Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison With Randomized Trial Results

Sebastian Schneeweiss, MD, ScD[*], **Amanda R. Patrick, MS**[*], **Til Stürmer, MD, MPH**[*], **M. Alan Brookhart, PhD**[*], **Jerry Avorn, MD**[*], **Malcolm Maclure, ScD**[*], **Kenneth J. Rothman, DMD, DrPH**[†], and **Robert J. Glynn, PhD, ScD**[*]

[*]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

[†]RTI-Health Solutions, Research Triangle Park, North Carolina

Abstract

Background—The goal of restricting study populations is to make patients more homogeneous regarding potential confounding factors and treatment effects and thereby achieve less biased effect estimates.

Objectives—This article describes increasing levels of restrictions for use in pharmacoepidemiology and examines to what extent they change rate ratio estimates and reduce bias in a study of statin treatment and 1-year mortality.

Methods—The study cohort was drawn from a population of seniors age 65 years and older enrolled in both Medicare and the Pennsylvania Pharmaceutical Assistance Contract for the Elderly (PACE) between 1995 and 2002. We identified all users of statins during the study period and assessed the time until death within 1 year. The following progressive restrictions were applied: (1) study incident drug users only, (2) choose a comparison group most similar to the intervention group, (3) exclude patients with contraindications, (4) exclude patients with low adherence, and (5) restrict to specific high-risk/low-risk subgroups represented in randomized trails (RCTs).

Results—The basic cohort comprised 122,406 statin users, who were on average 78 years old and predominantly white (93%) and showed an unadjusted rate ratio of 0.32 for statin users. When all 5 restrictions were applied (N = 11,673), the unadjusted rate ratio had increased to 0.72. Multivariable Cox regression adjusted rate ratios increased from 0.62 [95% confidence interval (CI), 0.58–0.66] to 0.79 (95% CI, 0.60–1.03). However, after the first 3 restrictions the effect size changed little. The final estimate is similar to that obtained as a pooled estimate of 3 pravastatin RCTs in patients age 65 years and older. We argue that restrictions 1 through 4 compromised generalizability little.

Conclusions—In our example of a large database study, restricting to incident drug users, similar comparison groups, patients without contraindication, and to adherent patients was a practical strategy, which limited the effect of confounding, as these approaches yield results closer to those seen in RCTs.

Keywords

pharmacoepidemiology; confounding; restriction; methods; statins

---

Results from pharmacoepidemiologic research often have immediate and far-reaching clinical, regulatory, and economic implications. Consequently, practitioners and policymakers must consider carefully whether any association between use of a prescription drug and health outcomes is causal. Although a variety of systematic errors may bias nonexperimental research,[1] confounding bias is of particular concern in epidemiologic studies of drug effects.[2]

Large health care utilization data sets are an efficient data source to analyze the relation between prescription drug use patterns in clinical practice and unintended and infrequent health outcomes.[3,4] Despite their advantages, pharmacoepidemiologic claims data studies have been criticized for the incompleteness of their information on potential confounders. Restricting study cohorts to patients who are homogeneous regarding their indication for the study drug will lead to more balance of patient predictors of the study outcome among exposure groups and thus will reduce confounding. Restricting study cohorts can also increase the likelihood that all included subjects will have a similar response to therapy, and therefore reduce the likelihood of effect modification. Randomized controlled trials (RCTs) commonly restrict their study population to patients with a presumed indication for the study drug and then randomly allocate the actual treatment. One might argue, therefore, that similar restriction of the study cohort in nonrandomized research can foster convergence between those findings and results from RCTs. Although restriction is based on measurable patient characteristics, selecting proxy characteristics may overcome the often-limited depth of information in health care utilization databases. The large size of such databases, however, makes it feasible to restrict on many criteria without reducing the study to a size that would meaningfully compromise the precision of effect estimates. A recent study suggested that the way investigators apply restriction varies substantially when they examine adverse drug effects in nonrandomized research, but it did not analyze the effects of different restriction strategies on the validity of findings.[5]

The objective of this article is to develop a structured approach to restricting patient populations in epidemiologic database studies of intended and unintended treatment effects. To illustrate how increasing restriction criteria affects the strength of an association between statin use and 1-year mortality, we used a cohort study of Medicare enrollees and compared results with those of randomized controlled trials. We used this example because statins are widely used in primary and secondary prevention of coronary heart disease and myocardial infarction (MI). It is a medication class for chronic use, and several randomized controlled trials are available in elderly patients showing a reduced risk of MI and suggesting improved mortality.

METHODS

Framework of 5 Decision Points

To illustrate the effects of restriction, we propose a sequence of 5 steps whereby investigators increasingly restrict their base populations to achieve a more homogeneous study population. Each step is intended to improve the balance of patient characteristics among exposure groups and thus reduces confounding. We start with describing the source population of our cohort study. The explanation and rational of each of the suggested restrictions will be followed by the specific implementation in our example study (Fig. 1).

Data Source of Empirical Analysis

The study cohort was drawn from a population of seniors age 65 years and older enrolled in both Medicare and the Pennsylvania Pharmaceutical Assistance Contract for the Elderly

(PACE) programs between 1995 and 2002. PACE is a state pharmaceutical benefits program for individuals with incomes below $14,000 and couples with incomes below $17,200; its data have been frequently used for pharmacoepidemiologic studies.[6,7] Drug dispensings required a nominal copayment of $6 and can be up to a 30-day supply. To ensure that drug and health care system use were correctly ascertained during the year before cohort entrance, we required that subjects have at least 1 prescription claim and 1 physician or hospital claim during each half of the year, indicating both Medicare and PACE enrollment. Specific cohorts drawn from this population are described below.

Study Exposure and Outcome

The initial exposure status of statin use, nonuse, or comparator drug use as determined from pharmacy claims was carried forward until censoring after 1 year or death, whichever came first. We analyzed the extent to which patients classified as nonusers started statins during follow-up and how many statin users discontinued, using a gap of 90 or more days without statin use in addition to the dispensed supply as the definition for statin discontinuation.

We used Medicare claims data to ascertain time of death. Death information from Medicare records is routinely cross-checked with Social Security data. Subjects were censored upon death or at the end of 365 days after drug initiation.

Cohort 0—Statin Users and Nonusers Matched on Calendar Time—We identified subjects who used a statin at any point between 1995 and 2002, assigning the date of first observed statin use (occurring after 1 year of eligibility) as an index date. On each statin user's index date, we sampled a subject who had not used a statin as of that date (ie, a nonuser) and assigned him or her the same index date.

First Decision Point: Study Incident Medication Users Only?

The population of statin users described above consists of a mix of incident drug users (ie, those starting on a statin) and prevalent users (ie, those taking a statin for some time). This design begins by identifying all patients in a defined population who were treated with the study medication at least once during a defined study period. Start of exposed person time begins at the first recorded dispensing of the study drug in the study period.

