Systematic Reviews and Meta- and Pooled Analyses

Bias in Observational Studies of Prevalent Users: Lessons for Comparative Effectiveness Research From a Meta-Analysis of Statins

Goodarz Danaei*, Mohammad Tavakkoli, and Miguel A. Hernán

* Correspondence to Dr. Goodarz Danaei, Department of Global Health and Population, Harvard School of Public Health, 665 Huntington Avenue, Bldg. 1, Room 1107, Boston MA 02115 (e-mail: gdanaei@hsph.harvard.edu).

Initially submitted March 7, 2011; accepted for publication August 3, 2011.

Randomized clinical trials (RCTs) are usually the preferred strategy with which to generate evidence of comparative effectiveness, but conducting an RCT is not always feasible. Though observational studies and RCTs often provide comparable estimates, the questioning of observational analyses has recently intensified because of randomized-observational discrepancies regarding the effect of postmenopausal hormone replacement therapy on coronary heart disease. Reanalyses of observational data that excluded prevalent users of hormone replacement therapy led to attenuated discrepancies, which begs the question of whether exclusion of prevalent users should be generally recommended. In the current study, the authors evaluated the effect of excluding prevalent users of statins in a meta-analysis of observational studies of persons with cardiovascular disease. The pooled, multivariate-adjusted mortality hazard ratio for statin use was 0.77 (95% confidence interval (CI): 0.65, 0.91) in 4 studies that compared incident users with nonusers, 0.70 (95% CI: 0.64, 0.78) in 13 studies that compared a combination of prevalent and incident users with nonusers, and 0.54 (95% CI: 0.45, 0.66) in 13 studies that compared prevalent users with nonusers. The corresponding hazard ratio from 18 RCTs was 0.84 (95% CI: 0.77, 0.91). It appears that the greater the proportion of prevalent statin users in observational studies, the larger the discrepancy between observational and randomized estimates.

bias (epidemiology); comparative effectiveness research; confounding factors (epidemiology); meta-analysis; prospective studies; selection bias

Abbreviations: CHD, coronary heart disease; CI, confidence interval; HR, hazard ratio; HRT, hormone replacement therapy.

Reliable evidence on the effectiveness and safety of clinical and public health interventions is central to the ongoing discussion of health care in the United States (1–6) and other countries (7, 8). Although randomized clinical trials are usually the preferred strategy for obtaining such evidence, they are not always feasible or timely. Clearly, much evidence on the comparative effectiveness and safety of clinical and public health interventions will have to be derived from observational studies.

Observational studies often yield effect estimates comparable to those of randomized trials (9–12), but the ability of observational analyses to provide valid effect estimates has been questioned in some high-profile cases. Prominent among these is the effect of postmenopausal hormone replacement therapy (HRT) on coronary heart disease (CHD). Although some observational studies have suggested harmful effects of HRT (13), most observational studies (14, 15) have found a lower CHD risk in prevalent users of HRT compared with nonusers. This finding was interpreted as supporting the existence of a protective effect of HRT on CHD risk. However, in a large randomized clinical trial, Manson et al. (16) found a higher CHD risk in incident users of HRT compared with nonusers, especially in the early period of follow-up. Reanalyses of the observational studies (17, 18) that, like the randomized trial, compared incident users with nonusers after applying a washout period found no overall beneficial effect of HRT on CHD risk.

The above findings support previous proposals to exclude prevalent users when using observational data to assess the comparative effectiveness and safety of clinical and public health interventions.
health interventions (19). Here we provide further support for eliminating prevalent users in observational studies on the effectiveness of statin therapy (3-hydroxy-3-methylglutaryl coenzyme-A reductase inhibitors) in reducing CHD risk and mortality. In most of the published observational studies, investigators have compared prevalent users with nonusers.

