



Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2011 June ; 20(6): 551–559. doi:10.1002/pds.2098.

The implications of propensity score variable selection strategies in pharmacoepidemiology – an empirical illustration

Amanda R. Patrick¹, Sebastian Schneeweiss¹, M. Alan Brookhart², Robert J. Glynn^{1,3}, Kenneth J. Rothman⁴, Jerry Avorn¹, and Til Stürmer²

¹Division of Pharmacoepidemiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

²Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, North Carolina

³Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

⁴RTI Health Solutions, Research Triangle Park, NC

Abstract

Purpose—To examine the effect of variable selection strategies on the performance of propensity score (PS) methods in a study of statin initiation, mortality and hip fracture assuming a true mortality reduction of <15% and no effect on hip fracture.

Methods—We compared seniors initiating statins with seniors initiating glaucoma medications. Out of 202 covariates with a prevalence > 5%, PS variable selection strategies included none, a priori, factors predicting exposure, and factors predicting outcome. We estimated hazard ratios (HRs) for statin initiation on mortality and hip fracture from Cox models controlling for various PSs.

Results—During one year follow-up, 2,693 of 55,610 study subjects died and 496 suffered a hip fracture. The crude HR for statin initiators was 0.64 for mortality and 0.46 for hip fracture. Adjusting for the non-parsimonious PS yielded effect estimates of 0.83 (95%CI:0.75–0.93) and 0.72 (95%CI:0.56–0.93). Including in the PS only covariates associated with a greater than 20% increase or reduction in outcome rates yielded effect estimates of 0.84 (95%CI:0.75–0.94) and 0.76 (95%CI:0.61–0.95), which were closest to the effects predicted from randomized trials.

Conclusion—Due to the difficulty of pre-specifying all potential confounders of an exposure-outcome association, data-driven approaches to PS variable selection may be useful. Selecting covariates strongly associated with exposure but unrelated to outcome should be avoided, because this may increase bias. Selecting variables for PS based on their association with the outcome may help to reduce such bias.

INTRODUCTION

Propensity score (PS) methods¹ have become a common analytic approach for controlling confounding in non-experimental studies of treatment effects^{2,3}. Propensity scores combine information on a large number of covariates into a single variable representing a subject's probability of receiving a particular treatment, given his/her measured characteristics. This

Corresponding author: Amanda Patrick Division of Pharmacoepidemiology & Pharmacoeconomics Brigham and Women's Hospital & Harvard Medical School 1620 Tremont Street, Suite 3030 Boston, MA 02120 Phone: (617) 278-0925 Fax: (617) 232-8602 arpatrick@partners.org.

The authors have no conflicts of interest to report.

score can be used for matching, stratification, as a weighting factor, or as an adjustment factor in multivariable regression^{4,5}. Ideally, investigators would have detailed knowledge of potential confounders and their association with the exposure and outcome of interest and would use PS methods to balance these confounders. In many practical settings, however, knowledge is incomplete and investigators face a large collection of potential confounders.

In contrast to variable selection for conventional outcome models⁶, relatively little has been written on variable selection for PS models. Simulations by Drake have shown that the omission of confounders from PSs results in exposure effect estimates that are biased to the same degree and in the same direction as estimates obtained omitting the same confounders from a conventional outcome model⁷. Rubin and Thomas have recommended that all variables related to the outcome be included in a PS regardless of their association with exposure,⁸ a recommendation supported by simulations by Brookhart et al.⁹ These simulations and further work have shown that including variables unrelated to the outcome in the PS increases the variance of exposure effect estimates and, when unmeasured confounders are present, may also increase bias compared with the crude estimate.^{9,10} In practice, PS models are often constructed to maximize prediction of exposure with c statistics reported in 38–52% of published papers.^{11,3}

In this study, we explored how variable selection strategies for PS models affect the estimate of the PS-adjusted exposure-disease association. We compared a strategy of including covariates in the PS based their outcome associations with several common PS model building strategies in a study on the effect of statin exposure on all-cause mortality and hip fracture. These examples were chosen because estimates of the effect of statins on all-cause mortality and hip fracture are available from randomized controlled trials, providing a gold standard against which our estimates could be measured.

SUBJECTS AND METHODS

Study population

We identified a population of seniors aged 65+ enrolled in both Medicare and the Pennsylvania Pharmaceutical Assistance Contract for the Elderly (PACE) program who initiated a statin or glaucoma medication in 1995–2002, with no use of either drug in the preceding 12 months. While the contrast of interest is statin initiation versus non-use, we selected a subset of the non-user population, glaucoma medication initiators, as a referent group because these patients, like statin initiators, initiated a preventive therapy, thus reducing the potential for healthy user effects^{12,13,14}. Each patient's initiation date was used as the baseline for follow-up¹⁵. Subjects were counted as initiators only on their first initiation during the observation period.