Mixed Prevalent and Incident User Cohorts

1. Studying mixed prevalent and incident user cohorts will lead to under-ascertainment of early events. Depending on the average duration (chronicity) of use, such cohorts may be composed predominantly of prevalent users and include few new users. The estimated average treatment effect will therefore underemphasize effects related to drug initiation and will overemphasize effects of long-term use.[8]

2. Prevalent users of a drug have by definition persisted in their drug use, similar to the concept of survivor cohorts in chronic disease epidemiology.[9] Being persistent or adherent is a characteristic found more frequently in patients who tolerate the drug well and who perceive some therapeutic benefit. Adherence is also associated with higher educational status and health seeking behavior, particularly if the study drug is treating an asymptomatic condition (eg, like statins treating hyperlipidemia) and because these characteristics are difficult to assess in claims data, they may lead to healthy user bias.[10–12]

3. The duration of use among prevalent users can differ by drug exposure; duration thus may cause bias if it remains unadjusted. Such a scenario is likely when newly marketed drugs are compared with competitors that have been available longer. In

database studies, duration of prior use can only be assessed by tracing back a continuous string of prescriptions to the initial prescription.

4. In studying prevalent users, investigators can assess patient characteristics only after the initial exposure; thus the drug under study may affect those characteristics. Adjusting for such factors that are on the causal pathway of the drug's action will lead to an underestimation of the drug effects.

"New User Design"—One begins an incident user design by identifying all patients in a defined population who start a course of treatment with the study medication. Exposed person-time begins at the start of treatment, which is identified as a dispensing of the index drug without a dispensing of that drug during the prior year or some other fixed time interval comparable with a wash-out period commonly used in RCTs. The advantage of the so-called "New User Design" has recently been summarized.[8] Although limiting the study population to drug initiators resembles one of several key characteristics of clinical trials, the limited number of incident users requires large source populations like health care utilization databases from which new starters can be identified efficiently. For some patients, it may not be the first time they take the study drug (ie, they are not really naive to the drug). Patients who know from earlier treatment courses that they tolerate and benefit from the drug are more likely to use the same drug again. The chance of an initiator being a true new user can be increased by requiring longer periods without use of the study drug before the index prescription.

Cohort 1—Incident Statin Users and Nonusers—Cohort 1a—Incident statin users and nonusers matched on calendar time. We restricted cohort 1 to incident statin users, defined as no statin use in the 12 months before the index date, and to their matched nonusers (exact date).

Cohort 1b—Incident statin users and nonusers matched on the date of another prescription or office visit. We matched the incident statin users from cohort 1a to nonusers who filled a prescription or had a physician visit on the index day (±30 days).

Second Decision Point: What Is the Most Adequate Comparison Group?

Choosing a comparison group is a complex and sometimes subjective issue. The ideal comparison group should comprise patients with identical distributions of measured and unmeasured risk factors of the study outcome.

Patients With the Same Treatment Indication: "Alternative Drug Users"—Selecting comparison drugs that have the same perceived medical indication for head-to-head comparisons of active drugs will reduce confounding by selecting patients with the same indication (eg, indication for using celecoxib vs. rofecoxib). Although one can rarely measure the indication directly—in the statin example we would need laboratory values of serum lipid levels that are not available in claims data—we infer the indication by the initiation of a treatment specific to the indication. However, new competitors within a class are often marketed for better efficacy, slightly expanded indications, or better safety [cyclo-oxygenase-2 inhibitors (coxibs) vs. nonselective nonsteroidal antiinflammatory drugs (NSAIDs)], influencing physicians' prescribing decisions.[13] In this way, new opportunities of confounding by indication can arise.

"Nonusers"—In some cases, there either is no comparator drug with a reasonably close indication to the study drug or a class effect is suspected such that the entire class is to be tested, requiring comparison subjects who did not use any drug of this class. The most obvious choice may be to identify study subjects who do not use the study drug and then to pick a random date as the index date, possibly matched by time to the index date of the first prescription among active drug users.

Obviously, patients on therapy most likely have a medical indication; by contrast a large proportion of nonusers have no medical indication. Patients initiating statin therapy are more likely to have elevated lipid levels, and therefore increased cardiac risks; nonusers as defined above may differ substantially from users of the index drug for both measured and unmeasured characteristics, even beyond the indication for the index drug.

As a case in point: although drug initiators have (presumably) been evaluated by a physician just before receiving that prescription, nonusers may not have seen a physician recently and, in fact, may have less contact with the health care system in general. Differential under-recording of health conditions in the nonuser comparison group makes members of the comparison group seem healthier than they really are and may lead to an overestimation of treatment effects.

Groups will be more comparable regarding access to health care, including health seeking behavior and disease surveillance, when choosing comparison patients who also had contact with the health care system in the form of a drug dispensing. Like patients starting the study drug, such patients have just been evaluated by a physician before the initial prescription. Adequate comparison groups for new statin initiators could, for example, be initiators of topical glaucoma drugs or thyroid hormone substitution. Both of these classes of pharmaceutics are unrelated to lowering serum lipid levels and are used for preventing the progression of an initially asymptomatic condition.

Cohort 2—Incident Statin Users and Incident Glaucoma Medication Users—We replaced the nonuser comparator group for cohort 1 with a specific active comparator group comprising subjects who initiated glaucoma medications (±30 days), with no use of a glaucoma medication or a statin in the preceding 12 months. We restricted our population of statin initiators to those who had not used a glaucoma medication in the past year. Glaucoma medication initiators were selected as a referent group in the interest of finding a group of patients who were similar to the exposed patients in having initiated a preventive therapy. Each patient's initiation date was used as the baseline for follow-up.

Third Decision Point: Excluding Patients With Contraindications?

In studies of the effectiveness of drugs, it is questionable whether to include patients who have a clear contraindication to the study drug. Such patients will be few and their experience will be unusual. Prudence dictates, therefore, excluding patients with contraindications or absolute indications, resulting in a situation similar to the therapeutic equipoise required for RCTs.[14]

Because reliably identifying contraindications in claims data is problematic, it is more promising to estimate appropriateness of treatment with an empirically fitted scoring algorithm. Propensity scores estimate each patient's probability of treatment given all measured covariates. Low propensity scores indicate low probability of treatment. Those with low scores will tend to be those who were not treated, and at the extreme end of the distribution there may be a range that is only populated by actual nonusers, because all users have higher propensity scores. Such nonusers are likely to have a contraindication for the study medication because no subject with such a low propensity score has actually received treatment. These patients should be deleted from the study population. Analogously, such trimming can be considered at the upper end of the propensity score, excluding patients who will always be treated.

Cohort 3—Incident Statin and Glaucoma Medication Users, Trimmed for Propensity Score Nonoverlap—We estimated exposure propensity scores for each subject by fitting a logistic model to predict exposure, in this case statin initiation, as a function of baseline covariates. We used the first percentile of the propensity score distribution in the exposed as a lower cut-point and the 99th percentile of the distribution of propensity scores in

the unexposed as an upper cut-point, restricting our analysis to subjects with propensity scores falling between these 2 numbers.

