HHS Public Access

Author manuscript

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2018 July ; 27(7): 771–780. doi:10.1002/pds.4435.

Classifying medical histories in U.S. Medicare beneficiaries using fixed vs. all-available look-back approaches

Mitchell M Conover, MSPH[1], **Til Stürmer, MD, PhD**[1], **Charles Poole, ScD**[1], **Robert J Glynn, ScD, PhD**[2], **Ross J Simpson Jr, MD, PhD, FACC, FACP**[3], **Virginia Pate, MS**[1], and **Michele Jonsson Funk, PhD**[1]

[1]Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A

[2]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, U.S.A

[3]Division of Cardiology, Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A

Abstract

Purpose—Evaluate use of fixed and all-available look-backs to identify eligibility criteria and confounders among Medicare beneficiaries.

Methods—We identified outpatient visits (2007–2012) with recently documented (≤180 days) cardiovascular risk and classified patients according to whether the exposure (statin) was initiated within 14 days. We selected each beneficiary's first eligible visit (in each treatment group) that met criteria during the respective look-backs: continuous enrollment (1 or 3 years for fixed look-back; 180 days for all-available), no cancer history, and no statin claims. We estimated crude and standardized mortality ratio weighted (SMRW) hazard ratios (HR) for the effect of statin initiation on incident 6-month cancer (a known null effect) and 2-year mortality, separately, adjusting for covariates assessed using each look-back.

Results—Analyzing short-term cancer, the estimated HR from the all-available approach (HR=0.90, 95%CI: 0.83, 0.98) was less biased than the 1-year look-back (HR=0.79, 95%CI: 0.73, 0.84), which included beneficiaries with prevalent cancer. The 3-year look-back (HR=1.05, 95%CI: 0.90, 1.21) was somewhat less biased than the all-available estimate but less precise due the exclusion of a large proportion of observations without sufficient continuous enrollment (62.0% and 59.9% of initiators and non-initiators, respectively). All approaches produced similar estimates of the effect on all-cause mortality. Alternative look-backs did not differ in their ability to control confounding.

CORRESPONDING AUTHOR: Mitchell M Conover, MSPH, University of North Carolina at Chapel Hill, Campus Box 7521, Department of Epidemiology / Chapel Hill 27599-7521, Phone: (919) 843-0384, Fax: (919) 843-7364, mconover@email.unc.edu.

PRESENTATIONS: Portions of these results were presented at the 32nd Annual International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE) on August 26, 2016 in Dublin, Ireland.

Conclusions—The all-available look-back performed nearly as well as the 3-year fixed, which produced the least biased point estimate. If 3-year look-backs are infeasible (e.g. due to power/sample), all-available look-backs may be preferable to short (1-year) fixed look-backs.

Keywords

administrative claims; healthcare; bias; classification; confounding variables; database; epidemiologic methods; longitudinal studies

INTRODUCTION

Clinical research is increasingly relying on secondary health data to evaluate the safety and effectiveness of medical therapies in real world populations.[1–3] To ensure comparable accuracy of information across comparator groups, longitudinal studies are routinely restricted to those who are continuously observed within the database for some uniform time period before exposure.[4] Potentially informative data occurring before this time period are discarded.[5] These fixed (or uniform) look-back periods are frequently used to define study eligibility criteria (e.g., no observed history of exposures or outcomes, no recent cardiovascular events) and also to capture baseline covariates used to adjust for confounding.

Selecting a fixed look-back period requires investigators to weigh competing priorities. A longer period allows for a more thorough characterization of database enrollees but also selects narrower, smaller cohorts. In many cases, at least in the US, database enrollment depends on a range of complex variables (e.g. employment, socioeconomic status, marital status / family structure, health status, age). It is unclear whether enrollment restrictions, which inadvertently condition on these characteristics, might impact findings. Despite widespread use of methods that clearly favor the principal of comparative information-accuracy in epidemiology, methodologists have debated its importance relative to other threats to validity, such as covariate misclassification or selection bias, which may be reduced by using all of the available data.[6–10] Observing all historical (pre-exposure) information available in a database while requiring only minimal baseline continuous enrollment has been proposed as a possible compromise which might improve capture of relevant medical history and selection of more inclusive, representative cohorts.[6,7] The common argument against using all-available look-backs is that, for many research questions, we might expect the completeness and longitudinal breadth of available data to vary informatively between exposure (e.g. when comparing users to non-users) or outcome groups, threatening validity of estimates.

To date, there has been limited research exploring the use of all-available data to characterize patient medical histories, primarily using simulations of simplified scenarios.[7,10] Research does exist demonstrating that effect estimates may vary depending on the length of *fixed* look-backs used to exclude (or washout) patients with prior exposures.[11,12] Only one paper has been published exploring use of all-available look-backs in actual data with multiple interrelated covariates but it does not address the issue of cohort selection.[13] Thus, we sought to evaluate the application of multiple look-back approaches to select patients and classify covariates in an observational cohort study set in the Medicare claims

database. In this study, we estimate the effects of statin initiation (compared to non-initiation) after an outpatient office visit on 1) a null outcome (6-month cancer incidence) and 2) a protective outcome (2-year all-cause mortality).