We chose statins as a case study because the beneficial effect of statins on CHD risk and mortality has been proven beyond any doubt in more than 3 dozen randomized clinical trials and several large meta-analyses (20–32), but no systematic review and meta-analysis of observational studies of statin therapy is available. We compared estimates from randomized and observational studies according to the type of intervention (primary prevention vs. secondary prevention) and classified observational studies according to whether they included prevalent users, incident users, or both. Our goal was to draw some general conclusions to improve the analyses of observational studies for comparative effectiveness research.

MATERIALS AND METHODS

Search strategy

We searched PubMed using “hydroxymethylglutaryl-CoA reductase inhibitors” and “myocardial ischemia” or “mortality” as Medical Subject Headings. We searched separately for “epidemiologic studies” and “randomized controlled trials” and identified studies that had been published before November 2010. We limited our search to studies conducted in humans and in adults over 18 years of age (≥19 years). We imposed no restrictions on the language of the publications. We also searched Embase using the same search terms and limitations for observational studies and for randomized trials published after August 2010, which was the end-of-search date in a recent meta-analysis of randomized trials that included Embase (27). Two authors (G. D. and M. T.) screened the articles using the title and abstract and subsequently retrieved and reviewed the whole article. Any discrepancies between the two screeners were resolved by coming to consensus on whether to include or exclude the abstract or article. If a consensus could not be reached, the situation was discussed with a third author (M. A. H.). We screened the bibliographies of the selected articles to find other relevant studies.

We excluded articles comparing different statins or studying the dose-response of a single statin; randomized trials that used other interventions (such as percutaneous coronary interventions) as the control group; trials that used cerivastatin; case-control studies not nested in a prospective cohort study; studies on short-term effects of statins (periprocedural, in-hospital effects, or follow-up periods of ≤6 months); studies where the investigators had not reported clinical endpoints or had reported only cerebrovascular events; studies on patients with defibrillators, heart failure, heart transplants, familial hypercholesterolemia, or chronic kidney disease; studies that did not have 1 arm for treatment with statin only; and studies with extended follow-up or post-hoc and subgroup analyses of previously published randomized trials.

Data extraction

We divided the selected studies into studies of primary prevention and secondary prevention. When a study included subjects both with and without previous cardiovascular events, we classified it as a secondary prevention study when more than 30% of participants had previous cardiovascular events at baseline. The outcome of interest for primary prevention studies was CHD, defined as nonfatal myocardial infarction or death from CHD. For secondary prevention studies, we chose all-cause mortality as the outcome because few observational studies reported information on recurrent myocardial infarction or CHD.

We extracted data on characteristics of the study population, including age, proportion of male participants, sample size, eligibility criteria, intervention or comparison groups, follow-up duration, compliance with or adherence to treatment, outcome definition and variables, and methods used to adjust for confounding. For randomized trials, we also extracted the point estimate for the intention-to-treat hazard ratio comparing treated persons with controls (and its confidence interval) for primary and secondary endpoints. If investigators had only reported results for several subgroups separately (e.g., men and women), we extracted the effect estimates for each subgroup. We excluded studies for which the article did not include enough information to calculate incidence rates and their confidence intervals.

For observational studies, we extracted the crude and adjusted hazard ratio estimates (and their 95% confidence intervals) comparing users with nonusers for the primary and secondary outcomes considered in each study. Statin users were defined as prevalent users if they had initiated statin therapy prior to their inclusion in the study and incident users if they had initiated statin therapy at or after their inclusion in the study. We classified observational studies into 3 categories based on these definitions: 1) studies that compared prevalent (current) users with nonusers, 2) studies that compared incident users with nonusers, and 3) studies that compared a combination of prevalent and incident users with nonusers. The latter group mostly included hospital-based studies that had assigned an initiation time (usually at discharge from the hospital) but did not exclude patients who were taking statins prior to admission. We excluded studies that compared persistent (long-term prevalent) users with nonpersistent users because the definitions of persistent or long-term use were not consistent across these studies.