Fourth Decision Point: Excluding Patients With Very Low Adherence?

Patients dropping out of RCTs for reasons related to the study drug may cause bias. Noninformative drop-out causes bias towards the null in intention-to-treat (ITT) analyses. The medical profession and regulatory agencies accept such a bias because its direction is known and trial results are considered conservative regarding the drug's effectiveness. Discontinuation of treatment may also be associated with study outcomes. Obvious reasons are lack of perceived treatment effect or intolerance. Both factors may lead to early stopping but can cause discontinuation at any time later during the course of treatment. Another factor that may lead to discontinuation of medications, particularly those used to treat asymptomatic conditions, is overall frail health status that requires multiple medications to treat the more symptomatic conditions. For example, cancer patients may discontinue statins to reduce polypharmacy in favor of more urgently needed drugs.[11]

RCTs try to minimize bias from nonadherence by frequently reminding patients and by run-in phases before randomization aimed to identify and exclude nonadherent patients. In routine care, adherence to drugs is unfortunately substantially lower than in RCTs. Studies have shown that, for statin medications, only 50–60% of elderly patients refill their prescriptions after 6 months.[15]

Starting follow-up after the third fill of a chronic medication will exclude patients who are least adherent. Unlike RCTs in which run-in phases are often done with placebo,[16] patients in routine care experience their first exposure to a new drug and may discontinue use because of a lack of effectiveness or intolerance during what may be the most vulnerable period for some medication-outcome relations. As long as that proportion is small and most patients discontinue for reasons not directly related to the study drug(s), this issue should be minor.

Cohort 4—Adherent Statin and Glaucoma Initiators—We further restricted cohort 3 to those subjects who were likely to be adherent to their medications, with adherence defined as filling a second and third prescription within 180 days after initiation. Follow-up began on the date of the third prescription.

Fifth Decision Point: Restriction to Specific High-Risk/Low-Risk Subgroups Represented in RCTs?

For RCTs all restrictions and most subgroup analyses need to be predetermined because RCTs are by design prospective and their size must be limited because of their high costs. Repeating an RCT with other inclusion criteria is usually not feasible. Therefore, the restrictive inclusion criteria of RCTs are used to focus on a population that may benefit most from an experimental treatment but are often compromises.[17]

In population-based database studies, analysts do not need to restrict the study population in the way RCTs do because the incremental cost for substantially expanding the study population is usually negligible. In database studies, several keystrokes will suffice to perform subgroup analyses to identify modifications of drug treatment effect that can be important for routine clinical practice.

Cohort 5—Subgroup Analysis of Typical Trial-Eligible Subjects—We restricted our final analysis to subjects from cohort 4 who would have been eligible to participate in the PROSPER trial[18] based on their baseline covariates as measured in claims data. In addition to meeting the PROSPER age criteria (70–82 years), subjects were required to have evidence of

one of the following risk factors as recorded in claims from the past year: angina, intermittent claudication, hypertension, diabetes, history of stroke, transient ischemic attack, MI, arterial surgery, amputation for vascular disease, or current smoking. Subjects who had had a stroke, transitory ischemic attack, MI, arterial surgery, or amputation for vascular disease in the past 6 months were excluded, as were those with a hospitalization for congestive heart failure, arrhythmia, or atrial fibrillation in the past 6 months. Finally, we eliminated subjects with evidence of dementia or cancer in the past year. These baseline characteristics did not include cholesterol measurements, which was an inclusion criterion from PROSPER.

Statistical Analysis

We obtained baseline demographic, health services use, and health status information from Medicare and PACE enrollment files and claims during the year before cohort entrance. Covariates included age, sex, race, receipt of preventive care services, use of laboratory tests, hospitalizations, number of distinct drugs used as well as use of drugs within specific classes, number of medical visits, nursing home residence, and presence of medical conditions as ascertained from inpatient and outpatient diagnosis codes. A full list of covariates and their definitions appears in the Appendix. Subjects without a record for a given condition were classified as not having that condition, and subjects with missing information on age, sex, and race were excluded.

We calculated unadjusted, age–sex adjusted, and fully adjusted rate ratios (RRs) from Cox proportional hazards models. Fully adjusted RRs came from a model adjusted for continuous exposure propensity scores. An exposure propensity score was calculated for each cohort by fitting a logistic model to predict exposure, in this case statin initiation, as a function of baseline covariates.[19,20] Similar to an ITT analysis in an RCT, our analysis assumed that patients remained on continuous therapy for the duration of the study. We did not censor patients discontinuing treatment from the analysis because of the potential for over- or underestimation by possible copredictors for discontinuation and outcome, but the current analysis could lead to some underestimation analogous to that seen in ITT analyses of RCTs.

This study was approved by the Brigham and Women's Hospital Institutional Review Board and was conducted under an ongoing data use agreement from the Centers for Medicare and Medicaid Services (CMS).

RESULTS

We identified 61,204 patients who used statins at least once during the study period. The majority (65%) were prevalent users of statins; 35% were new initiators of statins. Individual statins included simvastatin (33.3%), atorvastatin (28.8%), pravastatin (18.7%), fluvastatin (9.8%), lovastatin (6.9%), and cerivastatin (2.5%).

The mean age was 74.9 for all statin users and 80.3 for nonusers. Statin users were more likely than nonusers to have angina, coronary heart disease, diabetes, and hyperlipidemia. Baseline patient characteristics are presented in Table 1. Some of these imbalances became slightly less extreme as we applied more restriction criteria. For example, comparing cohort 0 with cohort 3, the difference in the prevalence of coronary atherosclerosis disease between exposed and comparison patients dropped from 12.5% to 8.2%, ischemic heart disease from 13.4% to 8.3%, and angina from 8.7% to 5.1%. Stronger reductions in the imbalance of patient characteristics were observed for risk factors for coronary heart disease, including diabetes (7.4–1.8%), and noncardiac comorbidities that may be generic markers for frailty, including chronic obstructive pulmonary disease (3.8–0.7%), osteoporosis (3.6–0.7%), urinary tract infections (4.4–0.5%), or current nursing home stay (4–0.9%).

In parallel to the increasing balance of patient characteristics, the unadjusted mortality RR increased steadily from cohort 0 (0.32) through cohort 5 (0.72, see Table 2). The numerically largest increases in RR of death were from cohort 0 to cohort 1 (excluding prevalent statin users) and from cohort 1 to cohort 2 (limiting nonusers to patients initiating glaucoma drugs). In the most restricted analysis (cohort 5), the unadjusted RR was 0.72. As the population size decreases with increasing restriction, the 95% confidence limits of the unadjusted analysis broadened as expected (Fig. 2A). RR estimates adjusted for age and sex also steadily increased. However, cohort 0 had an age- and sex-adjusted RR of 0.47, which was considerably higher than the unadjusted RR of 0.32; by contrast, the age- and sex-adjusted RR of 0.76 in cohort 5 was not much different from the unadjusted analysis in cohort 5.

