammatory drugs, and acetaminophen in gastric ulcer. A study of associations and of the effects of previous diagnosis on exposure patterns. Am J Epidemiol 1988;128:761–770.
37. Garcia Rodriguez LA, Hernandez-Diaz S: Relative risk of upper gastrointestinal complications among users of acetaminophen and nonsteroidal anti-inflammatory drugs. Epidemiology 2001;12:570–576.
38. Blot WJ, McLaughlin JK: Over the counter non-steroidal anti-inflammatory drugs and risk of gastrointestinal bleeding. J Epidemiol Biostat 2000;5:137–142.
39. Levy M, Miller DR, Kaufman DW, et al: Major upper

gastrointestinal tract bleeding. Relation to the use of aspirin and other nonnarcotic analgesics. Arch Intern Med 1988;148:281–285.
40. Jick H: Effects of aspirin and acetaminophen in gastrointestinal hemorrhage. Results from the Boston Collaborative Drug Surveillance Program. Arch Intern Med 1981;141:316–321.
41. Jerussi TP, Caubet JF, McCray JE, et al: Clinical endoscopic evaluation of the gastroduodenal tolerance to (R)-ketoprofen, (R)-flurbiprofen, racemic ketoprofen, and paracetamol: a randomized, single-blind, placebo-controlled trial. J Clin Pharmacol 1998;38(Suppl):19S–24S.
42. Konturek SJ, Obtulowicz W, Sito E, et al: Distribution of prostaglandins in gastric and duodenal mucosa of healthy subjects and duodenal ulcer patients: effects of aspirin and paracetamol. Gut 1981;22:283–289.
43. Langman MJS, Coggon D, Spiegelhalter D: Analgesic intake and the risk of acute upper gastrointestinal bleeding. Am J Med 1983;74:79–82.
44. Crofford LJ: Rational use of analgesic and anti-inflammatory drugs. N Engl J Med 2001;345:1844–1846.
45. Schneeweiss S, Sturmer T, Maclure M: Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf 1997;6(Suppl):S51–S59.
46. Wang J, Donnan PT, Steinke D, et al: The multiple propensity score for analysis of dose-response relationships in drug safety studies. Pharmacoepidemiol Drug Saf 2001;10:105–111.
47. Wang J, Donnan PT: Propensity score methods in drug safety studies: practice, strengths and limitations. Pharmacoepidemiol Drug Saf 2001;10:341–344.

ELSEVIER

Journal of Clinical Epidemiology ■ (2009) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies

Jaclyn L.F. Bosco[a,b,*], Rebecca A. Silliman[a,b], Soe Soe Thwin[a], Ann M. Geiger[c], Diana S.M. Buist[d], Marianne N. Prout[b], Marianne Ulcickas Yood[e], Reina Haque[f], Feifei Wei[g], Timothy L. Lash[a,b]

[a]*Department of Medicine, Geriatrics Section, Boston University School of Medicine, Boston, MA, USA*
[b]*Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*
[c]*Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA*
[d]*Group Health Center of Health Studies, Seattle, WA, USA*
[e]*Yale School of Public Health, New Haven, CT, USA*
[f]*Kaiser Permanente Southern California, Pasadena, CA, USA*
[g]*HealthPartners Research Foundation, Minneapolis, MN, USA*

Accepted 2 March 2009

Abstract

Objective: To evaluate the effectiveness of methods that control for confounding by indication, we compared breast cancer recurrence rates among women receiving adjuvant chemotherapy with those who did not.

Study Design and Setting: In a medical record review-based study of breast cancer treatment in older women ($n = 1798$) diagnosed between 1990 and 1994, our crude analysis suggested that adjuvant chemotherapy was positively associated with recurrence (hazard ratio [HR] = 2.6; 95% confidence interval [CI] = 1.9, 3.5). We expected a protective effect, so postulated that the crude association was confounded by indications for chemotherapy. We attempted to adjust for this confounding by restriction, multivariable regression, propensity scores (PSs), and instrumental variable (IV) methods.

Results: After restricting to women at high risk for recurrence ($n = 946$), chemotherapy was not associated with recurrence (HR = 1.1; 95% CI = 0.7, 1.6) using multivariable regression. PS adjustment yielded similar results (HR = 1.3; 95% CI = 0.8, 2.0). The IV-like method yielded a protective estimate (HR = 0.9; 95% CI = 0.2, 4.3); however, imbalances of measured factors across levels of the IV suggested residual confounding.

Conclusion: Conventional methods do not control for unmeasured factors, which often remain important when addressing confounding by indication. PS and IV analysis methods can be useful under specific situations, but neither method adequately controlled confounding by indication in this study.

Keywords: Confounding by indication; Propensity score; Instrumental variable; Nonrandomized studies; Breast cancer; Chemotherapy

1. Introduction

Confounding by indication remains an often intractable threat to validity in observational studies [1]. Although confounding is best controlled by a randomized design, randomization is not always feasible. For example, patients cannot be randomized to receive placebo when an efficacious therapy is available [2]. Furthermore, trials often exclude patients with preexisting conditions [3], particularly older adults [4]. Nonrandomized designs must evaluate the effectiveness of therapies whose efficacy has been established in select groups by clinical trials, but not in broader populations that might react differently to the therapy. For these and other reasons [3], nonrandomized studies of therapy effectiveness will remain important [1]. In addition, generalizing results from clinical trials with select patient populations may actually cause harm in the heterogeneous populations treated in clinical practice [5].

As an example, clinical trials of adjuvant chemotherapy in women aged 40–59 years with early-stage breast cancer demonstrate its efficacy with reductions in 5-year mortality between 20% and 40% [6], but it is uncertain whether these benefits extend to older women, who bear the majority

* Corresponding author. Department of Medicine, Geriatrics Section, Boston University Medical Center, 88 East Newton Street, Robinson 2, Boston, MA 02118, USA. Tel.: +617-638-8998; fax: +617-638-8387.
E-mail address: jfong@bu.edu (J. L.F. Bosco).