METHODS

Study population

We used a 20% random sample of Medicare fee-for-service beneficiaries with at least 1 month concomitant parts A, B, and D coverage, to identify all outpatient visits observed between 2007 to 2012 when the patient could have received a new statin prescription. For all look-back approaches, we required a minimum of six months of continuous Part A, B, and D enrollment before the potential index visit (see exposure below) and at least one Part D claim within this period. During the six months preceding the index visit, patients were required to have a diagnosis or procedure code indicative of elevated cardiovascular risk and no medications or diagnosis codes indicative of strong contraindications for statin therapy. These eligibility criteria were meant to imitate those of the Heart Protection Study.[14]

We identified three cohorts by applying different look-back periods to the set of potential index visits identified using the 6-month period above. For the all-available database history approach, we required no additional continuous enrollment, but excluded all visits preceded by any observable statin claims or cancer (other than non-melanoma skin cancer) diagnosis/treatment. When applying the conventional one- or 3-year fixed look-back periods, we further restricted the cohort to those continuously enrolled throughout the entire look-back and then excluded visits with prevalent statin use or cancer history within these look-back periods. When beneficiaries had multiple eligible outpatient visits, we selected the first eligible visit within each exposure group (i.e. the first eligible initiation visit and the first eligible non-initiation visit). A study schematic illustrating the overall study design is presented in Fig-1.

Exposure

We classified each index outpatient visit as either a statin initiation or non-initiation by evaluating whether there was a claim for a statin dispensing at a pharmacy in the subsequent 14 days.

Outcomes and follow-up

In separate analyses, we evaluated the effect of statin initiation on two outcomes 1) incident cancer within six months and 2) all-cause mortality within two years. For both, follow-up began on the day after the 14-day exposure assessment window (15 days after the index outpatient visit). Individuals with either outcome during this 14-day window (≈0.4% of visits) were excluded. For both outcomes, we censored follow-up when individuals disenrolled from the study database or the end of available data, December 31, 2012. For the short-term cancer outcome, we also censored follow-up when patients died or switched exposures. Exposure switching was defined as a statin fill for non-initiators and 14 days without medication coverage for initiators.

Covariates

We used the index visit claim to assess information on patient demographics (age, sex, race, geographic region, and calendar year). Then, using the various look-back approaches, we assessed historical claims to classify baseline health behaviors, diagnoses and procedures using CPT, HCPCS and ICD-9 codes associated with Part A and B claims and baseline medication use using NDC codes associated with Part D claims. We described utilization variables as rates (e.g. # outpatient visits per month).

Statistical analyses

Within each cohort, we evaluated covariate imbalance between initiators and non-initiators using the average standardized mean difference[15] and then used multivariable logistic regression to estimate a propensity score (i.e. baseline probability of statin initiation conditional on baseline covariates)[16] corresponding to each index visit in the cohort. Propensity score models included all variables that were identified as risk factors for the outcome using any look-back approach. A more detailed description of the approach to variable selection for the propensity score model is available in eAppendix 1 and the sets of selected variables for each outcome are given in the footnote of Table-1.

In each analysis, we estimated crude and adjusted hazard ratios for the effect of interest using Cox proportional hazards models. We used the robust variance to estimate confidence intervals to account for beneficiaries who entered the cohort twice (for an initiation and non-initiation).[17] We adjusted estimates to account for differences in measured baseline covariates using standardized mortality ratio weighting (SMRW) with and without 1% asymmetric trimming of the propensity score.[18–20] In a sub-analysis of the cancer outcome, we accounted for competing risk of mortality by fitting the Fine and Grey subdistribution hazards model.[21,22] We used the cumulative hazard function to plot cumulative incidence curves estimates of the risk difference (i.e. the difference in cumulative incidence at each point in time) over the course of follow-up.

For the 6-month cancer analysis, we anticipated a null effect, since it is implausible for any statin exposure to have a causal effect on the incidence of clinically-detectable cancer within such a short interval after initiation.[23] While this effect should be null, we expected estimates to be biased by uncontrolled differences in selection, baseline cancer risk and cancer surveillance during follow-up. Thus, we estimated mean squared error (MSE) using the equation: $\text{MSE} = (1 - \log\text{-HR})^2 + (\text{Standard Error}_{\log\text{-HR}})^2$. For the analysis evaluating the effect of statins on mortality, the results of two meta-analyses served as alloyed gold standards.[24,25]

Sub-analyses

Unlike the primary analysis, which applied the same look-back uniformly for all study components (e.g. exclusion of prevalent statin users, assessing confounders for adjustment), we conducted a sub-analysis varying each component individually and holding the others fixed. This allowed us a more granular exploration of the mechanisms through which look-backs might alter findings. We also conducted a sub-analysis with an active comparator, i.e. high-potency statins vs. low-potency statins.

This study was reviewed and approved by University of North Carolina's institutional review board (study: 16–1066). All analyses were conducted in SAS 9.4 (SAS Institute, Cary, NC, USA) and figures were produced using SAS 9.4 or R 3.3.1 (R Foundation for Statistical Computing; Vienna, Austria).

RESULTS

The all-available cohort (71,347 initiators, 476,832 non-initiators) was slightly smaller than the 1-year fixed cohort (86,923 initiators, 559,471 non-initiators) and much larger than the 3-year fixed cohort (18,918 initiators, 204,249 non-initiators) (Table-1). As implemented here, the all-available look-back had a far less restrictive continuous enrollment requirement compared to the 1-year look-back. However, the all-available cohort was smaller than the one year because it excluded more patients with identifiable history of statin use and/or cancer (Fig-2). With respect to the proportions of patients excluded for having prior statin use and cancer history, the all-available approach was less restrictive than the 3-year approach, but much more restrictive than the 1-year approach (Fig-S1). Among non-initiators, cancer incidence during follow-up was elevated in cohorts selected using shorter fixed look-backs (1-year: 2.0% vs. 3-year: 1.5%). Cancer incidence in the all-available cohort most closely resembled that of the 3-year fixed cohort. For all look-backs, the inclusion criteria for recently elevated cardiovascular risk was most frequently met by the presence of either diabetes or stroke.