Given multivariate adjustment using propensity scores of all covariates (in Table 1), the RRs started out at a higher level in cohort 0 (RR = 0.62) than in any other cohort, and they did not change substantially until we limited the comparison group to users of glaucoma medications (cohort 2, RR = 0.79). The RRs remained almost unchanged thereafter for all remaining cohorts (RR = 0.78, 0.80, 0.79, see Table 2 and Fig. 2B).

During 12 months of follow-up, 3.6% of matched nonusers initiated a statin during follow-up; 3.8% of nonusers matched on prescription or visit date started a statin; and 12% of glaucoma medication initiators started statins during follow-up. Thirty-six percent of incident statin users and 22.5% of prevalent users stopped taking their statin.

DISCUSSION

In an empirical analysis of the effect of statin use on 1-year mortality, we found that restricting the study population to incident statin users and the choice of comparison group both altered the effect estimates closer to the findings of RCTs in comparable populations. Subsequent restrictions to patients without contraindications and excluding nonadherent patients did not change this association in multivariate adjusted analyses. Our results need to be understood in the context of internal validity and generalizability.

Validity

Any discussion of validity requires a definition of a gold standard against which results are measured. For drug studies of this sort, the ultimate gold standard is the true biologic and physiologic effect of a drug in a given patient. RCTs are often seen as the gold standard method that comes closest to this goal. When compared with results from large population-based pharmacoepidemiologic studies, the question arises of whether the RCT results based on more selected populations can be generalized and continue to serve as the gold standard, or whether the observational data, not collected in a randomized trial setting, may provide more clinically relevant information. In our example, we assumed that RCT results are the gold standard for evaluating the validity of RR estimates at the fifth level of restriction, which was designed to mimic typical RCTs on statins. For all earlier restrictions, it is less clear how much the observed change in RR estimates result from effect modification by the restriction factors, bias, or a combination of both.

How closely did our analysis after the last level of restriction (restriction 5) match the findings of the RCTs? A meta-analyses of 3 RCTs of pravastatin showed a moderate overall mortality effect for patients 65 years and older [RR = 0.78; 95% confidence interval (CI), 0.68–0.89]. [21] The PROSPER trial,[18] an RCT of 5804 patients 7–82 years that compared pravastatin with placebo and that was not included in the above meta-analysis, found no overall mortality effect (RR = 0.97; 95% CI, 0.83–1.14). The Heart Protection Study,[22] a randomized controlled trial of 20,536 patients, compared simvastatin to placebo. The investigators found a mortality effect RR = 0.87 (95% CI, 0.81–0.94) in all age groups combined (40–80 years). The authors did not

report the effect estimate for mortality stratified by age, so we could not determine an estimate for the age group comparable to ours for this study. The 4S trial, an RCT of 4444 patients with coronary heart disease including 1021 patients ages 65 and older, showed a strong mortality effect of RR = 0.66 (95% CI, 0.48–0.90) in seniors.[23] The RCTs that most closely mimic our study population are those included in the pravastatin meta-analysis limited to patients 65 years and older, covering a wide range of baseline lipid levels and including primary and secondary prevention. The investigators reported a pooled pravastatin RR effect of 0.78 that was quite close to our results after 5 restrictions (RR = 0.79). The 4S trial was limited to secondary prevention and its somewhat stronger protective effect may therefore not generalize well to our population.

Assuming a moderate protective effect of statins based on RCTs, even our unadjusted results after applying all restrictions would have eliminated most of the bias. The multivariate adjusted analyses produced results comparable to those of RCTs after the second restriction (ie, limiting the study population to incident users of statins and limiting the comparison group to initiators of glaucoma drugs). We cannot claim with certainty that our results are entirely unbiased. Nonetheless, even if the gold standard estimate for our cohort 5 population had been the PROSPER trial results showing no effect on mortality, our combined restrictions seem to have reduced any bias substantially.

The point estimates of the multivariate-adjusted analyses did not change after the second restriction. This is partially because restrictions 3 (propensity score trimming) and 5 (restriction to typical trial exclusion criteria) are based on measured covariates that are adjusted for in the multivariate analysis. It is also attributable to the fact that statins have few contraindications, so that very few patients were excluded; exclusions based on contraindications can be expected to be more substantial in other settings.

Some might argue that pharmacoepidemiologic analyses of prevalent drug users should not be performed because they tend to underascertain events occurring shortly after drug initiation, and it becomes unclear whether measured patient characteristics are the cause or consequence of drug use. To address this issue, we purposely included the step from a mixed prevalent and incident users cohort to an incident user cohort to study that question. We found that this change resulted in an observable but not substantial change in RR estimates in our example.

Such empirical data on the changes caused by increasing restriction of the study population can guide decisionmakers in interpreting findings from pharmacoepidemiologic studies when RCT data are absent. In the end, however, it may not be possible to determine whether the changes in the effect estimate seen with each additional restriction demonstrate the likely magnitude of residual confounding or instead result from true effect modification by the restriction factors.

Generalizability

To guide our thinking about generalizability, it is useful to specify the patient to whom we wish to generalize our results. From a patient and physician perspective, the most relevant and frequently asked question is, "What is the effectiveness and safety of a particular drug that I am about to start and continue to use, compared with not starting therapy, or compared with starting an alternative drug?" From this viewpoint, restricting studies to initiators of drug therapy (restriction 1) does not limit generalizability. Instead, it avoids under-representation of treatment effects that occur shortly after initiation. Patients with known contraindications (or their clinicians) would usually not have to confront this hypothetical question because prescribing the drug in the first place would contravene current medical knowledge. Therefore, restriction 3 places little limits on generalizability.

In making a prescribing decision, physicians must assume that patients will take a drug as directed. If clinicians knew beforehand that a patient would not take a prescribed medication, they would not ponder the appropriateness of the drug in the first place. Consequently, excluding patients who are nonadherent to their treatment—independent of intolerance or treatment failure—will not limit generalizability to the question raised above (restriction 4). However, the situation is quite different if we restrict the study population by disease severity, comorbidities, polypharmacy, and other risk factors for the study outcome (restriction 5). Data based on such restrictions will limit physicians when making prescribing decisions concerning the excluded patient subgroups. The obvious solution to this problem is to stratify analyses according to relevant clinical subgroups, rather than restricting them out of the analysis altogether, and then evaluate the extent to which treatment effects differ across groups.[24] The large size of health care utilization databases can allow performing such subgroup analyses with substantial numbers of subjects, and represents an attractive alternative to wholesale restriction.