Statistical analysis

We pooled the reported hazard ratio estimates and computed their 95% confidence intervals separately for studies of primary and secondary prevention and also separately for randomized trials and observational studies. Pooled estimates were similar regardless of whether we used a fixed-effects model or a random-effects model (33). We used results from the latter in all meta-analyses. We evaluated the role of outliers among secondary prevention studies by dropping the two studies with the smallest and largest effect estimates and pooling the other studies. This analysis showed that our results were robust to the presence of outliers (results not shown). We
Figure 1. Processing of A) randomized trials and B) observational studies in a review and meta-analysis of observational studies of statin therapy. One randomized trial and 2 observational studies included comparisons of both primary and secondary prevention. (CHD, coronary heart disease).
used a funnel plot and a regression asymmetry test (34) to assess small-study bias. When the estimates differed substantially among the pooled studies, we used meta-regression (35) to explore the following potential predictors of heterogeneity: duration of follow-up (mean or median), mean age of participants at baseline, proportion of male participants, context (hospital-based vs. community- or population-based), location (United States or elsewhere), and adjustment methods (multivariate outcome models vs. propensity score methods). All analyses were conducted using Stata, version 10.1 (StataCorp LP, College Station, Texas).

RESULTS

Our search identified 425 randomized clinical trials and 778 observational studies, of which we selected 34 randomized trials and 84 observational studies after reviewing their titles and abstracts. A bibliographic search on the selected articles identified 21 additional randomized trials and 3 additional observational studies that matched our inclusion criteria. One observational study did not provide enough information to estimate the variance of the effect estimate and was excluded (36). We extracted information for at least 1 disease outcome from 40 randomized trials and 41 observational studies. See Figure 1 for a flowchart of studies processed in this review and Web Tables 1 and 2 (which appear on the Journal’s Web site (http://aje.oxfordjournals.org/)) for characteristics of the studies that contributed to one or more of the pooled estimates.

Secondary prevention

Eighteen randomized trials (37–54) estimated the mortality hazard ratio for statin initiation versus no initiation, either after a cardiovascular event or in a population with a high prevalence of cardiovascular disease. The pooled mortality hazard ratio was 0.84 (95% confidence interval (CI): 0.77, 0.91) (Figure 2). The pooled hazard ratio for recurrent CHD in 17 randomized trials of secondary prevention (37, 40–46, 50–52, 54–58) was 0.75 (95% CI: 0.70, 0.80) (Web Figure 1).

The pooled hazard ratio for all-cause mortality in observational studies comparing prevalent users with nonusers (59–71) was 0.44 (95% CI: 0.27, 0.72) before adjustment for potential confounders and 0.54 (95% CI: 0.45, 0.66) after adjustment (Figure 3). Not all observational studies reported the unadjusted hazard ratios, but when we restricted the analysis to those which did, our findings were not materially

Figure 2. Hazard ratio (HR) for mortality (squares) according to initiation of statin use in secondary prevention randomized clinical trials and pooled HR for treatment versus control status (diamond). Bars, 95% confidence interval (CI). (GISSI, Gruppo Italiano per lo Studio della Sopravvenza nell’Infarto Miocardico; HPS, Heart Protection Study; LIPID, Long-term Intervention with Pravastatin in Ischaemic Disease; MAAS, Multicentre Anti-Atheroma Study; 4S, Scandinavian Simvastatin Survival Study).
affected (Web Figure 2). Studies that compared a combination of prevalent and incident users with nonusers (72–84) had a pooled hazard ratio for all-cause mortality of 0.47 (95% CI: 0.32, 0.69) before adjustment for confounders and 0.70 (95% CI: 0.64, 0.78) after adjustment (Figure 4). Only 4 studies (59, 85–87) compared incident users with nonusers, and the pooled hazard ratio from these studies was 0.77 (95% CI: 0.65, 0.91) after adjustment for confounders (Figure 5).