In the all-available cohort, non-initiators had less available Part A/B history (median: 23 months, IQR: 19–38) compared to initiators (median: 31 months, IQR: 21–47). The same was also true for Part D database enrollment history among non-initiators (median: 20 months, IQR: 14–30) and initiators (median: 27 months, IQR: 18–41). The amount of available database history was nearly identical across levels of both the cancer and mortality outcomes.

In Fig-3 we present the proportion of the cohort with observable history of statin claims (Fig-3a) or cancer (Fig-3b) when all-available data was considered, stratified by the calendar year of the index visit. The corresponding figure for the active comparator sub-analyses is available in Fig-S2. Compared to non-initiators, initiators in the 1-year look-back cohort were more likely to have identifiable history of statin use; however, in the 3-year look-back cohort, the two groups were similar. In the 1-year look-back cohort, 46% and 30% of initiators and non-initiators (respectively) had identifiable baseline statin use when all available data was considered. For both fixed look-back approaches, non-initiators were more likely to have identifiable cancer history than initiators. Misclassification was less frequent in the cohorts selected using longer look-backs. Due to the left-truncation of the Medicare data in calendar time (in 2007), the all-available approach was less informative in earlier calendar years (i.e. since in earlier calendar years less data history was available).

It is important to note that most beneficiaries who entered the study twice entered the study as a non-initiator prior to entering as an initiator. The proportion of initiators who had a dual-entry in the cohort as a non-initiator did not vary widely by look-back approach, ranging from 70% of initiators for the 3-year approach to 75% for the 1-year (Table-S1).

Compared to non-initiators, initiators were younger, used more preventive health services / screening, and were more likely to be diabetic (Table-S2). Broadly speaking, the all-available approach tended to identify greater imbalance in measured covariates compared to fixed look-back approaches, although in most cases not by much (Fig-4). For all look-back approaches, covariates were well balanced (standardized difference <5%) after SMR-weighting. Propensity score distributions under each look-back approach are presented in Fig-S3 and Fig-S4.

In analyses of the 6-month cancer outcome, SMRW-adjusted estimates of the hazard ratio generated using fixed look-backs ranged from 0.79 (95% CI: 0.73–0.84, MSE: 1.54) for the 1-year to 1.05 (95% CI: 0.90–1.21, MSE: 0.92) for the 3-year fixed look-back (Table-1). The SMRW-adjusted HR estimate for the all-available approach (HR: 0.90, 95% CI: 0.83–0.98, MSE: 1.22) was more biased than the 3-year approach but more precise. In the 6-month cancer analysis, SMRW-adjustment had little impact on estimates, especially in the case of the 1-year look-back.

For the outcome of 2-year all-cause mortality, we observed substantial confounding in the crude estimates (Table-1). Crude HR estimates were very similar between the look-backs, spanning from 0.47 to 0.50. Point estimates of the HR were similar for all look-back approaches after applying SMRW adjustment. The adjusted estimate produced by the all-available approach (HR: 0.77, 95% CI: 0.74–0.80) was similar to the estimate produced by the 3-year fixed look-back (HR: 0.82, 95% CI: 0.76–0.88), but was more precise. All results were consistent after 1% asymmetric propensity score trimming (data not shown). Results from the active comparator sub-analyses are presented in Table-S5.

In the sub-analysis independently varying the look-back to define different study components, estimates were generally insensitive to look-back choice (Table 1). An important exception is that in the 6-month cancer analysis, estimates dramatically (and significantly) improved when we excluded patients with prior cancer history using the all-available approach (HR=0.94, 95% CI: 0.84,1.04) or 3-year fixed look-back (HR=0.93, 95% CI: 0.84, 1.03) instead of a short 6-month look-back (HR=0.69, 95% CI: 0.65, 0.74). In the 2-year mortality analysis, estimates were most sensitive to the choice of look-back used to exclude prevalent statin users. Using all-available or longer fixed look-backs moved estimates towards the null and increased the observed mortality in the cohort. Independent variation in the continuous enrollment requirement and assessment of confounders (for adjustment in propensity scores) resulted in negligible movement in estimates.

In Fig-5, we present cumulative estimates of the risk difference over the course of the 6-month follow-up for each look-back approach. (The corresponding cumulative incidence curves are available in Fig-S5, Fig-S6). Risk differences estimated using all-available and 3-year fixed look-backs were generally closer to the presumed truth (null) than the estimates produced using 1-year fixed look-backs. Throughout most of follow-up, the adjusted 3-year look-back estimate is the closest to the true null though, by the end of follow-up, the magnitude of the bias in the all-available estimate was comparable. The results of the short-term cancer analysis accounting for the competing risk of mortality were identical to the primary analysis (data not shown). Fig-6 presents the cumulative risk difference estimates

for the 2-year mortality analysis. Throughout follow-up, estimates produced by the different look-back approaches overlapped one another nearly perfectly.

DISCUSSION

For the effects explored in these analyses, differences in estimates produced using all-available and 3-year fixed look-backs were small, with substantial overlap in confidence intervals (Table-1). Point estimates produced by the 3-year look-back were slightly less biased than the all-available approach, but less precise. In claims studies, bias is typically of greater concern than precision. However, it is still necessary to understand trade-offs in bias and precision, since their relative importance will depend on the specific study question and population. Generally speaking, the all-available approach tracked closely with the 3-year look-back in sub-analyses where we independently varied specific look-back components (holding the others fixed).