0895-4356/09/$ – see front matter
doi: 10.1016/j.jclinepi.2009.03.001

What is new?

Key Findings

(1) The implementation of propensity score (PS) adjustment does not guarantee comparability between the exposure groups;

(2) A strong instrumental variable (IV) may be confounded;

(3) Restricting to a more homogenous population remains an effective way to control for confounding.

What this adds to what was known

The use of PSs and IV methods is not universally effective across all observational settings.

What is the implication, what should change now

Researchers should understand the limitations and appropriateness of the PS and IV methods in relation to their data before implementing and interpreting results that may be as biased as results generated by conventional methods.

burden of breast cancer occurrence [7]. Nonrandomized studies of older women with early-stage breast cancer suffered from differences in prognosis between women who received adjuvant chemotherapy and women who did not receive adjuvant chemotherapy [8,9], and thus are potentially biased by confounding by indication.

When the validity of a study is threatened by confounding by indication, it is not straightforward to determine which method of adjustment, if any, is most effective in obtaining a valid and precise estimate of effect. Conventional methods to adjust for confounding, such as restriction and multivariable regression, leave residual confounding because of unmeasured factors. Thus, propensity score (PS) adjustment and the instrumental variable (IV) approach have become increasingly popular [10–13], with the intent to address this residual confounding by simulating a randomized environment. PS adjustment theoretically increases comparability between the comparison groups by creating pseudorandomization of measured confounders [14]. The goal of the IV approach is to reduce confounding by indication through the use of a variable that is associated with the exposure, unrelated to the confounders, and has no direct association with the outcome other than through the exposure [15]. However, several investigators have cautioned that these alternative methods are not universal solutions to the problem of confounding by indication [10,11,13,16–18].

Three observational studies used the SEER-Medicare [19] linked data set and found that adjuvant chemotherapy decreased the rate of breast cancer-specific mortality [20] and all-cause mortality [20–22] in older women, with the greatest benefit seen in women with node-positive, estrogen receptor negative tumors [20,21]. Based on these results and those of clinical trials among middle-aged women [6], we expect adjuvant chemotherapy to be protective against breast cancer recurrence in older women. With this prior information in mind, we compared methods used to reduce confounding. We implemented restriction, multivariable regression, PS adjustment, and an IV-like method to estimate incidence rates of breast cancer recurrence in women who received adjuvant chemotherapy compared with women who did not, in the Breast Cancer Treatment Effectiveness in Older Women (BOW) cohort [8,9,23].

2. Methods

2.1. Study population

The BOW cohort study was conducted at six integrated health care systems that are part of the 14-system consortium of the Cancer Research Network (CRN) [24]. The overall goal of the CRN is to increase the effectiveness of preventive, curative, and supportive interventions for major cancers through a program of collaborative research, and to determine the effectiveness of cancer control interventions that span the natural history of major cancers among diverse populations and health systems. The six systems were Group Health Cooperative, Seattle, WA; Fallon Clinic Worcester, MA; Kaiser Permanente Southern California, CA; Lovelace Health System, New Mexico; HealthPartners, Minneapolis, MN; and Henry Ford Health System, Detroit, MI. The institutional review boards of each health care system and the Boston University Medical Center approved this study.

Detailed data collection methods have been described previously [23]. Briefly, our cohort included women 65 years of age or older diagnosed with early-stage (I–IIB) breast cancer between 1990 and 1994 at one of these six integrated health systems. Women with bilateral cancer or other malignancies except nonmelanoma skin cancer were excluded if their diagnosis was within 5 years before, or 30 days after, their initial breast cancer diagnosis. Our exposure of interest was adjuvant chemotherapy, therefore women who received only a biopsy ($n = 22$), neoadjuvant chemotherapy ($n = 3$), or had implausible chemotherapy start and stop dates recorded ($n = 13$) were excluded from this analysis. We will refer to this population as the ‘‘unrestricted cohort.’’

2.2. Data collection

Demographic and tumor characteristics, breast cancer treatments, recurrence, and comorbid conditions were collected via medical record reviews conducted up to 10 years postdiagnosis. Details of the medical record review are described by Thwin et al. [25].

2.3. Analytic variables

2.3.1. Adjuvant chemotherapy

Women who received adjuvant chemotherapy were considered the index group. Among women who received adjuvant chemotherapy, the median length of time to last adjuvant chemotherapy course was 183 days after diagnosis. Type of chemotherapy, start and stop dates, number of courses, and completion were also collected. Women who were not referred, not recommended, refused, or did not receive adjuvant chemotherapy comprised the reference group. Women with no mention of chemotherapy in the medical records were assigned to the reference group.

2.3.2. Follow-up time

We defined the start of follow-up as the date of last adjuvant chemotherapy course (index) or 183 days after diagnosis (reference), and follow-up continued until the diagnosis of breast cancer recurrence, death from any cause, disenrollment from the health care system, or the completion of 10 years of follow-up, whichever came first.

2.3.3. Breast cancer recurrence

Breast cancer recurrence was defined as a tumor pathologically or clinically diagnosed during the follow-up period. Tumors that occurred in the same breast as the original tumor or in any lymph node or distant site were classified as a recurrence. Women with recurrence ($n = 16$) or death ($n = 6$) that occurred before the last date of chemotherapy course or before 183 days after diagnosis were excluded from the analyses.

2.3.4. Patient characteristics

Demographic, tumor, and breast cancer treatment characteristics were considered potential confounders in the association between adjuvant chemotherapy and recurrence. Women were categorized by age at diagnosis (65–69; 70–74; 75–79; ⩾80 years old), race/ethnicity (non-Hispanic White; Hispanic and/or Other Race), tumor size (<1; 1 to <2; 2 to <3; ⩾3 centimeters [cm]), node positivity (negative [no presence of breast cancer in lymph nodes]; 1–3 positive nodes; ⩾ 4 positive nodes; not determined), histologic grade (well differentiated; intermediate or moderately differentiated; poorly differentiated, undifferentiated, or anaplastic; not determined or stated), primary therapy (breast conserving surgery [BCS] only; BCS plus radiation therapy; mastectomy), estrogen receptor (ER) expression (positive; negative; other), progesterone receptor (PR) expression (positive; negative; other), tamoxifen (prescribed; not prescribed), and baseline Charlson Comorbidity Index score (0; 1; ⩾2) [26]. Women who did not have an axillary lymph node dissection were similar to women who were node-negative and the two groups were combined. Women who were recorded as ''other'' for ER expression or PR expression were combined with ER positive expression and PR positive expression, respectively. Women who were prescribed tamoxifen or another hormonal agent ($n = 2$) were classified as having received tamoxifen.