Based on knowledge of the clinical outcomes and their risk factors, we identified a priori a number of potential confounders of the exposure / outcome associations in question. These included age, sex, race, calendar year, receipt of preventive services and laboratory tests, hospitalizations, number and types of drugs used, number of medical visits, nursing home residency, comorbidity score¹⁶ and the presence of specific medical conditions. These covariates were defined based on Medicare and PACE enrollment files and claims during the year before initiation of therapy. In addition to covariates selected a priori, we identified a list of potential confounders empirically. We selected medical conditions identified by 3-digit diagnosis code occurring in >5% of the population and drugs at the generic entity level prescribed with the same frequency. The arbitrary 5% cutpoint was selected because confounding by a dichotomous covariates is in part a function of its prevalence, such that a low prevalence bounds the potential for confounding. The final covariate list (including those selected a priori or on prevalence) included 202 covariates. With the exception of age,

which was modeled using a linear and quadratic term, all continuous covariates were modeled by including quintile indicators. Selection criteria were applied to indicators individually; in secondary analyses, indicators were selected as a group.

We used Medicare claims data to ascertain time to hip fracture (defined by a hospitalization or outpatient visit with a diagnosis code of 820.x–821.x plus an ICD-9 procedure code of 79.x5 or CPT-4 code of 27230–27248) and mortality during a one-year follow-up period. The positive predictive value (PPV) of the hip fracture definition has been estimated at 87 – 98% within the Medicare population¹⁷. Although the specificity could not be calculated in this validation study, the high PPV indicates a high specificity, limiting the potential for bias in relative measures of association.¹⁸ We restricted fracture analyses to subjects without a hip fracture in the year before baseline. Follow-up was censored at the first of 365 days, death, or development of the outcome.

Analytical strategies—We estimated propensity scores for each patient by fitting a logistic regression model to predict statin versus glaucoma medication initiation, as a function of baseline covariates. We constructed eight different PSs for each outcome: a “non-parsimonious” PS including all 202 baseline covariates, an “a priori” PS including only predictors thought, a priori, to be independently associated with the outcome, a “stepwise” PS, and four scores including covariates selected based on their associations with the outcome. We included age, a quadratic age term, and sex in all PS models.

Selection of variables based on association with the exposure: Because stepwise selection procedures are used in PS construction by some researchers¹¹, we included a “stepwise” PS in our analysis. Variables were selected from the pool of 202 baseline covariates using a stepwise selection procedure with an inclusion criterion of $p \leq 0.2$ and retention criterion of $p \leq 0.1$.

Selection of variables based on association with the outcome: We employed several strategies to identify covariates based on their association with the outcome of interest. We began by entering each covariate, individually, into a Cox proportional hazards model predicting the outcome of interest adjusted for age, age², sex, and exposure. To select covariates for the “outcome +/- 20%” and “outcome +/- 30%” models, we examined the point estimates of the covariate-outcome associations from these Cox models. For the “outcome +/- 20%” model, we selected covariates with an outcome HR of > 1.2 or $< 1/1.2$ to make our criteria symmetric on the log scale. Our more stringent “outcome +/- 30%” criteria selected covariates with an HR > 1.3 or $< 1/1.3$. The “outcome $p < 0.2$ ” and “outcome $p < 0.1$ ” relied on the significance of covariate-outcome associations.

Selection of variables based on their association with the outcome or the exposure: In this hybrid PS, we included all covariates that were selected into either the “outcome +/- 20%” or the “stepwise” PS.

Selection of variables based on subject-matter knowledge: Because the selection of variables based on detailed clinical knowledge of exposure-outcome relationships might seem preferable to data-driven approaches, we tested the approach of pre-specifying our covariate list. “A priori” PS covariates were identified based on the current medical literature. For hip fracture, this list included factors known to contribute to low bone mineral density or falls (see appendix). Indicators of major causes of mortality and potential indicators of frailty were selected for inclusion in the mortality model (appendix).

Estimation of exposure effect estimates: The relation between statin initiation and each outcome was estimated using a Cox proportional hazards model, with glaucoma drug

initiators serving as the referent group. We estimated exposure / outcome associations stratified on PS decile. In keeping with the typical application of PS methods, our primary analyses were conducted on a trimmed dataset such that subjects with PSs falling in regions of non-overlap (i.e. statin exposed subjects with PSs higher than those of any unexposed subjects and vice versa) were excluded. We conducted a series of secondary analyses (see appendix) exploring the effects of trimming, propensity score parameterization, and the inclusion of interaction terms in the PS. Lastly, we examined the variables responsible for differences in the exposure effect estimates obtained from different models.