Two meta-analyses evaluating the effect of statin use (vs. non-use) on 5-year mortality among elderly patients with established cardiovascular risk estimated risk ratios of 0.85 (95% CI: 0.78, 0.93)[25] and 0.78 (95% CI: 0.65, 0.89).[24] After SMRW-adjustment and trimming, all look-back approaches produced point estimates for 2-year mortality HR that fell in the plausible range between the point estimates for the risk ratios estimated by these meta-analyses. Two randomized double-blinded trials evaluating effects over shorter follow-up (two[26] and three[27] years) produced estimates of 0.76 (95% CI: 0.51, 1.00) and 0.75 (95% CI: 0.49, 0.99), respectively. Trial estimates may provide a reasonable benchmark. However, we cannot use them to assess the bias of the estimates produced in our study since we are evaluating statin effectiveness, not efficacy, in a broader, more heterogeneous population than was evaluated in the trials. Furthermore, given that treatment adherence is likely worse in an observational setting, the plausible range for estimates in our study may be closer to the null than the estimates produced by the trials.

In the analyses we present, there were four key aspects of the cohort that were affected by the look-back period (Table-1 presents results of individually varying each component): the continuous enrollment requirement, exclusion of prevalent statin users, exclusion of patients with a history of the cancer outcome, and assessment of confounders. We discuss the way in which the look-back approaches affected each of these in turn.

Imposing continuous enrollment requirements

We compared statin initiators and non-initiators because it seemed especially plausible that these exposure groups would exhibit striking differences in the accuracy/availability of database information (e.g., as a function of health services utilization and available database history). Indeed, due to our design, we observed less database history among non-initiators, with the median Part A/B look-back being about 8 months shorter among non-initiators. We did not observe meaningful variation in available database history with respect to either the cancer or the mortality outcome. In sub-analyses, independently varying the continuous enrollment requirement had little impact on crude or adjusted effect estimates (Table-1).

Excluding prevalent statin users

Proper exclusion of prevalent statin use is necessary to correctly align time at risk after true initiation. A substantial proportion of cohorts selected using short fixed look-backs had identifiable prior statin use when all available data was considered. Unrecognized prior statin exposure appeared non-differential when using a longer fixed look-back but was more common among initiators when using a short fixed look-back. This may indicate that short fixed look-backs are prone to including prevalent users (e.g. patients paying out-of-pocket, recent/short-term discontinuers). Presumably, these patients were identified and excluded by the longer 3-year look-back. Independently varying the look-back for excluding prevalent statin users produced changes in estimate in the 2-year mortality analysis but not the 6-month cancer analysis (since the true effect in the cancer analysis is null) (Table-1).

Excluding prevalent cancer cases

Considering all-available data, the short 1-year look-back cohort incorrectly included 18% and 23% of initiators and non-initiators (respectively) who had observable cancer history in the database (Fig-3b). A possible explanation for why initiators had less unidentified cancer history might be that they were younger and that approximately 70% of initiators entered the cohorts as non-initiators prior to entering as initiators. It may also be driven by differential surveillance. Initiators were more likely to have undergone cancer and other health screenings. Initiators' superior cancer surveillance *within* the fixed look-back period may reduce the number of unrecognized cancers in the cohort that can be reclassified using data *outside* the look-back period. Failing to properly exclude patients with observable cancer history in the database is more likely to bias estimates of the effect of statins on short-term cancers, where the truth is known to be null. We observed this in the sub-analysis independently varying exclusion for patients with a history of the cancer outcome, producing meaningful improvements in estimates when using longer look-backs (e.g. 3-year or all-available approaches) to exclude these patients (Table-1). This is the most plausible explanation for why the all-available and 3-year fixed analyses of the short-term cancer outcome produced less biased estimates than the 1-year fixed look-back.

Assessment and control for confounding

To informally evaluate the impact of different look-backs on identifying and adjusting for confounding, we can observe change in crude estimates after SMRW adjustment. Unfortunately, in the evaluation of the short-term cancer outcome, the only analysis where we can reasonably estimate bias and MSE, SMRW adjustment had a nearly negligible impact on estimates (Table-1). However, in the mortality analysis, where SMRW adjustment produced large changes in estimates (indicating a more prominent role of measurable confounding), we observed substantial overlap in crude estimates and substantial overlap in adjusted estimates. The fact that the change-in-estimate due to adjustment was similar for all of the look-back approaches indicates that, in this setting, the information obtained from distal database history captured by longer look-backs is of limited use. This finding is consistent with the findings of Nakasian et al. who compared short fixed look-backs to all-available approaches in an analysis of a commercial claims database.[13] In the sub-analysis independently varying the look-back used to assess confounders, the all-available (HR=0.79)

and 3-year (0.80) look-back estimates for 2-year mortality were only slightly lower compared to those produce by shorter fixed look-backs (1-year: HR=0.83).

This study has some important limitations. Since this paper explores an applied example in real-world data, it is difficult to know the truth or evaluate true bias as earlier simulation work has. Single empirical examples have, however, previously been successfully used to compare different study designs.[28] A unique limitation for cancer analysis, where the true effect is null, is that imprecise approaches (e.g. a 3-year fixed look-back) will be more likely to produce correct inference (i.e. confidence intervals containing the null). Also, it is likely that analyses of the short-term cancer outcome remains confounded by variables that we could not measure in the Medicare data. Minimal change in the cancer estimates before and after adjustment indicates a limited ability to control for confounding when using claims data. However, in analyses of the mortality outcome, where SMRW adjustment resulted in substantial changes in estimates, all look-backs produced similar estimates. Furthermore, we selected a population with recently-observed elevated cardiovascular risk in order to assure that everyone would have a plausible indication for statin therapy. However, it is possible that our estimates remain confounded factors that we measure within the claims data, which may lead a physician to withhold statins from an otherwise indicated patient (e.g. frailty). Our design allowed the same patient to enter as both a statin initiator and non-initiator, and the great majority who did entered first as a non-initiator, i.e., with less available look-back. It is unlikely this impacted the relative performance of the different look-backs since the frequency of repeated patients in the cohort did not vary widely by look-back approach. Furthermore, we adjusted estimates using SMRW (which weights to the treated population), preventing us from double-counting patients who were eligible to enter the cohort in both exposure groups, since they can only appear once as an initiator. Finally, determinants of continuous enrollment, and thus performance of different look-back methods, may vary across different study questions, populations, and databases, which may limit the generalizability of our findings.