2.4. Data analysis

Descriptive statistics for demographic, tumor, and treatment characteristics were calculated using univariate statistics. These characteristics were also evaluated as potential confounders of the association between adjuvant chemotherapy and breast cancer recurrence using contingency table analyses.

We compared several methods in their ability to obtain valid and precise results, using the prior from trials of younger breast cancer patients as a guide for the expected direction of the effect. Figure 1 illustrates the analytic samples used for each of the analytic methods described later. All analyses were performed using SAS statistical software version 9.1 (SAS Institute, Cary, NC, USA).

2.4.1. Unadjusted analysis

Using Cox proportional hazards regression on the unrestricted cohort, we estimated the hazard ratio (HR) associating receipt of adjuvant chemotherapy vs. not receiving adjuvant chemotherapy.

2.4.2. Restricted analysis

Within the unrestricted cohort, we identified a restricted subset of women as at high risk for recurrence using the St. Gallen [27] criteria from the calendar time of diagnosis (1992). These criteria combine tumor size, node positivity, histologic grade, and ER and PR expression to identify women who are considered at high risk for recurrence. A woman was classified as at high risk if she was node-positive, or node-negative with one of the following three tumor characteristics: (1) poorly differentiated, grade III histology; (2) ER negative and ⩾1 cm diameter; or (3) ER positive and >2 cm diameter. Using this restricted cohort to reduce confounding, we conducted Cox proportional hazards regression to estimate the association between adjuvant chemotherapy and breast cancer recurrence.

2.4.3. Restriction and multivariable regression

Using the restricted cohort, we adjusted for demographic characteristics (age group, race/ethnicity, health care system, baseline Charlson Comorbidity Index score [26]), tumor characteristics (tumor size, node positivity, histologic grade, ER expression, and PR expression), and treatment characteristics (primary therapy, tamoxifen prescription) to estimate the HR of breast cancer recurrence comparing those who received chemotherapy with those who did not.

2.4.4. Propensity score method

A PS is a summary confounder score that is modeled using the exposure as the dependent variable [14,28,29].

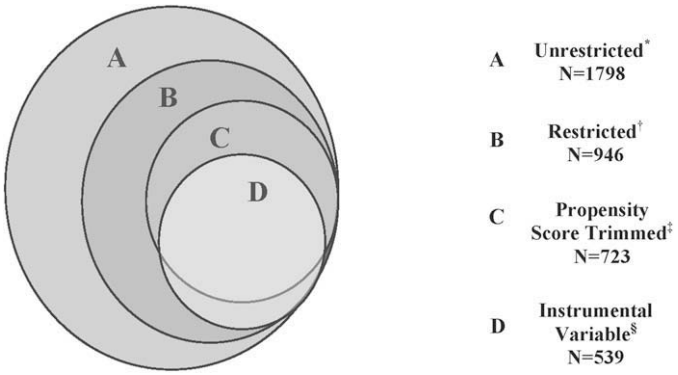

Fig. 1. Venn diagram of analytic sample sizes for each adjustment method used to control for confounding by indication in a study of older women with breast cancer. *BOW cohort exluding women with biopsy only (n=22), neoadjuvant chemotherapy (N=3); inconsistent chemotherapy dates (n=13), and women who had a recurrence (n=16) or died (n=6) before the start of adjuvant chemotherapy or 183 days after date of diagnosis. †Restricted to women who are classified as high-risk for recurrence by the 1992 St. Gallen Criteria.[25] ‡Trimmed the sample to exclude women who did not have overlapping propensity scores. §Excludes 253 women because they were the only patient seen by their surgeon or because they were the chronologically first patient for their surgeon, in our dataset.

Using logistic regression with the restricted cohort, we modeled the probability of receiving adjuvant chemotherapy as a function of the variables included in the multivariable adjusted model. To increase comparability between our index and reference groups, we trimmed the data to include only women with overlapping scores between the index and reference groups. With the trimmed data set, we used Cox proportional hazards regression to model the association between adjuvant chemotherapy and recurrence, using three PS adjustment approaches. First we divided the trimmed sample into PS quintiles. We adjusted for PS quintiles and used the lowest quintile as the reference. Second, we adjusted for the continuous PS measure in the Cox proportional hazards model. Last, we used a doubly robust adjustment, in which we adjusted for the continuous PS and the variables used to predict the probabilities of receiving adjuvant chemotherapy.

As recommended by Sturmer et al. [10], we evaluated the distribution of patient characteristics within the PS quintiles among women who received chemotherapy and women who did not. To assess whether our trimmed data set differed from the restricted cohort, we performed multivariable adjustment on the trimmed data set and compared these results to the results of the restricted cohort.

2.4.5. Instrumental variable-like method

The IV method has been used in analyses when confounding by indication is suspected [18,30–33]. Specifically, the use of our IV-like approach was intended to control for the confounding by unmeasured indications for chemotherapy. Using an approach similar to Brookhart et al.'s preference-based IV method [13,33], in the restricted cohort, we used each patient's surgeon's chronologically preceding patient's receipt of adjuvant chemotherapy (preceding patient within our data set) as the IV within strata of stage and ER expression to estimate the effect of receipt of adjuvant chemotherapy on time to breast cancer recurrence. We used a surgeon's preceding patient's receipt of adjuvant chemotherapy as a surrogate for a medical oncologist's preceding patient's prescription of adjuvant chemotherapy because we did not have information on each patient's medical oncologist (in addition, some patients did not see a medical oncologist). We assigned the IV by stratifying the data set by surgeon. Within each surgeon, the data were sorted by the patient's date of diagnosis in chronological order. Patients of surgeons who only treated one participant in our data set were excluded from the IV-like analysis. For surgeons with more than one patient, the chronologically preceding patient's receipt of chemotherapy was assigned. The chronologically first patient for each surgeon was excluded so that each patient would have an IV defined.