We took RCT evidence as the gold standard for treatment efficacy. We assumed that statins reduce all-cause mortality by 15% as reported in a recent meta-analysis of RCT data from elderly adults¹⁹ and have no effect on hip fracture, as reported in another recent meta-analysis of data from secondary analyses of four statin trials (OR: 1.03, 95% CI: 0.91 – 1.16)²⁰. We further assumed that the effectiveness in our unselected population would be attenuated relative to the RCT efficacy.²¹

Model-based standard errors obtained from a regression analysis are known to underestimate the true uncertainty surrounding a treatment effect estimate when data-driven model selection strategies are used. To determine whether this is an issue in using data-driven approaches to PS variable selection, we performed bootstrapping analyses, re-sampling our data with replacement 1,000 times, and conducting the model selection strategies and analyses on each sample. We used the standard-deviations of the resulting distributions of statin effect estimates to construct empirical confidence intervals for the estimates from our main analyses.

RESULTS

During 1995–2002, 40,721 PACE-Medicare enrollees initiated statins and 14,889 initiated glaucoma medications. Compared with glaucoma medication initiators, statin initiators were slightly younger, and were more likely to have cardiovascular disease (see table 1)

During up to one year of follow-up, 2,693 subjects (4.8%, table 2) died. Statin exposure was associated with a 36% (HR: 0.64, 95% CI: 0.59–0.69) reduction in mortality in unadjusted analyses. Adjustment for the non-parsimonious PS shifted the estimate to 0.83 (0.75 – 0.93). Adjustment for a PS including only covariates associated with a 20% increase or reduction in outcome rate (outcome \pm 20%) yielded an estimate of 0.84 (0.75 – 0.94). PSs constructed using the p-value based and stricter outcome association criteria yielded estimates of 0.81 to 0.83.

Among the 55,276 subjects without a prior hip fracture, 496 (0.8%) experienced hip fractures during follow-up. In unadjusted analyses, statin use was associated with a 54% reduction in the risk of hip fracture (HR: 0.46 95% CI: 0.39 – 0.55). Adjustment for nonparsimonious PS moved the effect estimate to 0.72 (0.56 – 0.93). The stepwise PS and stepwise \pm 20% PSs yielded similar estimates. Adjustment for the outcome \pm 20% PS yielded an estimate of 0.76 (0.61 – 0.95). Again, PS models including covariates meeting stricter or p-value based outcome-association criteria yielded larger apparent protective effects.

As shown in table 3, the difference between the statin effect estimates obtained adjusting for the “outcome \pm 20%” PS and those obtained adjusting for the non-parsimonious PS was driven largely by the inclusion of a glaucoma diagnosis covariate in the latter PS. This covariate was strongly inversely associated with statin exposure (OR=0.07) because of the comparison group used, but was not independently associated with hip fracture or mortality under our “outcome \pm 20%” criteria. The covariate was included when the outcome

association criteria were relaxed slightly (outcome HR > 1.175 or < 1/1.175). When the glaucoma diagnosis covariate was forced into the outcome \pm 20% PS and this PS was used to control for confounding, the resulting statin effect estimates for hip fracture and mortality were 0.69 and 0.82 – close to the values obtained adjusting for the non-parsimonious PS.

In addition to being associated with a change in statin effect estimate, addition of the glaucoma diagnosis covariate to the PS was associated with a jump in c statistic. The “outcome \pm 20%” PSs for hip fracture and mortality had c statistics of 0.81 and 0.82. These c statistics increased to 0.89 and 0.90 when the glaucoma diagnosis covariate was forced into the model. The highest c statistics (0.91) were consistently obtained from the non-parsimonious, stepwise, and “stepwise or \pm 20%” PSs. The greater discrimination between exposed and unexposed by the non-parsimonious PS compared with the outcome \pm 20% PS can be seen in the PS distributions in figure 1.

The model-based standard errors for statin effects estimated adjusting for PSs constructed using outcome-association variable selection criteria were consistently underestimated (table 4). This problem was greatest using the “outcome \pm 20%” PS, where the standard errors for the estimated effect of statin use on hip fracture and mortality were underestimated by 19% and 14% respectively. Excluding the prior glaucoma diagnosis covariate from the covariate list considerably reduced the degree of underestimation. The model-based standard errors for the non-parsimonious, a priori and stepwise PS models were correct.

Our results were unchanged in secondary analysis exploring the effects of not trimming for PS non-overlap, propensity score parameterization as splines versus deciles, selection of covariate quintile indicators as a group vs. individually, and the inclusion of interaction terms in the PS (see appendix).