This study's primary goal was to demonstrate quantitatively the effects of 5 subsequent restrictions applied to a study population on observed drug effects—in this case, the estimate of association between statin use and mortality. We have chosen one of several possible sequences of the 5 restrictions. Applying the restrictions in another order may have resulted in slightly different results. We limited the study to a drug for chronic use for which discontinuation would not be related to either a cure or symptom relief, and used death as the ultimate and most generalizable outcome. We might have observed quite different effects of the restrictions if we had studied acute myocardial infarction, an outcome more strongly associated with the indication for statin use. We further simplified the statistical analysis by not modeling the drug exposure as a time-varying variable, but rather assuming constant use after initiation. This approach may lead to a bias towards the null. The choice of a comparison medication (ie, initiation of glaucoma drug use) may work better in our example of elderly patients[10,12,25] than in younger populations.

In conclusion, we identified a set of restrictions that analysts should consider in epidemiologic studies of the safety and effectiveness of therapeutics when using large observational databases. Such restrictions will place few limits on generalizability of research finding for most clinically relevant treatment choices.

Acknowledgments

REFERENCES

1. Maclure M, Schneeweiss S. Causation of bias: the episcope. Epidemiology 2001;12:114–122. [PubMed: 11138805]

2. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity. II. Observational studies. Lancet 2001;357:455–462. [PubMed: 11273081]

3. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215–1218. [PubMed: 8634569]

4. Schneeweiss S, Avorn J. Using health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–337. [PubMed: 15862718]

5. Perrio M, Waller PC, Shakir SAW. An analysis of the exclusion criteria used in observational pharmacoepidemiological studies. Pharmacoepidemiol Drug Saf 2007;16:329–336. [PubMed: 16741894]

6. Schneeweiss S, Glynn RJ, Avorn J, et al. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. J Clin Epidemiol 2005;58:98–102. [PubMed: 15649677]

7. Solomon DH, Schneeweiss S, Glynn RJ, et al. The relationship between selective COX-2 inhibitors and acute myocardial infarction. Circulation 2004;109:2068–2073. [PubMed: 15096449]

8. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003;158:915–920. [PubMed: 14585769]

9. Rothman, KJ. Epidemiology. An Introduction. New York, NY: Oxford University Press; 2002.

10. Glynn RJ, Knight EL, Levin R, et al. Paradoxical relations of drug treatment with mortality in older persons. Epidemiology 2001;12:682–689. [PubMed: 11679797]

11. Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. N Engl J Med 1998;338:1516–1520. [PubMed: 9593791]

12. Glynn RJ, Monane M, Gurwitz JH, et al. Aging, comorbidity, and reduced rates of drug treatment for diabetes mellitus. J Clin Epidemiol 1999;52:781–790. [PubMed: 10465323]

13. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577–581. [PubMed: 2057656]

14. Sturmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. Pharmacoepidemiol Drug Saf 2006;15:698–709. [PubMed: 16528796]

15. Benner JS, Glynn RJ, Mogun H, et al. Long-term persistence in use of statin therapy in elderly patients. JAMA 2002;288:455–461. [PubMed: 12132975]

16. Pablos-Mendez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA 1998;279:222–225. [PubMed: 9438743]

17. Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. JAMA 1992;268:1417–1422. [PubMed: 1512909]

18. Shepherd J, Blauw GJ, Murphy MB, et al. The design of a prospective study of Pravastatin in the Elderly at Risk (PROSPER). PROSPER Study Group. PROspective Study of Pravastatin in the Elderly at Risk. Am J Cardiol 1999;84:1192–1197. [PubMed: 10569329]

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

20. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757–763. [PubMed: 9382394]

21. Simes J, Furberg CD, Braunwald E, et al. for the Prospective Pravastatin Pooling Project investigators. Effects of pravastatin on mortality in patients with and without heart disease across a broad range of cholesterol levels. Eur Heart J 2002;23:207–215. [PubMed: 11792135]

22. Heart Protections Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 2002;360:7–22. [PubMed: 12114036]

23. Miettinen TA, Pyorala K, Olsson AG, et al. Cholesterol-lowering therapy in women and elderly patients with myocardial infarction or angina pectoris: findings from the Scandinavian Simvastatin Survival Study (4S). Circulation 1997;96:4211–4218. [PubMed: 9416884]

24. Rothwell PM. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. Lancet 2005;365:176–186. [PubMed: 15639301]

25. Glynn RJ, Schneeweiss S, Wang P, et al. Selective prescribing can lead to over-estimation of the benefits of lipid-lowering drugs. J Clin Epidemiol 2006;59:819–828. [PubMed: 16828675]
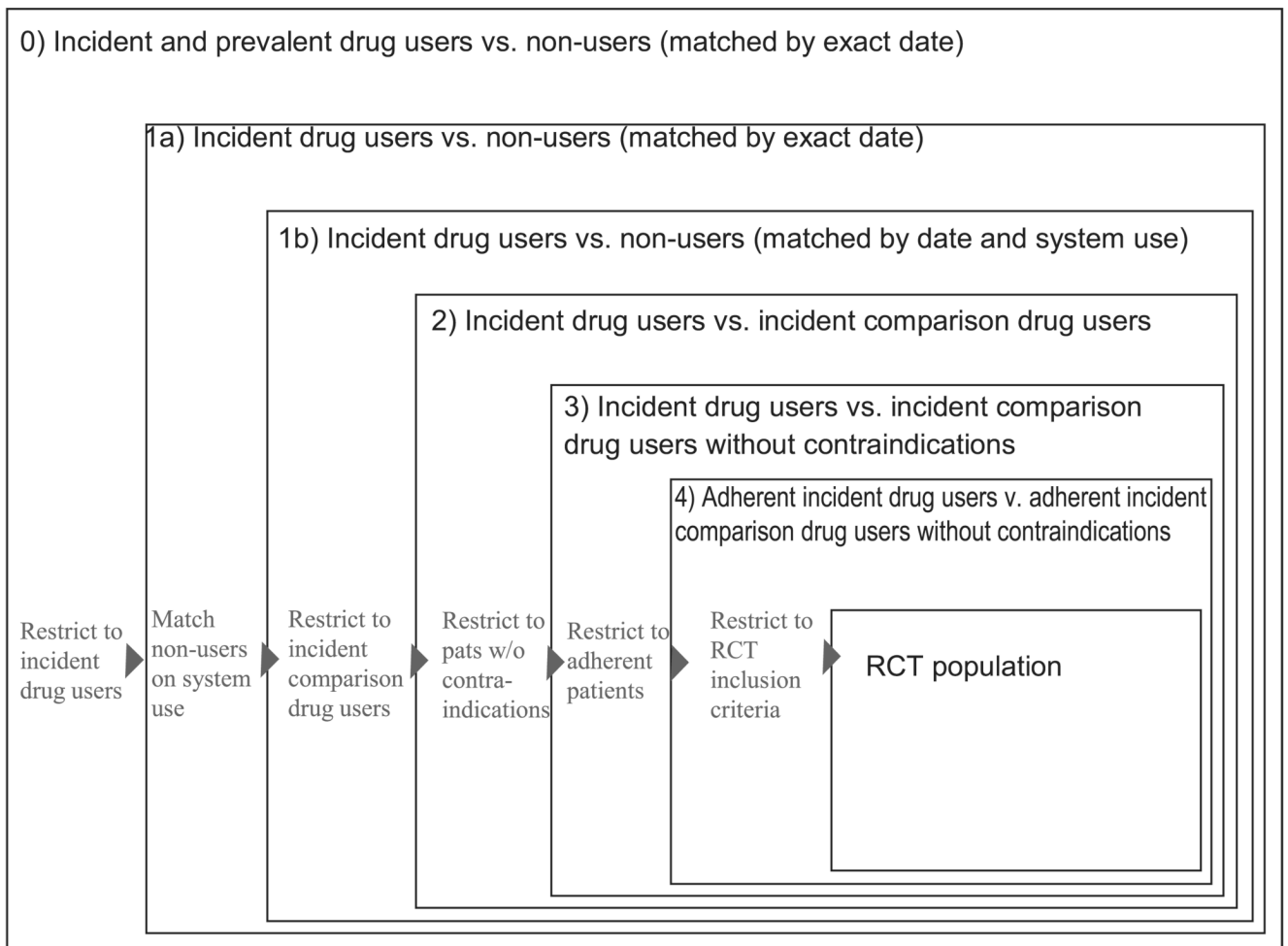
FIGURE 1.
Five consecutive population restrictions to approximate RCTs in pharmacoepidemiologic database studies.