IV-like estimation requires a two-step process. The first step used logistic regression to estimate the probability of receiving adjuvant chemotherapy given the preceding patient's receipt of adjuvant chemotherapy, and included patient characteristics (demographic, tumor, and treatment) in the model. The second step predicted time to recurrence from the probabilities calculated in the first step, using Cox proportional hazards regression and adjusting for patient characteristics.

Using a patient's surgeon's preceding patient's receipt of adjuvant chemotherapy as the IV, we relied on three key assumptions about the properties of the IV: (1) surgeon's previous patient's receipt of adjuvant chemotherapy was independent of the unmeasured risk factors in the current patient (IV not associated with confounders); (2) surgeon's previous patient's receipt of adjuvant chemotherapy was independent of the outcome in the current patient (IV had no direct effect on outcome); and (3) surgeon's previous patient's receipt of adjuvant chemotherapy varies within surgeons (IV associated with exposure).

Following methods outlined by Brookhart and Schneeweiss, we assessed the validity and interpretation of our estimate from our IV-like approach [13]. The strength of

the IV was estimated by performing simple linear regression with the IV as the independent variable and receipt of chemotherapy as the dependent variable in the model. Weassessed the strength of our IV by comparing it with the strength reported by Brookhart and Schneeweiss [13].

We used measured patient characteristics as proxies for unmeasured variables. To evaluate whether our IV assumptions were violated, we calculated the prevalence differences of patient characteristics between the levels of the IV and the prevalence differences of patient characteristics between the two levels of receiving chemotherapy. We assessed the imbalance of these characteristics by calculating prevalence difference ratios between the IV relative to receipt of chemotherapy. Prevalence difference ratios less than the null value of 1 indicated that the patient characteristics were more balanced across the levels of the IV than across the levels of the exposure. The prevalence difference ratios were compared with the strength of the IV. If the prevalence difference ratios were less than the strength of the IV, then the estimate for the association between adjuvant chemotherapy and recurrence using the IV-like method would result in a less biased estimate than using conventional methods [12]. Then we looked at the prevalence differences across the IV. For each characteristic, if the prevalence difference across the IV was not close to zero (no difference), then the IV remained confounded by that characteristic, and residual confounding could not be ruled out.

The widths of the 95% confidence intervals (CIs) around the HRs for each analytic method were calculated as the ratio of the upper limit to the lower limit. Larger widths were interpreted as having less precision.

We repeated each analytic method to assess whether the rate of recurrence varied by type of chemotherapy regimen.

3. Results

Frequencies for demographic and tumor characteristics for the unrestricted cohort who received primary therapy ($n = 1798$), the cohort restricted to women at high risk for recurrence ($n = 946$), the PS analytic sample ($n = 723$), and the IV analytic sample ($n = 539$) are presented in Table 1 by receipt of chemotherapy. For women classified as at high risk, 20% experienced a breast cancer recurrence. In the unrestricted, restricted, PS, and IV samples, a higher proportion of women who received adjuvant chemotherapy were in the youngest age category (65−69 years), had a baseline Charlson score of 0, and were node-positive, whereas a lower proportion of women were ER-positive compared with those who did not receive adjuvant chemotherapy. These differences in distributions illustrate the potential for confounding by indication. Adjustment for tumor characteristics had the largest impact on the effect estimates. Node positivity had that highest magnitude of confounding of 1.7, followed by histology, tumor size, and ER status (each have a magnitude of confounding of about 1.3).

In the unrestricted cohort, receipt of adjuvant chemotherapy was crudely associated with recurrence (HR = 2.6; 95% CI = 1.9, 3.5). After restricting the cohort to women at high risk for recurrence, the HR relating recurrence to receipt of chemotherapy (HR = 1.8; 95% CI = 1.3, 2.5) seemed to be confounded by indications for receipt of chemotherapy, presuming the prior based on clinical trials demonstrating a protective effect holds true in this population [6]. We observed a modest increased hazard rate of breast cancer recurrence in women who received adjuvant chemotherapy compared with those who did not after multivariable regression (HR = 1.1; 95% CI = 0.7, 1.6).

The PS distributions among women who received chemotherapy vs. those who did not showed no substantial overlap (Fig. 2), even after trimming the extreme probabilities of receiving (‘‘All Exposed’’) and not receiving chemotherapy (‘‘All Unexposed’’). Our PS trimmed sample consisted of 723 women at high risk for recurrence. The crude estimate for the PS analytic sample was HR = 1.7 (95% CI = 1.2, 2.5). The PS quintile adjustment method yielded a slightly higher HR (HR = 1.3; 95% CI = 0.8, 2.0) than the multivariable regression method. Both the continuous and the doubly robust PS adjustment methods yielded a HR = 1.1 (95% CI = 0.7, 1.7). The multivariable adjusted association in the PS trimmed sample was similar to what we observed using the multivariable method on the restricted cohort (HR = 1.1; 95% CI = 0.7, 1.7).