DISCUSSION

In a study of the effects of statin use on mortality and hip fracture, we found that the effect estimates obtained varied across a range of commonly-employed PS variable selection strategies. Estimates of the effect of statins on mortality varied from 0.81 (0.73 – 0.90) to 0.84 (0.75 – 0.94) and estimates of the effect of statins on hip fracture varied from 0.63 (95 % CI: 0.52 – 0.76) to 0.76 (0.61 – 0.95). While our findings are not unexpected given the theory and existing simulation-based evidence on PS model construction, they illustrate the implications of PS variable selection strategy in a practical pharmacoepidemiologic setting for the first time.

For both outcomes, adjusting for the non-parsimonious PS yielded apparent protective effects that were larger than those reported in meta-analyses of RCTs: i.e. a 15% reduction in all-cause mortality²² and no protection (RR=1.03) for hip fracture.²³ Because effectiveness is likely to be less than efficacy in our population,²⁴ we assumed the true effect for mortality in our population would be no more than a 15% reduction. Adjusting for a PS including only covariates empirically associated with the outcome – i.e. those associated with a greater than 20% increase or reduction in outcome rates, adjusted for age, sex, and exposure – yielded estimates that were closer to the assumed true effects.

Although these differences were not pronounced, our findings are consistent with the results of recent simulations which suggested that covariates associated with the outcome should be included in the PS regardless of their association with exposure and that covariates strongly associated with exposure and unassociated or only weakly associated with the outcome should be avoided, as these covariates can increase the variance and bias of effect estimates.⁹ Our results are also consistent with those of a recent paper demonstrating that the

inclusion of covariates strongly associated with exposure and unassociated or only weakly associated with the outcome in a PS can increase bias¹⁰. Our findings were mainly driven by the presence of a covariate strongly predictive of exposure status and likely unrelated to hip fracture and mortality. They argue in favor of considering outcome associations when selecting covariates to be entered into the PS. A model including only covariates specified a priori did not perform as well, attesting to the potential value of empirical variable selection strategies unless all confounders are known.

Our work adds to evidence that the c statistic may not be a useful measure of the ability of a PS to adjust for confounding. Work by Weitzen et al., has shown that model discrimination tests were not useful in detecting the omission of a confounder from a PS model.²⁵ In our case, the addition of the prior glaucoma diagnosis covariate to a PS increased its c statistic considerably, but also increased the bias. Rather than using the c statistic as a quality criterion, our findings suggests that pronounced changes in the c statistic of the PS model after inclusion of a single covariate may be a useful tool in detecting covariates strongly predictive of exposure that should then be subjected to further scrutiny with respect to their potential as a confounder.

Our study suffers from several limitations. We examined only a handful of possible approaches that could be applied to the PS variable selection problem. We did not use cross-validation in our variable selection procedures because existing literature suggests that the estimated PS generally performs better to control for confounding than the true PS.^{26,27} Despite its advantages³⁰, the use of an active control group may increase the risk of encountering a covariate that is strongly related to exposure but unrelated or only weakly related to the outcome. However, this type of variable, a prime example of which is any variable that might be used as an instrument, such as physician preference or distance to a provider, is likely to exist in other settings as well.²⁸ It is possible that covariate values were misclassified in our study due to our reliance on administrative claims data from the year preceding medication initiation. Disease history is captured only to the extent that diagnoses are accurately captured in the billing process, and it is possible that some chronic conditions were missed due to our use of a 12-month versus longer covariate assessment period. Misclassification may partly explain the divergence of our results from those reported in RCTs. In addition, it should be noted that while a meta-analysis of RCT data reported no-effect of statins on hip fracture, observational studies have generally,^{29,30,20} although not always^{31,20} reported protective effects. While confounding has been frequently cited as an explanation for the divergence of findings, it is also possible that effect measure modification or other factors are responsible.

Our consideration of covariate-outcome associations in the first step of some PS building strategies seems to differ from the design notion proposed by Rubin³² of creating comparable groups of exposed and unexposed using the PS without information on the outcome. While Rubin mentions that “variables that are effectively known to have no possible connection to the outcomes” should not be included in a list used to define comparable groups of exposed and unexposed, he argues that the outcome should be locked away during the design phase.³² Our use of outcome information to select covariates for which our groups of exposed and unexposed should be balanced differs somewhat from this approach, but it should be noted that we are using outcome information only to estimate the association between the potential covariate and the outcome and not to assesses which covariates have what confounding effect. While not locking away the outcome when implementing our method, we are still agnostic about the exposure-outcome association and the effects of the covariates on this association. An alternative approach could be to evaluate covariate / outcome associations within the comparator group only.^{33, 34}