Figure 3. Hazard ratio (HR) for mortality (squares) according to statin use in secondary-prevention observational studies comparing prevalent users with nonusers and pooled HR for users versus nonusers (diamond). A) unadjusted results; B) adjusted results. Bars, 95% confidence interval (CI).
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Favors Statins</th>
<th>Favors Control</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen-Maycock, 2002 (72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt;65 years</td>
<td></td>
<td></td>
<td>0.52 (0.35, 0.79)</td>
</tr>
<tr>
<td>Ages 65−79 years</td>
<td></td>
<td></td>
<td>0.53 (0.38, 0.73)</td>
</tr>
<tr>
<td>Age ≥80 years</td>
<td></td>
<td></td>
<td>0.49 (0.26, 0.93)</td>
</tr>
<tr>
<td>De Luca, 2006 (75)</td>
<td></td>
<td></td>
<td>0.16 (0.09, 0.32)</td>
</tr>
<tr>
<td>Leeper, 2007 (79)</td>
<td></td>
<td></td>
<td>0.81 (0.68, 0.96)</td>
</tr>
<tr>
<td>Muhlestein, 2004 (80)</td>
<td></td>
<td></td>
<td>0.44 (0.29, 0.65)</td>
</tr>
<tr>
<td>Overall (I² = 84.2%, P = 0.000)</td>
<td></td>
<td></td>
<td>0.47 (0.32, 0.69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Favors Statins</th>
<th>Favors Control</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen-Maycock, 2002 (72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt;65 years</td>
<td></td>
<td></td>
<td>0.70 (0.46, 1.07)</td>
</tr>
<tr>
<td>Ages 65−79 years</td>
<td></td>
<td></td>
<td>0.56 (0.41, 0.77)</td>
</tr>
<tr>
<td>Age ≥80 years</td>
<td></td>
<td></td>
<td>0.50 (0.26, 0.96)</td>
</tr>
<tr>
<td>Carrier, 2009 (73) (bilateral ITA)</td>
<td></td>
<td></td>
<td>1.05 (0.73, 1.51)</td>
</tr>
<tr>
<td>Carrier, 2009 (73) (unilateral ITA)</td>
<td></td>
<td></td>
<td>0.74 (0.64, 0.85)</td>
</tr>
<tr>
<td>Cooke, 2009 (74)</td>
<td></td>
<td></td>
<td>0.74 (0.63, 0.88)</td>
</tr>
<tr>
<td>De Luca, 2006 (75)</td>
<td></td>
<td></td>
<td>0.24 (0.12, 0.47)</td>
</tr>
<tr>
<td>Fintel, 2007 (76)</td>
<td></td>
<td></td>
<td>0.68 (0.53, 0.87)</td>
</tr>
<tr>
<td>Furukawa, 2008 (77)</td>
<td></td>
<td></td>
<td>0.70 (0.58, 0.85)</td>
</tr>
<tr>
<td>Kulik, 2008 (78)</td>
<td></td>
<td></td>
<td>0.85 (0.72, 0.94)</td>
</tr>
<tr>
<td>Leeper, 2007 (79)</td>
<td></td>
<td></td>
<td>0.65 (0.53, 0.80)</td>
</tr>
<tr>
<td>Muhlestein, 2004 (80)</td>
<td></td>
<td></td>
<td>0.49 (0.33, 0.74)</td>
</tr>
<tr>
<td>Nagashima, 2007 (81)</td>
<td></td>
<td></td>
<td>0.64 (0.45, 0.90)</td>
</tr>
<tr>
<td>Sheng, 2009 (82)</td>
<td></td>
<td></td>
<td>1.06 (0.76, 1.47)</td>
</tr>
<tr>
<td>Stenestrand, 2001 (83)</td>
<td></td>
<td></td>
<td>0.75 (0.63, 0.89)</td>
</tr>
<tr>
<td>Zhang, 2009 (84)</td>
<td></td>
<td></td>
<td>0.58 (0.36, 0.93)</td>
</tr>
<tr>
<td>Overall (I² = 57.5%, P = 0.002)</td>
<td></td>
<td></td>
<td>0.70 (0.64, 0.78)</td>
</tr>
</tbody>
</table>