For the IV-like method, to ensure that an instrument was assigned for each patient, 253 women were excluded because they were the only patient, in our data set, seen by their surgeon or because they were the chronologically first patient, in our data set, for their surgeon. The final analytic sample included 539 women at high risk. The crude estimate for the IV analytic sample was HR = 2.1 (95% CI = 0.1, 3.8). The IV adjusted estimate was HR = 0.9 (95% CI = 0.2, 4.3), but confounding was not completely controlled. Although all of our prevalence difference ratios were less than the strength of the IV of 23.7%, residual associations between the IV and several measured characteristics—such as histology, tumor size, and node positivity—remained (Table 2). Our prevalence difference ratios were both above and below the null, indicating that for some characteristics (age, comorbidity, tamoxifen prescription, ER expression, and PR expression) the IV was more balanced across levels of the characteristic than the observed exposure, but for others (race, tumor size, node positivity, histology, and primary therapy) the IV was less balanced than the observed exposure. For example, the imbalance in tumor size <1 cm was an absolute difference of 2.99 between those who received adjuvant chemotherapy and those who did not. The imbalance was reduced to 0.67 for the IV prevalence difference, resulting in a prevalence difference ratio of 0.22. Some of these characteristics are important prognostic markers for recurrence risk, so these residual associations portend the potential for residual confounding by indication. Figure 3 depicts the estimates and standard errors for the association between

Table 1
Demographic, tumor, and treatment characteristics in the subjects for the unrestricted cohort, restricted cohort, PS analytic sample, and instrumental variable analytic sample

	Unrestricted cohort[a] ($n = 1798$)				Restricted cohort[b] ($n = 946$)				PS cohort[c] ($n = 723$)				IV cohort[d] ($n = 539$)			
	Chemo		No chemo		Chemo		No chemo		Chemo		No chemo		Chemo		No chemo	
Characteristic	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Age categories (yr)																
65–69	106	63	506	31	99	63	256	32	85	60	244	42	60	66	143	32
70–74	47	28	484	30	47	30	237	30	44	31	199	34	25	27	145	32
75–79	11	6.6	291	18	9	5.7	148	19	9	6.4	91	16	4	4.4	81	18
80+	3	1.8	351	22	3	1.9	148	19	3	2.1	48	8.3	2	2.2	79	18
Race/ethnicity																
Non-Hispanic White	143	86	1325	81	137	87	651	83	121	86	474	81	79	87	379	85
Hispanic and/or Other Race	24	14	302	19	21	13	133	17	20	14	104	18	12	13	65	15
Charlson Comorbidity Index																
0	134	80	1093	67	126	80	531	67	112	79	428	74	76	84	299	67
1	32	19	454	28	31	20	216	27	28	20	146	25	14	15	124	28
2+	1	0.6	85	5.2	1	0.6	42	5.3	1	0.7	8	1.4	1	1.1	25	5.6
Tumor size (cm)																
<1	11	6.6	365	22	10	6.3	65	8.2	9	6.4	52	8.9	5	5.5	38	8.5
1 to <2	52	31	693	42	49	31	254	32	43	31	196	34	31	34	130	29
2 to <3	47	28	368	23	44	28	276	35	40	28	196	34	19	21	162	36
3+	57	34	206	13	55	35	194	25	49	35	138	24	36	40	118	26
Node positivity																
None	43	26	1308	80	40	25	472	60	40	28	290	50	21	23	255	57
1 to 3 nodes	64	38	262	16	60	38	255	32	55	39	231	40	35	38	155	35
4+ nodes	60	36	62	3.8	58	37	62	7.9	46	33	61	10	35	38	38	8.5
Not determined	1	0.6	336	21	1	0.6	110	14	1	0.7	28	4.8	1	1.1	64	14
Histologic grade																
Well differentiated	14	8.4	279	17	13	8.2	66	8.4	13	9.2	55	9.5	9	9.9	45	10
Intermediate/moderate	49	29	615	38	47	30	240	30	44	31	184	32	25	27	141	31
Poorly differentiated/undifferentiated/ anaplastic	89	53	326	20	85	54	316	40	72	61	241	41	53	58	169	38
Not determined/stated	15	9.0	412	25	13	8.2	167	21	12	8.5	102	18	4	4.4	93	21
Estrogen receptor expression																
Positive	71	43	1227	75	68	43	564	71	68	48	404	69	47	52	356	79
Negative	90	54	192	12	85	54	164	21	69	49	140	24	40	44	55	12
Indeterminate/other	6	3.6	213	13	5	3.2	61	7.7	4	2.8	38	6.5	4	4.4	37	8.3
Progesterone receptor expression																
Positive	57	34	980	60	54	34	450	57	53	38	321	55	33	36	274	61
Negative	102	61	385	24	97	61	251	32	82	58	206	35	53	58	125	28
Indeterminate/other	8	4.8	257	16	7	4.4	88	11	6	4.3	55	9.5	5	5.5	49	11
Primary therapy																
BCS only	3	1.8	211	13	2	1.3	75	9.5	2	1.4	17	2.9	1	1.1	49	11
BCS plus radiation therapy	40	24	590	36	40	25	209	26	38	27	171	29	23	25	126	28
Mastectomy	124	74	831	51	116	73	505	64	101	72	394	68	67	74	273	61
Tamoxifen prescribed																
Yes	102	61	1080	66	97	61	624	79	94	67	454	78	63	69	373	83
No	65	39	557	34	61	39	165	21	47	33	128	22	28	31	75	17

Chemotherapy regimen															
CMF	112	67	—	—	109	69	—	—	98	70	—	—	60	66	—	—
Adriamycin-based	47	28	—	—	42	27	—	—	36	26	—	—	27	30	—	—
Other	8	4.8	—	—	7	4.4	—	—	7	5.0	—	—	4	4.4	—	—
Chemotherapy completed																
CMF	96	57	—	—	93	59	—	—	84	60	—	—	53	58	—	—
Adriamycin-based	44	26	—	—	39	25	—	—	34	24	—	—	24	26	—	—
Other	6	3.6	—	—	6	3.8	—	—	6	4.3	—	—	4	4.4	—	—
Not complete	21	13			20	13			17	12			10	11		
Diagnosed with recurrence																
Yes	49	29	191	12	49	31	138	17	42	30	107	18	31	34	78	17
No	118	71	1441	88	109	69	651	83	99	70	475	82	60	66	370	83

Abbreviations: PS, propensity score; CMF, cyclophosphamide−methotrexate−flourouracil; IV, instrumental variable; BCS, breast conserving surgery.