Further research exploring these approaches is needed. Formal quantitative bias analysis may be a promising method to explore (and/or bound) the impact that differential database history might have on the performance of different look-backs.[29] Our decision to select each beneficiary's first eligible visit may reduce the benefit of using all-available database information and potentially increases differential information accuracy by exposure status. Our motivation for using this approach was to provide a conservative evaluation of all-available look-backs in a potentially problematic setting. However, further research is needed exploring the performance of different look-back approaches when using alternative cohort selection strategies (e.g. randomly sampling across person-time). Our study design and choice of comparators prevented us from doing so here.

Within this applied setting, we contribute evidence that the all-available look-back is a tenable alternative to using long 3-year look-backs, which produced the least biased point estimate, to characterize patients in longitudinal database studies. Both approaches outperformed the widely used 1-year fixed look-back. This indicates that in frequently encountered settings where 3-year fixed look-backs are not feasible (e.g. due to the statistical power required to estimate effects or the structure of the database), the all-

available look-back may be the preferred method. The case for all-available look-backs is made stronger by the fact that the comparability of information accuracy in study groups being compared can be empirically evaluated (e.g. the amount of available baseline data, or the frequency of healthcare interactions), at least to some degree. The look-backs did not appear to vary substantially with respect to their ability to control for confounding. However, selecting a study population using all-available look-backs produced a cohort with less prevalent exposure and cancer reducing bias in analyses where exclusion of patients with prior cancers was essential. By not requiring long periods of continuous enrollment, cohorts selected using the all-available approach were broader and more clearly defined than cohorts selected using fixed look-backs, enhancing the precision of estimates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

SPONSORS: This work was partly funded by R01 HL118255 from the National Heart Lung and Blood Institute and R01 AG023178 and R01 AG056479 by the National Institute on Aging. The database infrastructure used for this project was funded by the Department of Epidemiology, UNC Gillings School of Global Public Health; the Cecil G. Sheps Center for Health Services Research, UNC; the CER Strategic Initiative of UNC's Clinical Translational Science Award (UL1TR001111); and the UNC School of Medicine.

This work was partly funded by R01 HL118255 from the National Heart Lung and Blood Institute and R01 AG023178 and R01 AG056479 by the National Institute on Aging. While conducting this work, MMC received research funding via a student fellowship from Amgen, Inc, as a research assistant on a grant from the National Heart Lung and Blood Institute (NHLBI, R01 HL118255), and via a post-doctoral research fellowship from GlaxoSmithKline.

MJF receives investigator-initiated research funding and support as Principal Investigator from NIH National Heart Lung and Blood Institute (NHLBI, R01 HL118255) and Reagan Udall Foundation; as Co-Investigator from the NIH National Institute on Aging (NIA, R01 AG023178), the NIH National Center for Advancing Translational Sciences (NCATS, 1UL1TR001111), AstraZeneca, and the Patient Centered Outcomes Research Institute (PCORI, 1IP2PI000075). Dr. Jonsson Funk does not accept personal compensation of any kind from any pharmaceutical company, though she receives salary support from the Center for Pharmacoepidemiology in the Department of Epidemiology, Gillings School of Global Public Health (current members: GlaxoSmithKline, UCB BioSciences, Merck, Shire). Dr. Jonsson Funk is a member of the Scientific Steering Committee (SSC) for a post-approval safety study of an unrelated drug class funded by GlaxoSmithKline. All compensation for services provided on the SSC is invoiced by and paid to UNC Chapel Hill.

TS receives investigator-initiated research funding and support as Principal Investigator (R01 AG023178, R01 AG056479) from the National Institute on Aging (NIA), and as Co-Investigator (R01 CA174453, R01 HL118255, R01 MD011680, R21-HD080214), National Institutes of Health (NIH). He also receives salary support as Director of the Comparative Effectiveness Research (CER) Strategic Initiative, NC TraCS Institute, UNC Clinical and Translational Science Award (UL1TR001111) and as Director of the Center for Pharmacoepidemiology (current members: GlaxoSmithKline, UCB BioSciences, Merck, Shire) and research support from pharmaceutical companies (Amgen, AstraZeneca) to the Department of Epidemiology, University of North Carolina at Chapel Hill. Dr. Stürmer does not accept personal compensation of any kind from any pharmaceutical company. He owns stock in Novartis, Roche, BASF, AstraZeneca, and Novo Nordisk.

CP has received salary support from the Center for Pharmacoepidemiology in the Department of Epidemiology, Gillings School of Global Public Health (current members: GlaxoSmithKline, UCB BioSciences, Merck, Shire).

RJG has received research support from AstraZeneca, Novartis, Kowa, and NIH.

RJS is a paid consultant for Merck, Pfizer and Amgen; has given lectures for Merck and Pfizer; and has received research funding from Pfizer, Merck and Amgen.