**Figure 4.** Hazard ratio (HR) for mortality (squares) according to statin use in secondary-prevention observational studies comparing a combination of prevalent and incident users with nonusers and pooled HR for users versus nonusers (diamond). A) unadjusted results; B) adjusted results. Bars, 95% confidence interval (CI). (ITA, internal thoracic artery graft).
There was substantial heterogeneity among these groups of studies, as evidenced by large $I^2$ values (91% for studies of prevalent users, 58% for studies of prevalent and incident users combined, and 84% for studies of incident users). The meta-regression analyses did not find any strong predictors of heterogeneity in studies of prevalent users. In studies that compared a combination of prevalent and incident users with nonusers, the log hazard ratio increased by 0.09 (95% CI: 0.02, 0.16) per additional year of mean/median follow-up and decreased by 0.038 (95% CI: 0.018, 0.074) per 10% increase in proportion of male participants at baseline.

**Primary prevention**

Eleven randomized trials of primary prevention of CHD (88–98) estimated the hazard ratio for statin initiation versus no initiation. The pooled hazard ratio was 0.69 (95% CI: 0.60, 0.79) (Web Figure 3). There were only 2 observational studies (99, 100) comparing incident users with nonusers, and the pooled adjusted hazard ratio for CHD from these studies was 0.80 (95% CI: 0.63, 1.02). Of the 4 studies that compared prevalent users with nonusers, 2 studies reported only the hazard ratio for all-cause mortality (79) or cardiovascular mortality (101), and 2 studies reported hazard ratios for myocardial infarction, with considerable heterogeneity (the adjusted hazard ratio was 0.35 in one study (102) and 1.41 in the other one (99)); thus, we did not pool data from those 2 studies.

**Small-study bias**

The funnel plots for randomized trials and observational studies (Web Figure 4) showed little evidence of small-study bias. The $P$ values from the asymmetry test were 0.11 and 0.24 for primary- and secondary-prevention randomized trials, respectively; 0.30 for secondary-prevention observational studies of prevalent users; 0.15 for studies that combined prevalent and incident users; and 0.46 for studies of incident users. Furthermore, exclusion of angiographic trials (4 on primary prevention and 6 on secondary prevention) did not change the results of our meta-analyses (results not shown).

**Adherence to treatment**

Figure 6 presents findings for treatment discontinuation among statin initiators/users in primary and secondary prevention studies by year of follow-up and type of study. In primary prevention studies, on average, 21% of statin initiators discontinued treatment in randomized trials versus 41% in observational studies. In secondary prevention studies, the corresponding proportions were 15% in randomized trials and 32% in observational studies. In only 2 observational studies did investigators report the proportion of nonusers who started treatment during follow-up, so we could not make any meaningful comparison of imperfect adherence among nonusers.

**DISCUSSION**

According to our meta-analysis of randomized trials of statins for secondary prevention, the mortality hazard ratio for initiation of statin therapy versus no initiation was 0.84 (95% CI: 0.77, 0.91). That is, statin therapy reduces mortality by 16% in patients with a history of cardiovascular disease. These results are generally consistent with prior meta-analyses of randomized trials (20–31). In contrast, our meta-analysis of observational studies of prevalent statin users found a hazard ratio of 0.54 (95% CI: 0.45, 0.66).
The difference between the pooled hazard ratio for randomized trials and that for observational studies with prevalent users is unlikely to be due to chance (because their 95% confidence intervals do not overlap), use of different drugs across studies (because commonly used statins have similar effects (25)), differences in adherence to treatment (because lower adherence in observational studies should have resulted in a bias towards the null), or different distributions of effect

*Am J Epidemiol.* 2012;175(4):250–262
modifiers (because subgroup analyses of many randomized trials have rarely found differences (23, 25, 103)). On the other hand, the randomized-observational discrepancy may be due to selection bias or residual confounding, as discussed below.