A Unadjusted mortality rate ratio estimates:
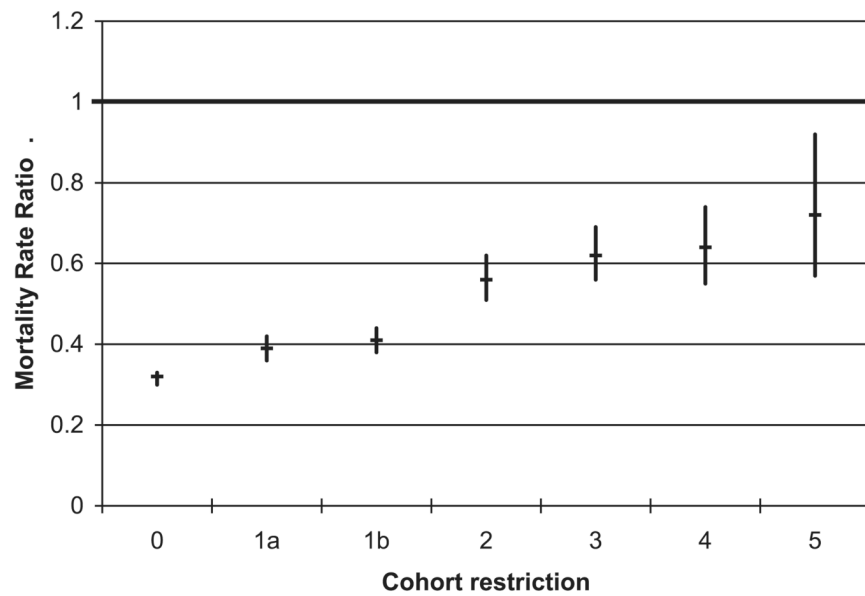
B Multivariate adjusted mortality rate ratio estimates:
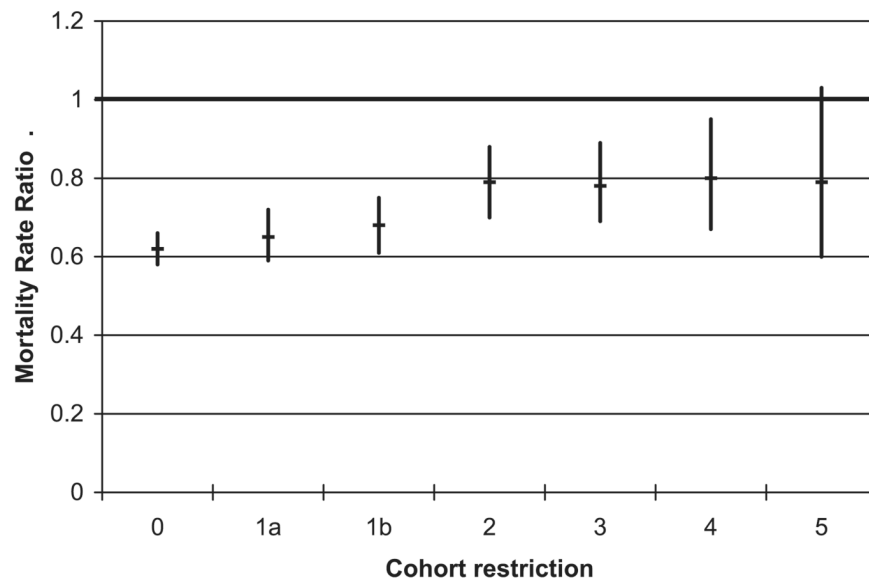
FIGURE 2.
Unadjusted and multivariate adjusted 1-year mortality rate ratios for 6 increasingly restricted cohorts of statin users. A, Unadjusted mortality rate ratio estimates; B, Multivariate adjusted mortality rate ratio estimates.

TABLE 1

Selected Baseline Characteristics of the Increasingly Restricted Study Cohorts[*]