VP receives investigator-initiated research funding and support as Co-Investigator from the NIH National Institute on Aging (NIA, R01 AG023178), the NIH National Center for Advancing Translational Sciences (NCATS, 1UL1TR001111), the Comparative Effectiveness Research (CER) Strategic Initiative, the NIH National Heart Lung and Blood Institute (NHLBI, R01 HL118255). She also receives research support from pharmaceutical companies (Amgen, AstraZeneca, Merck, Shire), via the Department of Epidemiology, Gillings School of Global Public Health.

The database infrastructure used for this project was funded by the Department of Epidemiology, UNC Gillings School of Global Public Health; the Cecil G. Sheps Center for Health Services Research, UNC; the CER Strategic Initiative of UNC's Clinical Translational Science Award (UL1TR001111); and the UNC School of Medicine.

References

1. Institute P-COR. [Accessed September 26, 2017, 2017] PCORnet: The National Patient-Centered Clinical Research Network. 2017. https://www.pcori.org/research-results/pcornet-national-patient-centered-clinical-research-network
2. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System--a national resource for evidence development. N Engl J Med. 2011; 364(6):498–499. [PubMed: 21226658]
3. Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science--big data rendered fit and functional. N Engl J Med. 2014; 370(23):2165–2167. [PubMed: 24897081]
4. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association : JAMIA. 2007; 14(1):1–9. [PubMed: 17077452]
5. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. 2013
6. Rothman, KJ, Greenland, S, Lash, TL. Case-Control Studies. In: Seigafuse, S, Bierig, L, editors. Modern Epidemiology. 3. Philadelphia (PA): Lippincott Williams & Wilkins; 2008. 121–127.
7. Brunelli SM, Gagne JJ, Huybrechts KF, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. Pharmacoepidemiology and drug safety. 2013; 22(5):542–550. [PubMed: 23526818]
8. Wacholder S. When measurement errors correlate with truth: surprising effects of nondifferential misclassification. Epidemiology (Cambridge, Mass). 1995; 6(2):157–161.
9. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. American Journal of Epidemiology. 1992; 135(9):1019–1028. [PubMed: 1595688]
10. Conover MM, Jonsson Funk M. Uniform vs. all-available look-backs to identify exclusion criteria in observational cohort studies [abstract]. 2015
11. Riis AH, Johansen MB, Jacobsen JB, Brookhart MA, Sturmer T, Stovring H. Short look-back periods in pharmacoepidemiologic studies of new users of antibiotics and asthma medications introduce severe misclassification. Pharmacoepidemiol Drug Saf. 2015
12. Li X, Girman CJ, Ofner S, et al. Sensitivity analysis of methods for active surveillance of acute myocardial infarction using electronic databases. Epidemiology. 2015; 26(1):130–132. [PubMed: 25390030]
13. Nakasian SS, Rassen JA, Franklin JM. Effects of expanding the look-back period to all available data in the assessment of covariates. Pharmacoepidemiol Drug Saf. 2017; 26(8):890–899. [PubMed: 28397352]
14. Heart Protection Study Collaborative G. The effects of cholesterol lowering with simvastatin on cause-specific mortality and on cancer incidence in 20,536 high-risk people: a randomised placebo-controlled trial [ISRCTN48489393. BMC Medicine. 2005; 3
15. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS®. 2012
16. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.

17. Lin DY, Wei L-J. The robust inference for the Cox proportional hazards model. Journal of the American Statistical Association. 1989; 84(408):1074–1078.
18. Brookhart MA, Wyss R, Layton JB, Stürmer T. Propensity score methods for confounding control in nonexperimental research. Circ Cardiovasc Qual Outcomes. 2013; 6(5):604–611. [PubMed: 24021692]
19. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. American Journal of Epidemiology. 2005; 162(3):279–289. [PubMed: 15987725]
20. Patorno E, Grotta A, Bellocco R, Schneeweiss S. Propensity score methodology for confounding control in health care utilization databases. Epidemiology, Biostatistics and Public Health. 2013; 10(3)
21. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association. 1999; 94(446):496–509.
22. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. American Journal of Epidemiology. 2009; 170(2):244–256. [PubMed: 19494242]
23. Setoguchi S, Glynn RJ, Avorn J, Mogun H, Schneeweiss S. Statins and the risk of lung, breast, and colorectal cancer in the elderly. Circulation. 2007; 115(1):27–33. [PubMed: 17179016]
24. Afilalo J, Duque G, Steele R, Jukema JW, de Craen AJ, Eisenberg MJ. Statins for secondary prevention in elderly patients: a hierarchical bayesian meta-analysis. Journal of the American College of Cardiology. 2008; 51(1):37–45. [PubMed: 18174034]
25. Roberts CG, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. The journals of gerontologySeries A, Biological sciences and medical sciences. 2007; 62(8):879–887.
26. Jukema JW, Bruschke AV, van Boven AJ, et al. Effects of lipid lowering by pravastatin on progression and regression of coronary artery disease in symptomatic men with normal to moderately elevated serum cholesterol levels. The Regression Growth Evaluation Statin Study (REGRESS). Circulation. 1995; 91(10):2528–2540. [PubMed: 7743614]
27. Pitt B, Mancini GB, Ellis SG, Rosman HS, Park JS, McGovern ME. Pravastatin limitation of atherosclerosis in the coronary arteries (PLAC I): reduction in atherosclerosis progression and clinical events. PLAC I investigation. J Am Coll Cardiol. 1995; 26(5):1133–1139. [PubMed: 7594023]
28. 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. 2005; 161(9): 891–898. [PubMed: 15840622]
29. Lash, TL, Fox, MP, Fink, AK. Applying Quantitative Bias Analysis to Epidemiologic Data. Springer; New York: 2009.