[a] Excludes: 22 biopsy only; 3 neoadjuvant chemotherapy; 13 inconsistent chemotherapy dates; 16 recurrences; and 6 deaths before the start of adjuvant chemotherapy or 183 days after date of diagnosis.

[b] Restricted to women who are classified as high-risk for recurrence by the 1992 St. Gallen criteria.[25]

[c] PS analytic sample after trimming the nonoverlapping scores between the index and the reference groups.

[d] IV analytic sample after excluding patients of surgeons who only treated one patient and the chronologically first patient.

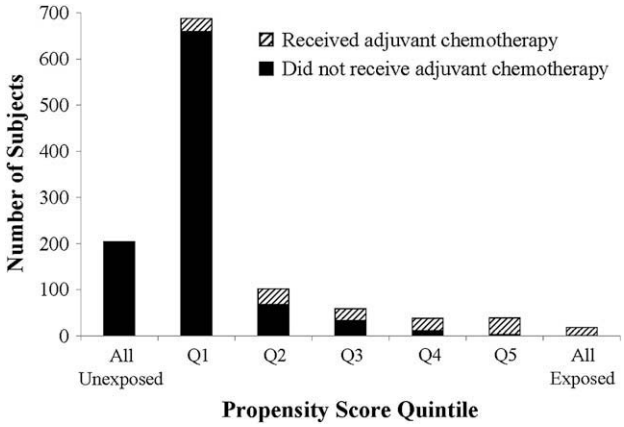

Fig. 2. Propensity score (PS) distribution for adjuvant chemotherapy in older women with breast cancer by quintile. The PS analytic sample trimmed the ''All exposed'' and ''All unexposed'' categories.

adjuvant chemotherapy and breast cancer recurrence for the unadjusted and adjusted methods.

Among the women who received chemotherapy, 67% received a cyclophosphamide−methotextrate−flourouracil (CMF) regimen, 28% received an adriamycin-based regimen, and 4.8% were classified as having another regimen. Because of small numbers in chemotherapy subgroups, we could only examine the effect of CMF chemotherapy regimen on the rate of recurrence. Using the unrestricted, restricted, and PS methods, the results did not change appreciably, except that there were wider intervals around the estimates. Using the IV-like method, the association between CMF and recurrence became slightly more protective (HR = 0.6; 95% CI = 0.1, 3.8).

4. Discussion

The association between receipt of adjuvant chemotherapy and recurrence risk in older women with breast cancer provides a useful example of the manner in which confounding by indication can complicate nonrandomized studies of treatments in general populations. When considering treatment recommendations to reduce breast cancer recurrence, oncologists treating geriatric patients take into account tumor prognostic factors and additional factors, such as life expectancy, physical function, and quality of life [34]. With minimal trial-based information available to inform clinical guidelines, which currently offer no guidance for treating older women with cancer [35], nonrandomized studies are vitally important. However, nonrandomized studies are only reliable when confounding by indication is handled adequately. When treatment with adjuvant chemotherapy among older patients is based on clinical judgment, controlling for prognostic factors alone leaves residual confounding by indication.

Although not intended to control for unmeasured confounding [10−12,14], PS adjustment has been implemented in studies for this reason [21,22]; however, consistent with

Table 2
Assessment of imbalance of measured patient characteristics across levels of IV and exposure (adjuvant chemotherapy) and prevalence difference ratios

Patient characteristics	Prevalence difference across levels of instrument[a]	Prevalence difference across levels of exposure[b]	Prevalence difference ratio[c]
Age categories (yr)			
65−69	18.16	34.01	0.53
70−74	−2.25	−4.90	0.46
75−79	−8.39	−13.68	0.61
80+	−7.52	−15.43	0.49
Race/ethnicity			
Non-Hispanic White	3.15	2.21	1.43
Hispanic or Other Race	−2.28	−1.32	1.73
Charlson Comorbidity Index			
0	9.66	19.78	0.49
1	−6.89	−12.30	0.56
2+	−2.77	−4.48	0.62
Tumor size (cm)			
<1	−0.67	−2.99	0.22
1 to <2	−6.1	5.05	1.21
2 to <3	4.07	−15.28	0.27
3+	2.69	13.22	0.20
Node positivity			
None or not determined	−15.22	−33.84	0.45
1−3 nodes	5.01	3.86	1.30
4+ nodes	10.21	29.98	0.34
Histologic grade			
Well differentiated	−1.62	−0.15	10.80
Intermediate/ moderate	18.96	−4.00	4.74
Poorly differentiated/ undifferentiated/ anaplastic	−6.34	20.52	0.31
Not determined/ stated	−11.01	−16.36	0.67
ER expression			
Positive	−24.3	−31.68	0.77
Negative	24.3	31.68	0.77
PR expression			
Positive	−22.16	−30.34	0.73
Negative	22.16	30.34	0.73
Primary therapy			
BCS only	−8.01	−9.84	0.81
BCS plus radiation therapy	−6.38	−2.86	2.23
Mastectomy	14.39	12.69	1.13
Tamoxifen prescribed			
Yes	−7.03	−14.03	0.57
No	7.03	14.03	0.57

Abbreviations: PR, progesterone receptor; IV, instrumental variable; ER, estrogen receptor.

[a] Prevalence of being assigned the index condition for the instrument minus the prevalence of being assigned the reference condition for the instrument.

[b] Prevalence of having the index condition (receiving adjuvant chemotherapy) minus the prevalence of not having the reference condition (not receiving adjuvant chemotherapy).

[c] Prevalence difference for the instrument divided by the prevalence difference for the exposure.