A comparison of prevalent users of statins with nonusers is subject to selection bias (18, 104) because prevalent users have by definition survived under treatment. If treatment decreases the risk of the outcome, the group of prevalent users will be progressively enriched with susceptible patients as compared with nonusers or never users (and conversely more resilient patients if treatment increases the risk of the outcome). In addition, including prevalent users in the analysis often implies that the confounders are measured after treatment initiation. If confounders are affected by prior treatment, often implies that the confounders are measured after treatment (see Figure 6) or by increased nonadherence over time (see Figure 6) or by increased selection bias due to inclusion of prevalent users.

To eliminate these selection biases, one can restrict the analysis to patients who have not used the drug for some period of time before the start of follow-up. Then the analysis would compare incident users (or initiators) with nonusers or never users (and conversely more resilient patients if treatment increases the risk of the outcome). In addition, including prevalent users in the analysis often implies that the confounders are measured after treatment initiation. If confounders are affected by prior treatment, adjustment for confounding will introduce selection bias (105).

Residual confounding is always a potential source of discrepancy between randomized trials and observational studies. In the secondary-prevention observational studies included in our analysis, statin users were consistently younger than nonusers (64, 65, 71, 106), had undergone more revascularizations (62, 70, 71, 106), and had more use of antihypertensive and antithrombotic drugs (62, 64, 65, 71), which implies better access to health care and possibly a better prognosis. As a result, statin use may be a marker for healthy status and/or high-quality medical care in secondary-prevention observational studies. This confounding might partly explain why the pooled crude hazard ratios (HR = 0.44 in studies with prevalent users only; HR = 0.47 in studies with a combination of prevalent and incident users) were slightly smaller than the fully adjusted ones (HR = 0.54 and HR = 0.70, respectively) (Figures 3 and 4).

Our main findings are focused on secondary prevention studies, and we had to focus on all-cause mortality because very few studies reported effects on recurrent CHD. Unfortunately, the number of primary prevention observational studies was too small for a meaningful analysis. The pooled hazard ratio from 2 observational studies of incident users suggested a smaller protective effect than in randomized trials (HR = 0.80 vs. HR = 0.69), possibly because of residual confounding by indication. A comparison of the crude and adjusted estimates from these 2 studies (HR = 1.67 vs. HR = 0.80) showed that the participants with a higher risk of the disease were more likely to receive treatment in primary prevention settings, which suggests that the pattern of confounding varies between the primary and secondary prevention settings.

Our study had several limitations. First, few observational studies had compared incident users of statin therapy with nonusers, which limited the precision of the estimates for these studies. Second, our search was limited to PubMed and Embase. To minimize the inadvertent exclusion of small studies, we conducted a comprehensive bibliographic search and reviewed previously published meta-analyses of randomized trials. Our funnel plots for secondary-prevention observational studies did not show any evidence of small-study bias. Third, the primary-prevention randomized trials used slightly different definitions of CHD, which may have increased the heterogeneity of the hazard ratios.

Mean follow-up time and proportion of male participants were the main determinants of heterogeneity in secondary-prevention observational studies that compared a combination of prevalent and incident users with nonusers. A larger protective effect for men was also observed in a recent meta-analysis of randomized trials (32). The smaller hazard ratio estimates with longer follow-up times may be partly explained by increased nonadherence over time (see Figure 6) or by increased selection bias due to inclusion of prevalent users.

In summary, our findings support the hypothesis that the greater the proportion of prevalent statin users in observational studies, the larger the discrepancy between observational and randomized estimates. In future observational studies of comparative effectiveness, investigators may reduce the potential for bias by attempting to emulate the design and analysis of a hypothetical trial by imposing the same eligibility criteria and comparing incident users of treatment with nonusers (99).

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Miguel A. Hernán, Goodarz Danaei); Department of Global Health and Population, Harvard School of Public Health, Boston, Massachusetts (Goodarz Danaei, Mohammad Tavakkoli); and Harvard-MIT Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology, Boston, Massachusetts (Miguel A. Hernán).

This work was funded by National Institutes of Health grant R01 HL080644 to Dr. Miguel A. Hernán.

Conflict of interest: none declared.

REFERENCES


Am J Epidemiol. 2012;175(4):250–262