KEY POINTS

1. Using a 3-year fixed or all-available look-back appears favorable to the widely-used 1-year fixed look-back, especially when exclusion of prior outcomes is necessary.
2. The 3-year fixed look-back produced the least biased point estimate, closely followed by the all-available approach.
3. The continuous enrollment required for the 3-year fixed look-back decreased the sample size substantially (excluding 62% initiators and 59% of initiators), reducing the precision of estimates.
4. The look-back approaches did not differ in their ability to control for confounding.
5. Cohorts selected using all-available look-backs were broader and clearly defined than those selected using short fixed look-backs.

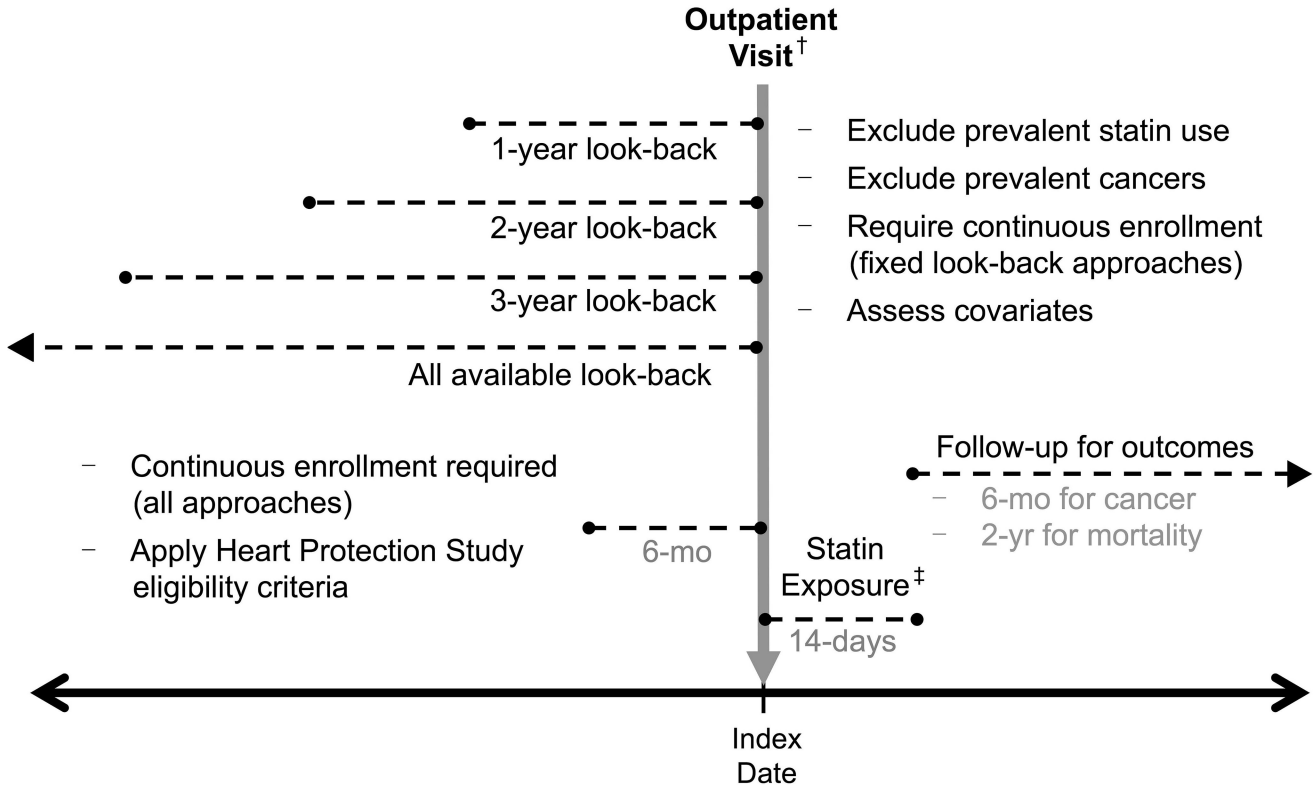

Figure 1.
Study schematic.