While automated PS variable selection approaches are appealing in their ease of execution and ability to identify confounders that may be missed by the study investigators, they suffer from some limitations and should not take the place of a thoughtful consideration of confounding by the study investigators³⁵. One disadvantage to any automated variable selection strategy, including those we used to select covariates associated with the outcome, is that it requires setting an arbitrary threshold. The performance of our method based on the covariate-outcome associations was sensitive to the threshold chosen and performed variably in bootstrapped samples, resulting in larger standard errors based on bootstrapped datasets compared with the model-based standard errors. Nevertheless, the relatively poor performance of the PSs constructed based on a priori hypotheses attests to the difficulty of correctly pre-specifying all empirical confounders of an exposure effect in administrative claims data. A middle ground might be to use a data-driven approach to identify a pool of potential covariates (i.e. based on prevalence) from which potentially problematic covariates (i.e. those strongly related to exposure and not thought to be related to the outcome) could be identified and excluded. In the absence of covariates strongly related to exposure, the same PS could be used to study multiple outcomes; in the presence of such covariates, the suitability of a generic PS would need to be evaluated for each outcome of interest. This approach would preserve some of the advantage of PSs over traditional outcome models – namely the ability to use a single score to study multiple outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by grants RO1 AG023178 and RO1 AG018833 from the National Institute on Aging at the National Institutes of Health.

References

1. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
2. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*. 2005; 58(6):550–9. [PubMed: 15878468]
3. Stürmer 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. [PubMed: 16632131]
4. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. Oct 15; 1998 17(19):2265–81. [PubMed: 9802183]
5. Stürmer 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*. May 1; 2005 161(9):891–8. [PubMed: 15840622]
6. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol*. Dec 1; 1993 138(11):923–36. [PubMed: 8256780]
7. Drake C. Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*. 1993; 48:1231–1236.
8. Rubin DR, Thomas N. Matching using estimated propensity score: relating theory to practice. *Biometrics*. 1996; 52:249–264. [PubMed: 8934595]
9. Brookhart MA, Schneeweiss S, Rothman KJ, Avorn J, Stürmer T. Variable selection in propensity score models. *Am J Epidemiol*. Jun 15; 2006 163(12):1149–56. [PubMed: 16624967]

10. Bhattacharya, J.; Vogt, WB. Do instrumental variables belong in propensity scores?. National Bureau of Economic Research; 2007. Working Paper 343
11. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf.* Dec; 2004 13(12):841–53. [PubMed: 15386709]
12. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology.* Nov; 2001 12(6):682–9. [PubMed: 11679797]
13. Schneeweiss S, Patrick AR, Stürmer T, Brookhart MA, Avorn J, Maclure M, Rothman KJ, Glynn RJ. Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison with Randomized Trial Results. *Med Care.* Oct; 2007 45(10 Supl 2):S131–42. [PubMed: 17909372]
14. Glynn RJ, Schneeweiss S, Wang PS, Levin R, Avorn J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. *J Clin Epidemiol.* Aug; 2006 59(8):819–28. [PubMed: 16828675]
15. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* Nov 1; 2003 158(9):915–20. [PubMed: 14585769]
16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis.* 1987; 40:373–83. [PubMed: 3558716]
17. Baron JA, Lu-Yao G, Barrett J, McLerran D, Fisher ES. Internal validation of Medicare claims data. *Epidemiology.* 1994; 5(5):541–4. [PubMed: 7986870]
18. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* 1977; 105:488–95. [PubMed: 871121]
19. Roberts CGP, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. *J Gerontol A Biol Sci Med Sci.* 2007; 62:879–887. [PubMed: 17702880]
20. Toh S, Hernández-Díaz S. Statins and fracture risk. A systematic review. *Pharmacoepidemiol Drug Saf.* Jun; 2007 16(6):627–40. [PubMed: 17286319]
21. Strom, BL., editor. *Pharmacoepidemiology.* 4th edition. John Wiley & Sons Ltd; Hoboken, NJ: 2005.
22. Roberts CGP, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. *J Gerontol A Biol Sci Med Sci.* 2007; 62:879–887. [PubMed: 17702880]
23. Toh S, Hernández-Díaz S. Statins and fracture risk. A systematic review. *Pharmacoepidemiol Drug Saf.* Jun; 2007 16(6):627–40.
24. Strom, BL., editor. *Pharmacoepidemiology.* 4th edition. John Wiley & Sons Ltd; Hoboken, NJ: 2005.
25. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Weaknesses of goodness-of-fit tests for evaluating propensity score models: the case of the omitted confounder. *Pharmacoepidemiol Drug Saf.* Apr; 2005 14(4):227–38. [PubMed: 15386700]
26. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics.* Jun; 1992 48(2):479–95. [PubMed: 1637973]
27. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol.* Aug 15; 1999 150(4):327–33. [PubMed: 10453808]
28. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology.* May; 2006 17(3):268–75. PubMed PMID: 16617275. [PubMed: 16617275]
29. de Vries F, de Vries C, Cooper C, Leufkens B, van Staa TP. Reanalysis of two studies with contrasting results on the association between statin use and fracture risk: the General Practice Research Database. *Int J Epidemiol.* Oct; 2006 35(5):1301–8. [PubMed: 17053011]
30. Chan KA, Andrade SE, Boles M, Buist DS, Chase GA, Donahue JG, Goodman MJ, Gurwitz JH, LaCroix AZ, Platt R. Inhibitors of hydroxymethylglutaryl-coenzyme A reductase and risk of fracture among older women. *Lancet.* Jun 24; 2000 355(9222):2185–8. [PubMed: 10881890]
31. LaCroix AZ, Cauley JA, Pettinger M, Hsia J, Bauer DC, McGowan J, Chen Z, Lewis CE, McNeeley SG, Passaro MD, Jackson RD. Statin use, clinical fracture, and bone density in