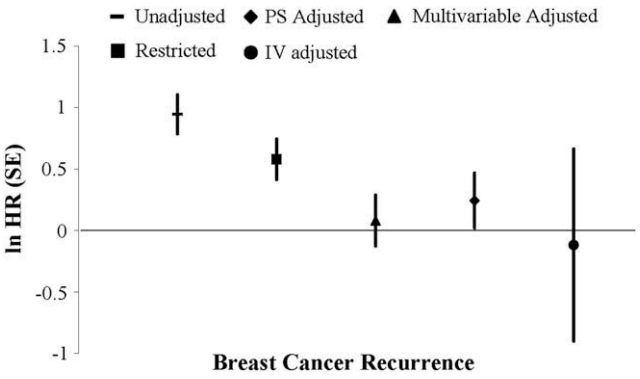

Fig. 3. Estimates and standard errors for the association between adjuvant chemotherapy and rate of breast cancer recurrence in older women.

other reports, our results that suggest PSs do not provide any better control for unmeasured confounding than multivariable regression [10,11]. Even after controlling for known prognostic factors, we obtained effect estimates in the causal direction, which would not be correct given the prior on the expected direction of effect, which is based on results from clinical trials in younger women [6]. Our results using the IV-like approach yielded a slightly protective estimate of the association; however, the imbalance of measured factors across levels of the IV indicated that our estimate remained confounded. Thus, no method of adjustment completely resolved this bias.

Selection bias and misclassification are unlikely explanations for our results. The potential for selection bias owing to barriers to care was reduced by using an unselected sample of Medicare-insured women with complete data on treatment from integrated health care systems [23]. The inter-rater reliability of medical record abstraction was ⩾90% overall [36]; with 90% sensitivity and 96% specificity for breast cancer recurrence classification and 90% sensitivity and perfect specificity for receipt of chemotherapy [36].

Another possibility may be that the protective effect of adjuvant chemotherapy seen in younger women does not apply to early-stage breast cancer in older women. The meta-analysis of 194 randomized controlled trials from 1985 to 2000 stratified by age yielded a similar finding as our IV-like result for women 70 years of age or older (recurrence rate ratio = 0.88) for 15 years of follow-up [6]. Yet, only 3.6% of the 95,403 women participating in the polychemotherapy trials were in this age category [6]. Thus, this meta-analysis finding should be interpreted with caution because geriatric women were underrepresented [3,4]. It is likely that the women who enrolled in these trials were healthier [4] than the general elder patient population living with breast cancer [37].

We explored whether the effect of chemotherapy on recurrence varied by type of regimen. When we repeated the analyses restricting the chemotherapy exposure to CMF regimen, other than less precision for the estimates, the HRs were nearly the same, except that the IV estimate

became slightly more protective. We could not perform subgroup analyses for the adriamycin-based regimen because of small numbers. However, in younger patients, for whom the data are adequate to assess the differences between adriamycin-based and nonadriamycin-based chemotherapy, the difference in recurrence between these types of chemotherapy is ~3% at 5 years after diagnosis [6].

We explored potential explanations for our PS and IV findings. Our PS quintile adjustment suggested a stronger association among women who received adjuvant chemotherapy and recurrence than the other PS methods. Subjects were not evenly distributed between the quintiles, which was because of the inability of the PS quintile adjustment to discriminate scores between subjects with the same probability of exposure. Thus, most of the subjects fell into the lowest quintile (Q1), which may explain why the continuous and doubly robust PS adjustments yielded better control.

We assessed whether our PS findings could be explained by differences in patient characteristics between the PS trimmed sample and restricted cohort by comparing multivariable regression results of the two analytic samples. We found nearly identical results, indicating that the distributions were similar. Additionally, our PS adjustment results were nearly equivalent to those yielded by the multivariable method. PSs are thought to be superior to multivariable regression models because they theoretically allow control for multiple measured confounders and increase comparability between the index and reference groups [10−12,14]. However, in a review by Sturmer et al., they found that only 13% of 69 studies had multivariable adjusted results >20% different than results from adjusting for PS [10]. Moreover, we found that even after trimming our data set to exclude nonoverlapping PSs, the distribution of the PSs among those who received chemotherapy (index) vs. those who did not (reference) still lacked comparability. This finding suggests residual confounding, which we could not examine using conventional methods. As expected, the PS method did not rectify the confounding by indication in our study; it persisted in the cohort of high-risk patients even after adjusting for measured prognostic factors that are considered when prescribing adjuvant chemotherapy.

We compared our IV-like method to Brookhart and Schneeweiss' example of a preference-based IV method to provide a better understanding of the validity of our IV result. They studied approximately 50,000 subjects [33], whereas after applying the exclusions required to implement our IV-like method, our analytic sample was 539 women. The IV acts as if we had randomized the exposure and like randomization, substantial departures in the data from the presumed balance of measured and unmeasured confounders is more likely in smaller studies.

More probable explanations may be violations of the IV assumptions, which Hernan and Robins have emphasized are unverifiable [17]. We initially questioned the strength of our IV because an IV that is weakly associated with exposure can bias the estimate more than not adjusting at all [17,18,38,39]. However, the strength of our IV was equivalent to the strength of the IV used by Brookhart and Schneeweiss (23%) [13] and similar to the strength of other preference-based IVs that they have encountered (M.A. Brookhart, unpublished data, 2008).

We then evaluated whether our IV was independent of unmeasured risk factors. We assessed the plausibility of confounding by unmeasured factors by comparing the prevalence differences of measured factors across levels of the IV. Imbalances remained among measured characteristics, suggesting that there may be clustering of patient risk factors within certain surgeons. Therefore, we cannot rule out that associations between important unmeasured factors and the IV may exist. The imbalance of measured patient characteristics across levels of the IV indicated that the IV is confounded. The IV estimate for the association between receipt of chemotherapy and breast cancer recurrence controlled for more confounding than the other methods, but did not completely resolve the bias.

The intervals around the estimates were wider using the PS method and IV method than conventional methods. The widths of the intervals (ratio of upper to lower limits) around the unrestricted, restricted, PS continuous and doubly robust, and PS quintile estimates were 1.8, 1.9, 2.4, and 2.5, respectively. The width of the interval around the IV estimate was substantially larger at 22. This demonstrates that our IV-like method was less statistically efficient than the conventional methods and, therefore, larger samples may be needed for IV methods to be feasible.