	Cohort 0		Cohort 1: Incident Statin Users			Cohort 2[†]	Cohort 3: PS Trimmed		Cohort 4: Incident and Adherent		Cohort 5: RCT Eligible	
	Statin Incident/ Prevalent Statin User	Compar. Nonuser	Statin 1a, b, 2 Incident Statin User	Compar. 1a. Nonuser	Compar. 1b. Nonuser Matched on Rx or Visit Date	Compar. Glaucoma Medication Initiators	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users
N	61,204	61,204	21,233	21,233	21,233	14,889	18,482	12,013	14,220	7455	7984	3689
Demographics												
Age (mean, SD)	74.9 (6.1)	80.3 (7.3)	75.8 (6.0)	80.4 (7.3)	80.5 (7.2)	80.4 (6.8)	76.3 (5.9)	79.1 (6.3)	76.2 (5.9)	79.0 (6.3)	75.8 (3.5)	76.8 (3.5)
Male	18.5	19.4	19	19.2	18.5	17.2	17.9	16.9	17.8	16.4	17	17.7
Race												
White	92.8	93.5	92.7	93.5	94	89.7	92.4	90.1	93	90.8	92.9	89.5
Black	6	5.4	6.2	5.4	5	9.2	6.5	8.7	5.9	8.2	6.1	9.5
Other	1.2	1	1.1	1.1	1	1.1	1.1	1.2	1.1	1	1	1
Comorbid conditions—cardiovascular												
Comorbidity score (mean, SD)	1.7 (1.9)	1.9 (2.1)	1.9 (2.0)	1.9 (2.1)	1.9 (2.1)	1.9 (2.1)	1.9 (1.9)	1.9 (2.0)	1.9 (1.9)	1.8 (1.9)	1.7 (1.7)	1.7 (1.8)
Prior MI	7.5	3.6	10	3.4	3.5	3.2	6.5	3.5	6.7	3.1	3.4	2.3
Prior CABG or PTCA	5.4	1	7.7	0.9	0.9	1	3.4	0.8	3.5	0.7	1.1	0.3
Angina	32.7	24	36.9	24.1	24.9	25.6	32.3	27.2	32.5	25.6	31.1	24.6
Atrial fibrillation	9.2	14.3	11.7	14.6	14.6	12.6	11.2	11.9	11.5	10.9	8.9	8.4
Cardiovascular diagnoses (mean, SD)	4.1 (4.0)	3.3 (3.4)	4.8 (4.4)	3.4 (3.4)	3.4 (3.3)	3.5 (3.4)	4.3 (4.0)	3.6 (3.5)	4.3 (4.0)	3.4 (3.3)	3.8 (3.2)	3.2 (2.8)
Cardiovascular system symptoms	11.3	9.3	13.2	9.2	9.2	9.9	12.3	10.3	12.1	10.3	10.7	9.2
Chest pain	27.5	23.3	33.2	22.9	23.2	24.6	29.2	25.3	28.3	23.8	24.7	21.4
Complications of heart disease	11.8	10.3	13.1	10.2	10.1	11.1	12.1	11.4	12	11.2	11.3	10.1
Conduction disorders	4.9	5.2	6.2	5.4	5.1	5.2	5.5	5.3	5.5	4.8	4	3.4
Heart failure	24.5	29.6	28.3	30	30.2	29.2	27.2	28.3	27.1	26.7	24.3	24.8
Coronary atherosclerosis	40.1	27.6	45	28.2	28.6	29.2	40.1	31.9	40.2	30.5	38	27.8
Diabetes	34.5	27.1	36.3	27.4	28	32.9	36.2	34.4	36.7	34	39.2	41.1
Hyperlipidemia	77.5	21.6	70.6	21.3	21.5	28.1	67.9	33.6	68.3	33.8	69.9	34.6

	Cohort 0		Cohort 1: Incident Statin Users			Cohort 2†	Cohort 3: PS Trimmed		Cohort 4: Incident and Adherent		Cohort 5: RCT Eligible	
	Statin Incident/ Prevalent Statin User	Compar. Nonuser	Statin 1a,b, 2 Incident Statin User	Compar. 1a. Nonuser	Compar. 1b. Nonuser Matched on Rx or Visit Date	Compar. Glaucoma Medication Initiators	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users
Hypertension	75.2	69.9	78.6	69.8	71.2	73.3	77.9	74.4	78.5	74.4	85.2	83.6
Ischemic heart disease	46.2	32.8	50.3	33.4	33.5	34.3	45.7	37.4	45.7	35.8	43.9	33.2
Palpitations	3.7	3.1	4.4	2.9	3.2	3.3	4.2	3.5	4	3.4	3.8	3.1
Peripheral vascular disease	20.5	24.2	23.5	24.4	24.5	24.5	22.9	23.7	22.7	22.9	22.3	22.3
Stroke, transient ischemic attack	15.1	11.6	17.8	11.6	11.8	12.8	16.4	13.7	16.6	13.6	13.9	11.7
Comorbid conditions—other												
Alzheimer disease	3.4	9.1	3.7	9.4	8.5	6.8	3.9	5.6	3.9	4.9	0	0
Cancer	20.9	24.6	23.1	24.7	24.8	25.7	23.4	25.3	23.4	24.6	22.4	23.8
Depression	4.8	6.2	5.5	6.3	6.1	5.9	5.4	5.6	5.2	5.1	4.2	4.3
COPD	15.5	19.3	18.7	19.9	19.7	19	17.9	18.6	17.6	17.8	16.8	18.3
Prior hip fracture (frailty marker)	0.4	1.3	0.4	1.3	1.2	1.1	0.4	0.7	0.4	0.6	0.3	0.5
Osteoporosis (frailty marker)	6.8	10.4	8.2	10.2	10.4	10.1	8.5	9.2	8.5	9.2	7.9	8.2
Urinary tract infection (frailty marker)	13	17.4	15.3	17.6	16.8	16.5	15.2	15.7	14.6	15.2	13.1	13.4
Renal disease	3.8	3.9	4.5	3.8	3.5	4	4.3	4	4.3	3.6	3.7	3.7
Parkinson disease	1	2.4	1.1	2.4	2.4	1.9	1.1	1.7	1.1	1.6	0.9	1.4
Use of medications												
Different drugs taken (mean, SD)	7.8 (4.8)	7.6 (4.9)	7.9 (5.0)	7.6 (5.0)	7.8 (5.0)	8.2 (5.1)	7.9 (5.0)	8.2 (5.1)	7.8 (4.9)	7.9 (5.0)	7.7 (4.8)	8.0 (5.0)
Cardiovascular	82.3	76.4	84.1	76.6	78.7	77.1	82.7	79.6	83.6	79.7	87.5	84.7
Loop diuretic use	15.2	18.6	16.8	19.3	19.4	18.8	16.7	18.1	16.8	17	15.8	16.4
Nonsteroidal anti-inflammatory drugs	27.2	27.2	28.8	26.1	27	31.9	29.3	31.7	29.5	31.6	29.5	33.5
Estrogen	5.9	4.4	5.5	4.6	4.6	4.5	5.5	4.8	5.6	4.8	5.4	4.7
Use of health care services												

	Cohort 0		Cohort 1: Incident Statin Users			Cohort 2†	Cohort 3: PS Trimmed		Cohort 4: Incident and Adherent		Cohort 5: RCT Eligible	
	Statin Incident/ Prevalent Statin User	Compar. Nonuser	Statin 1a, b, 2 Incident Statin User	Compar. 1a. Nonuser	Compar. 1b. Nonuser Matched on Rx or Visit Date	Compar. Glaucoma Medication Initiators	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users	Statin Incident Statin Users	Compar. Incident Glaucoma Med Users
Preventive care	64.9	61.9	67.9	62.1	63.2	66.3	69.2	67	70.6	68.7	71.6	70.7
Hospitalization with cardiac diagnosis	14.7	9.6	19.7	9.7	9.7	9.3	14.6	9.9	14.4	8.8	9.1	6.7
Electrocardiogram	55.5	51.6	62.5	51.6	51.3	57.7	59.8	57.7	59.5	53.8	57.3	52.4
Hospitalization	27.2	29.6	33.3	29.8	28.7	29	29	27.9	28.8	25.4	21.1	21.6
Laboratory test	68.8	61.5	70.1	61.1	61.2	65.2	69.6	65.3	69.6	65.3	70.8	66.7
Lipid test	42.4	18.4	40.3	17.3	17.3	21.8	38.9	24.9	38.9	25.1	41.2	27.1
Nursing home stay	2.8	6.8	3.9	7.1	6.4	5.7	3.7	4.6	3.6	4.1	1.8	2.6
Office visits (mean, SD)	8.6 (6.1)	8.5 (6.4)	9.5 (6.3)	8.5 (6.4)	8.7 (6.3)	9.8 (6.8)	9.3 (6.3)	9.7 (6.9)	9.2 (6.2)	9.6 (6.6)	9.2 (6.1)	9.9 (6.5)
With cardiovascular diagnosis	6.7 (7.0)	6.2 (6.9)	7.8 (7.5)	6.2 (6.8)	6.3 (6.8)	6.5 (6.9)	7.2 (7.2)	6.7 (6.9)	7.2 (7.0)	6.4 (6.6)	6.8 (6.2)	6.2 (5.7)