- postmenopausal women: results from the Women's Health Initiative Observational Study. *Ann Intern Med.* Jul 15; 2003 139(2):97–104. [PubMed: 12859159]
32. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med.* Jan 15; 2007 26(1):20–36. [PubMed: 17072897]
33. Hansen BB. The prognostic analogue of the propensity score. *Biometrika.* Jun; 2008 95(2):481–488.
34. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res.* Feb; 2009 18(1):67–80. Epub 2008 Jun 18. [PubMed: 18562398]
35. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care.* Jun; 2010 48(6 Suppl):S114–20. [PubMed: 20473199]
36. Roberts CGP, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. *J Gerontol A Biol Sci Med Sci.* 2007; 62:879–887. [PubMed: 17702880]
37. Toh S, Hernández-Díaz S. Statins and fracture risk. A systematic review. *Pharmacoepidemiol Drug Saf.* Jun; 2007 16(6):627–40.
38. Strom, BL., editor. *Pharmacoepidemiology.* 4th edition. John Wiley & Sons Ltd; Hoboken, NJ: 2005.
39. Roberts CGP, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. *J Gerontol A Biol Sci Med Sci.* 2007; 62:879–887. [PubMed: 17702880]
40. Toh S, Hernández-Díaz S. Statins and fracture risk. A systematic review. *Pharmacoepidemiol Drug Saf.* Jun; 2007 16(6):627–40.
41. Strom, BL., editor. *Pharmacoepidemiology.* 4th edition. John Wiley & Sons Ltd; Hoboken, NJ: 2005.

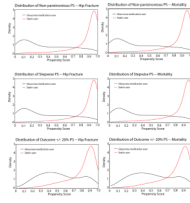


Figure 1.
Estimated density of propensity score

Table 1

Distribution of covariates prior to drug initiation

	Statin initiators*	Glaucoma Initiators*
N	40,721	14,889
Demographic		
Age (mean, SD)	76.8 (6.1)	80.4 (6.8)
Sex (female)	33417 (82.1)	12329 (82.8)
Race		
white	37925 (93.1)	13355 (89.7)
black	2335 (5.7)	1369 (9.2)
other	461 (1.1)	165(1.1)
Diagnoses		
Cardiovascular		
MI*	4289 (10.5)	478 (3.2)
Prior CABG or PTCA	3207 (7.9)	146 (1)
Angina	9393 (23.1)	1800 (12.1)
Ischemic Heart Disease	20764 (51.0)	5104 (34.3)
Asymptomatic CVD	5217 (12.8)	1649 (11.1)
Coronary atherosclerosis	18911 (46.4)	4351 (29.2)
Stroke / TIA	7068 (18.7)	1911 (12.8)
Hyperlipidemia	29229 (71.8)	4190 (28.1)
Diabetes	15255 (37.5)	4841 (32.5)
Hypertension	32896 (80.8)	10907 (73.3)
Conduction disorders	2718 (6.7)	778 (5.2)
Heart Failure	12507 (30.7)	4349 (29.2)
Atrial Fibrillation	5391 (13.2)	1872 (12.6)
Osteoporosis	4029 (9.9)	1497 (10.1)
Prior hip fracture [†]	177 (0.4)	157 (1.1)
Prior fracture of wrist, spine or humerus	1037 (2.5)	510 (3.4)
Disorders of refraction	108 (0.3)	53 (0.4)
Blindness	477 (1.2)	341 (2.3)
Cataract	14532 (35.7)	8500 (57.1)
Syncope	2962 (7.3)	1,062 (7.1)
Gait abnormality	2023 (5.0)	824 (5.5)
COPD	7806 (19.2)	2825 (19.0)
Rheumatoid arthritis	1309 (3.2)	577 (3.9)
Arthritis	14127 (34.7)	5404 (36.3)
Hyperthyroidism	1163 (2.9)	355 (2.4)
Hyperparathyroidism	201 (0.5)	76 (0.5)
Falls	1805 (4.4)	944 (6.3)