Alternative methods have been suggested to reduce confounding in observational studies, yet we found that conventional methods, such as restriction and multivariable regression, were as effective as the PS method. Our IV-like method was the only approach that yielded a protective association. However, we must be cautious in its interpretation because of the residual confounding in the distribution of measured factors across levels of the IV. The use of these alternative analytic methods to control for confounding by indication is not universal across all observational settings [10,13,16].

Nonrandomized studies of therapy effectiveness will remain important contributions to our scientific knowledge base. Such studies will, however, remain susceptible to confounding by indication, despite advancing methods to control this seemingly intractable bias. Understanding the limitations and appropriateness of the PS and IV methods is an essential step before implementing and interpreting results that may be as biased as results generated by conventional methods.

Acknowledgment

Source of financial support: Supported by Public Health Service Grant R01 CA093772 (Breast Cancer Treatment Effectiveness in Older Women, Rebecca A. Silliman, PI)

from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

The authors thank Dr. Alan Brookhart for his thoughtful review and guidance regarding the instrumental variable analysis. We also thank Dr. Terry Field, site-principal investigator at Fallon Clinic/Meyers Primary Care Institute, University of Massachusetts Medical School and Hans Petersen at Lovelace Respiratory Research Institute for Lovelace Health System for their contributions to the design and data collection aspects of the BOW cohort.

Preliminary results of this research were presented at the 40th annual meeting of the Society for Epidemiologic Research (June 2007).

References

[1] Giordano SH, Kuo Y-F, Duan Z, Hortobagyi GN, Freeman J, Goodwin JS. Limits of observational data in determining outcomes from cancer therapy. Cancer 2008;112:2456−66.

[2] Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Engl J Med 1994;331:394−8.

[3] Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: A critical comparison of trials with nonrandomized studies. Hepatology 2006;44:1075−82.

[4] Avorn J. In defense of pharmacoepidemiology - embracing the yin and yang of drug research. N Engl J Med 2007;357:2219−21.

[5] Gross CP, Steiner CA, Bass EB, Powe NR. Relation between prepublication release of clinical trial results and the practice of carotid endarterectomy. JAMA 2000;284:2886−93.

[6] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687−717.

[7] Lash TL, Silliman RA. Re: prevalence of cancer. J Natl Cancer Inst 1998;90:399−400.

[8] Geiger AM, Thwin SS, Lash TL, Buist DSM, Prout MN, Wei F, et al. Recurrences and second primary breast cancers in older women with initial early-stage disease. Cancer 2007;109:966−74.

[9] Ulcickas Yood M, Owusu C, Buist DSM, Geiger AM, Field TS, Thwin SS, et al. Mortality impact of less-than-standard therapy in older breast cancer patients. J Am Coll Surg 2008;206:66−75.

[10] Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol 2006;59:437−47.

[11] Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 2006;98:253−9.

[12] Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol 2003;158:280−7.

[13] Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat 2007;3:Article 14.

[14] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41−55.

[15] Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722−9.

[16] Terza JV, Bradford WD, Dismuke CE. The use of linear instrumental variables methods in health services research and health economics: a cautionary note. Health Serv Res 2008;43:1102−20.

[17] Hernan M, Robins JM. Instruments for causal inference. Epidemiology 2006;17:360−72.

[18] Martens E, Pestman W, de Boer A, Belitser S, Klungel O. Instrumental variables: application and limitations. Epidemiology 2006;17:260−7.

[19] Warren JL, Klabunde CN, Schrag D, Bach P. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care 2002;40(suppl):IV3−IV18.

[20] Giordano SH, Duan Z, Kuo Y-F, Hortobagyi GN, Goodwin JS. Use and outcomes of adjuvant chemotherapy in older women with breast cancer. J Clin Oncol 2006;24:2750−6.

[21] Elkin EB, Hurria A, Mitra N, Schrag D, Panageas KS. Adjuvant chemotherapy and survival in older women with hormone receptor-negativebreast cancer: assessing outcome in a population-based, observational cohort. J Clin Oncol 2006;24:2757−64.

[22] Du XL, Jones DV, Zhang D. Effectiveness of adjuvant chemotherapy for node-positive operable breast cancer in older women. J Gerontol A Biol Sci Med Sci 2005;60:1137−44.

[23] Enger SM, Thwin SS, Buist DSM, Field T, Frost F, Geiger AM, et al. Breast cancer treatment of older women in integrated health care settings. J Clin Oncol 2006;24:4377−83.

[24] Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, et al. Building a research consortium of large health systems: the cancer research network. J Natl Cancer Inst Monogr 2005;2005:3−11.

[25] Thwin S, Clough-Gorr K, McCarty M, Lash T, Alford S, Buist D, et al. Automated inter-rater reliability assessment and electronic data collection in a multi-center breast cancer study. BMC Med Res Methodol 2007;7:23.

[26] Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373−83.

[27] Glick JH, Gelber RD, Goldhirsch A, Senn H-J. Meeting highlights: adjuvant therapy for primary breast cancer. J Natl Cancer Inst 1992;84:1479−85.

[28] Sturmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. Am J Epidemiol 2005;161:891−8.

[29] Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757−63.

[30] Kahn KL, Tisnado DM, Adams JL, Liu H, Chen W-P, Hu FA, et al. Does ambulatory process of care predict health-related quality of life outcomes for patients with chronic disease? Health Serv Res 2007;42:63−83.

[31] Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: a case study on the estimation of causal effects. Am J Epidemiol 2008;167:1120−9.

[32] Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA 2007;297:278−85.

[33] Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 2006;17:268−75.

[34] Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. N Engl J Med 1998;338:1516−20.

[35] NIH consensus conference Treatment of early-stage breast cancer. JAMA 1991;265:391−5.

[36] Lash TL, Fox MP, Thwin SS, Geiger AM, Buist DSM, Wei F, et al. Using probabilistic corrections to account for abstractor

agreement in medical record reviews. Am J Epidemiol 2007;165: 1454−61.

[37] McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. BMJ 1999;319:312−5.

[38] Staiger D, Stock JH. Instrumental variables regression with weak instruments. Econometrica 1997;65:557−86.

[39] Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. J Am Stat Assoc 1995;90:443−50.