*Numbers represent prevalence estimates in percent if not otherwise specified.

†Statin user group is the same as in cohort 1.

TABLE 2

Association Between Statin Use and 1-Year Mortality in Increasingly Restricted Patient Populations

Cohort	No. Subjects	Person-Years	Events	Event Rate	Unadjusted	Rate Ratios Age, Sex, Adjusted	Multivariate Adjusted	95% CI[*]
0) Incident and prevalent statin users versus nonusers								
Nonuser	61,204	54,847	5594	0.102				
Statin	61,204	56,779	1828	0.032	0.32	0.47	0.62	0.58–0.66
1a) Incident statin users versus nonusers								
Nonuser	21,233	19,495	1964	0.101				
Statin	21,233	20,104	784	0.039	0.39	0.54	0.65	0.59–0.72
1b) Incident statin users versus nonusers matched on Rx or visit date								
Nonuser	21,233	19,561	1868	0.095				
Statin	21,233	20,104	784	0.039	0.41	0.56	0.68	0.61–0.75
2) Incident statin users versus incident glaucoma medication users								
Glaucoma	14,889	13,805	955	0.069				
Statin	21,233	20,067	784	0.039	0.56	0.78	0.79	0.70–0.88
3) PS trimmed incident statin users versus incident glaucoma medication users								
Glaucoma	12,013	11,192	698	0.062				
Statin	18,482	17,436	678	0.039	0.62	0.76	0.78	0.69–0.89
4) Incident and adherent statin users versus incident and adherent glaucoma medication users								
Glaucoma	7455	6927	311	0.045				
Statin	14,220	13,426	384	0.029	0.64	0.77	0.80	0.67–0.95
5) PROSPER eligible								
Glaucoma	3689	3468	110	0.032				
Statin	7984	7615	174	0.023	0.72	0.76	0.79	0.60–1.03

95% confidence interval for the multivariate adjusted rate ratio estimates.

APPENDIX

Covariates Included in Outcome Models

Covariate	ICD-9 and CPT-4 Codes
Demographic information	
Age (linear)	
Sex	
Race (white, African American, other)	
Comorbid conditions	
Comorbidity score (linear)	
Alzheimer disease or other dementia	290.xx, 294.x, 330.x, or 331.x
Angina	411.xx, 413.xx or use of nitrates
Atrial fibrillation	427.31
Cancer	140.xx–172.xx, 174.xx–239.xx
Cardiovascular diagnoses, no. (categorical)	
Cardiovascular system symptoms	785.xx
Cataract	366.xx
Chest pain	786.5x
Chronic obstructive pulmonary disease	491.20, 491.21, 492.0, 492.8, 493.2x, 494, 496
Complications of heart disease	429.x
Conduction disorders	426.xx
Coronary atherosclerosis	414.0x
Depression	311.xx
Diabetes mellitus	250.xx or use of diabetes medication
Disorders of refraction and accommodation	367.9
Fracture of hip	820.x or 821.x with ICD-9 procedure of 79.x5
Fracture of wrist, humerous, or spine	805.xx, 812.1x, 812.5x, 813.xx
Gait abnormalities	781.2, 781.3
Heart failure	428.xx
Hyperlipidemia	272.0, 272.2–272.4, 272.9
Hyperparathyroidism	252.0
Hypertension	401.xx, 402.xx
Hyperthyroidism	242.xx
Hypothyroidism	244.9
Ischemic heart disease	414.xx
Lupus	710.0
Myocardial infarction	410.xx
Osteoarthrosis	715.xx
Osteoporosis	733.0x
Palpitations	785.1
Parkinson disease	332.0, 332.1, 333.0
Peripheral vascular disease	440.xx, 443.9
Renal disease	250.4, 250.40, 250.41, 403.0, 403.00, 403.01, 403.1, 403.10, 403.11, 403.9, 403.90, 403.91, 403.96, 404.0,

Covariate	ICD-9 and CPT-4 Codes
	404.01–404.03, 404.1, 404.10–404.13, 404.2, 404.3, 404.9, 404.90–404.93, 585, 585.0, 585.00, 586, 586.0, 586.00, 586.1, 586.3, 586.9, 587, 587.0, 581, 581.x, 581.09, 581.1, 581.2, 581.4, 581.8, 581.81, 581.84, 581.89, 581.9, 581.90, 582, 582.0, 582.1, 582.2, 582.81, 582.85, 582.89, 582.9, 583, 583.0
Rheumatoid arthritis	714.x
Stroke, transient ischemic attack	430.xx–435.xx
Syncope	780.2 or 458.0
Urinary tract infection	599.0
Use of health care services	
Bone mineral density test	
Electrocardiogram	CPT = 93000, 93005, 93010
Laboratory test	
Lipid test	CPT = 82465, 0026T, 82172, 83715, 83716, 83718, 83719, 83721, 80061
Hospitalization	
Hospitalization with a cardiac diagnosis	DRG = 104–118, 121–140, 143, 478, 479, 514–518, 525–527, 535, 536
Nursing home residency	
Office visits, no. (categorical)	
Covariate	ICD-9 and CPT-4 Codes
Preventive care (gynecological exam, prophylactic vaccination, routine medical exam, or screening mammogram)	V72.3, V04.8x, V70.0, V76.12
Revascularization (CABG or PTCA)	ICD-9 procedure 36.xx OR CPT-4 procedure of 33510, 33511, 33512, 33513, 33514, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33530, 33533, 33534, 33535, 33536, 33545 OR DRG 106, 107, 112
Visits with a cardiovascular diagnosis, no. (categorical)	
Use of medications	
Corticosteroids	
Cardiovascular	
Estrogen	
Loop diuretics	
Nonsteroidal antiinflammatory drugs	
No. distinct generic entities used (categorical)	
Osteoporosis drugs	
Psychoactive medications	