	Statin initiators*	Glaucoma Initiators*
N	40,721	14,889
Cancer	9646 (23.7)	3833 (25.7)
Urinary tract infection	6382 (15.7)	2,455 (16.5)
Alzheimer's disease	1906 (4.7)	1019 (6.8)
Parkinson's disease	546 (1.3)	281 (1.9)
Depression	2454 (6.0)	877 (5.9)
Comorbidity Score (mean, SD)	1.93 (2.1)	2.03 (1.98)
Health System Service Use		
Use of preventive care services	28280 (69.4)	9878 (66.3)
Bone mineral density testing	695 (1.7)	245 (1.6)
Nursing home residence	1863 (4.6)	855 (5.7)
Hospitalization	14046 (34.5)	4318 (29.0)
Number of drugs used		
1-3	6119 (15.0)	2444 (16.4)
4-5	7505 (18.4)	2892 (19.4)
6-8	10622 (26.1)	3754 (25.2)
9-11	7388 (18.1)	2574 (17.3)
>12	9087 (22.3)	3225 (21.7)
Number of physician visits		
0-2	8419 (20.7)	2892 (19.4)
3-4	6979 (17.1)	2395 (16.1)
5-7	9167 (22.5)	3225 (21.7)
8-11	7987 (19.6)	2980 (20.0)
12+	8169 (20.1)	3397 (22.8)
Cardiovascular medication use	34918 (85.7)	11477 (77.1)
NSAIDS use	12240 (30.1)	4748 (31.9)
Hormone therapy	2261 (5.6)	674 (4.5)
Corticosteroid use	2785 (6.8)	1201 (8.1)
Loop diuretics	7852 (19.3)	2797(18.8)
Osteoporosis med use	3,327 (8.2)	1173 (7.9)
Psychoactive med use	8,799 (21.6)	3313 (22.3)

* N (%) unless stated otherwise; SD: standard deviation

† patients with prior hip fracture were excluded from statin-hip fracture analysis

Table 2

Effect estimates from unadjusted and propensity score adjusted models

	PS Characteristics				Outcome Effect Estimate [‡]		
	N*	N Events*	C statistic	N covariates	Estimate	SE [§]	HR 95% CI [†]
ALL-CAUSE MORTALITY							
Unadjusted	55610	2693			-0.45	0.04	0.64 0.59 – 0.69
Adjust for age and sex	55610	2693			-0.16	0.04	0.86 0.79 – 0.93
Adjusted for PS:							
Non-parsimonious	55466	2682	0.91	202	-0.18	0.06	0.83 0.75 – 0.93
A priori	55544	2683	0.82	67	-0.19	0.05	0.83 0.76 – 0.91
Step wise	55454	2683	0.91	82	-0.21	0.06	0.81 0.73 – 0.90
Outcome +/- 20%	55582	2686	0.82	143	-0.18	0.06	0.84 0.75 – 0.94
Outcome +/- 30%	55604	2692	0.81	127	-0.19	0.05	0.83 0.75 – 0.91
Outcome p < 0.2	55310	2667	0.90	168	-0.20	0.06	0.82 0.73 – 0.91
Outcome p < 0.1	55337	2669	0.90	158	-0.21	0.06	0.81 0.72 – 0.90
Stepwise or +/- 20%	55464	2682	0.91	166	-0.18	0.06	0.83 0.75 – 0.93
Hip FRACTURE							
Unadjusted	55276	496			-0.77	0.09	0.46 0.39 – 0.55
Adjust for age and sex	55276	496			-0.40	0.09	0.67 0.56 – 0.81
Adjusted for PS:							
Non-parsimonious	55115	495	0.91	201	-0.32	0.13	0.72 0.56 – 0.93
A priori	55264	496	0.75	56	-0.47	0.10	0.63 0.52 – 0.76
Stepwise	55112	494	0.91	85	-0.30	0.13	0.74 0.58 – 0.96
Outcome +/- 20%	55258	496	0.81	120	-0.27	0.14	0.76 0.61 – 0.95
Outcome +/- 30%	55273	496	0.80	87	-0.35	0.14	0.71 0.57 – 0.87
Outcome p < 0.2	55206	495	0.89	120	-0.38	0.13	0.69 0.54 – 0.88
Outcome p < 0.1	55256	495	0.81	96	-0.31	0.14	0.73 0.59 – 0.91
Stepwise or +/- 20%	55081	493	0.91	153	-0.32	0.13	0.73 0.57 – 0.94

* N subjects and events may vary due to exclusion of subjects with prior event from hip fracture analyses as well as trimming of subjects with propensity scores falling in areas of non-overlap. Hip fracture analyses were restricted to subjects without a hip fracture in the year prior to baseline.

[§] Standard error from 1,000 bootstrap

[‡] Based on the standard-deviation of the statin effect estimates over 1,000 bootstraps

[‡] Outcome effect estimates were compared to statin effect estimates from meta-analyses of randomized controlled trials: a 15% reduction in all-cause mortality (95% CI: 7% to 22%)³⁶ and no effect on hip fracture (OR: 1.03, 95% CI: 0.91 – 1.16)³⁷. We further assumed that the effectiveness in our unselected population would be attenuated relative to the RCT efficacy.³⁸

Table 3

Effect of outcome-association cut-point used as variable selection criterion

Model	PS Characteristics		Outcome Effect Estimate	
	N covariates	C statistic	Estimate	HR
	Mortality			
Non-parsimonious	202	0.91	-0.18	0.83
Outcome +/- 5%	181	0.91	-0.20	0.82
Outcome +/- 7.5%	172	0.90	-0.21	0.81
Outcome +/- 10%	163	0.90	-0.21	0.81
Outcome +/- 12.5%	157	0.90	-0.22	0.81
Outcome +/- 15%	151	0.90	-0.22	0.80
Outcome +/- 17.5%	149	0.90	-0.22	0.81
Outcome +/- 20% + glaucoma diagnosis	144	0.90	-0.20	0.82
Outcome +/- 20%	143	0.82	-0.18	0.84
Outcome +/- 22.5 %	139	0.81	-0.17	0.84
Outcome +/- 25%	135	0.81	-0.17	0.84
Outcome +/- 27.5%	130	0.81	-0.18	0.83
Outcome +/- 30%	127	0.81	-0.19	0.83
	Hip Fracture			
Non-parsimonious	201	0.91	-0.32	0.72
Outcome +/- 5%	177	0.90	-0.37	0.69
Outcome +/- 7.5%	170	0.90	-0.36	0.70
Outcome +/- 10%	160	0.90	-0.36	0.70
Outcome +/- 12.5%	151	0.90	-0.37	0.69
Outcome +/- 15%	146	0.90	-0.38	0.68
Outcome +/- 17.5%	137	0.89	-0.37	0.69
Outcome +/- 20% + glaucoma diagnosis	121	0.89	-0.37	0.69
Outcome +/- 20%	120	0.81	-0.27	0.76
Outcome +/- 22.5 %	108	0.81	-0.31	0.74
Outcome +/- 25%	100	0.81	-0.28	0.75
Outcome +/- 27.5%	91	0.80	-0.35	0.71
Outcome +/- 30%	87	0.80	-0.35	0.71

‡ Outcome effect estimates were compared to statin effect estimates from meta-analyses of randomized controlled trials: a 15% reduction in all-cause mortality (95% CI: 7% to 22%)³⁹ and no effect on hip fracture (OR: 1.03, 95% CI: 0.91 – 1.16)⁴⁰. We further assumed that the effectiveness in our unselected population would be attenuated relative to the RCT efficacy.⁴¹

Table 4

Model-based and bootstrapped standard errors

	Model-based SE	Bootstrapped SE	% underestimation*
HIP FRACTURE			
Unadjusted	0.090	0.091	1%
Age, sex adjusted	0.095	0.094	-1%
Adjusted for PS:			
Non-parsimonious	0.129	0.132	3%
A priori	0.100	0.100	<1%
Step wise	0.128	0.131	2%
Outcome +/- 20%	0.112	0.138	19%
Outcome +/- 30%	0.108	0.139	22%
Outcome p < 0.2	0.125	0.133	6%
Outcome p < 0.1	0.111	0.136	19%
Stepwise or +/- 20%	0.090	0.091	1%
MORTALITY			
Unadjusted	0.040	0.039	-2%
Age, sex adjusted	0.042	0.043	-1%
Adjusted for PS:			
Non-parsimonious	0.056	0.056	-1%
A priori	0.049	0.048	-3%
Stepwise	0.056	0.056	-1%
Outcome +/- 20%	0.049	0.057	14%
Outcome +/- 30%	0.048	0.051	6%
Outcome p < 0.2	0.056	0.057	1%
Outcome p < 0.1	0.056	0.055	-1%
Stepwise or +/- 20%	0.056	0.056	-1%

* % underestimation is calculated as 1 minus the ratio of the model-based SE to the bootstrapped SE.