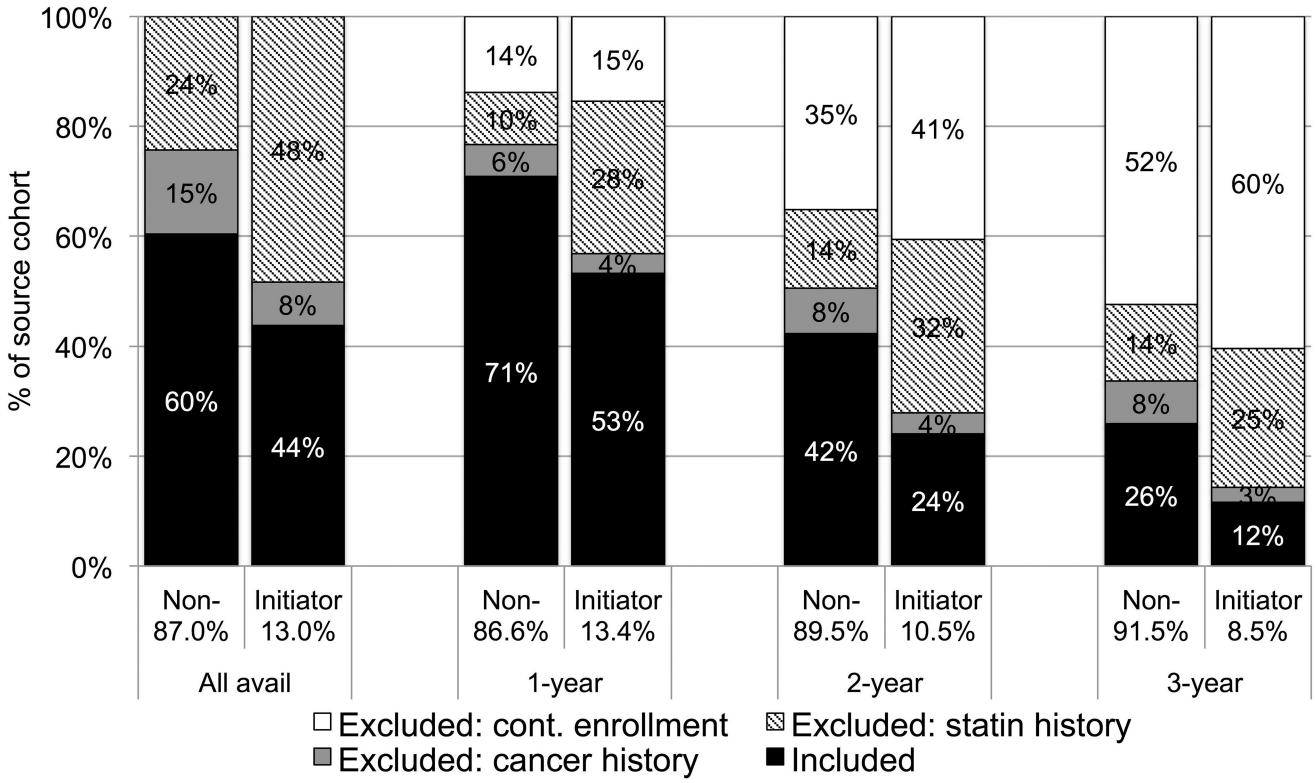

Figure 2.

Bar chart showing the proportion excluded for each of three eligibility criteria applied using different look-back approaches and the final proportion eligible for inclusion.

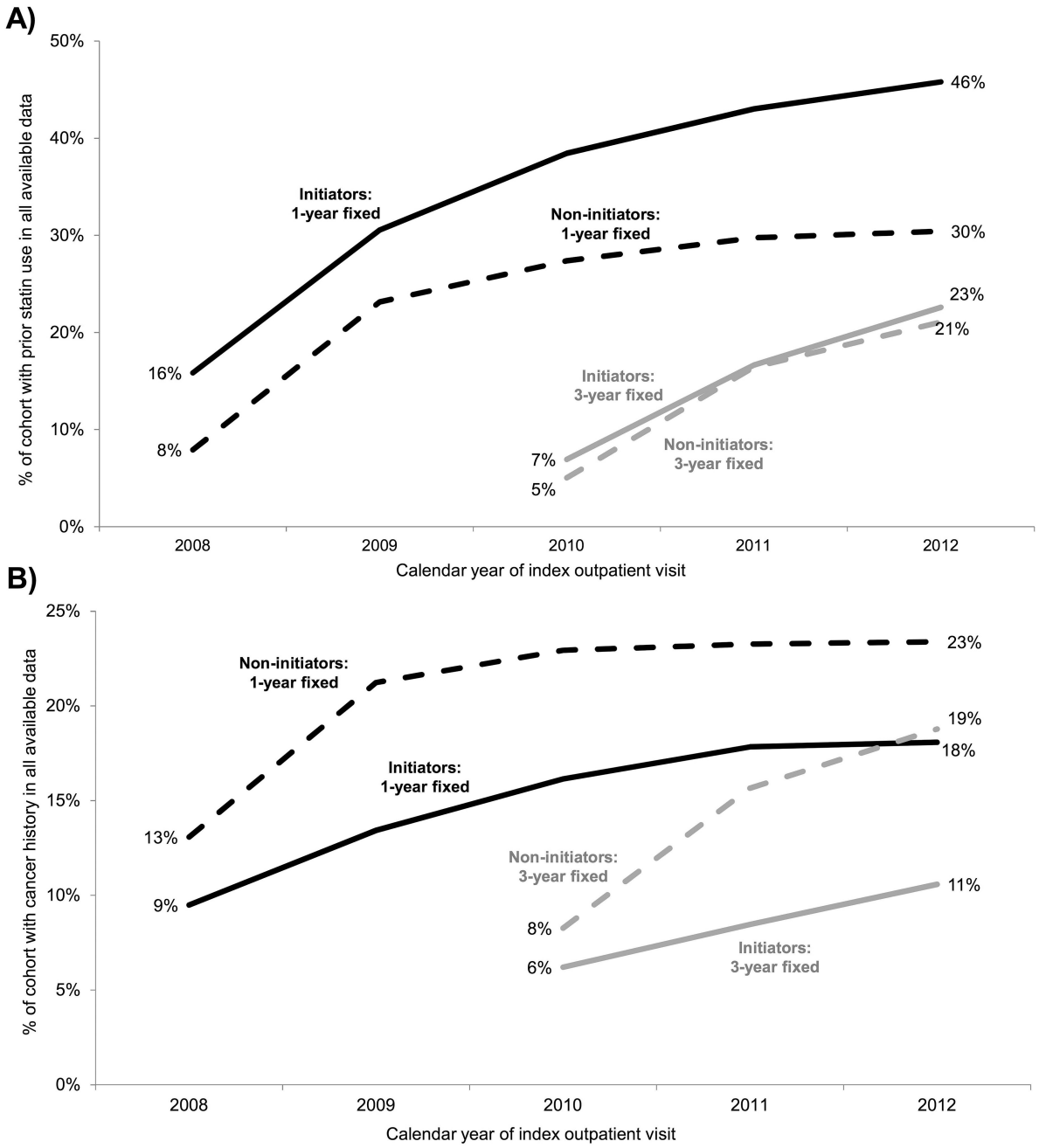

Figure 3.

a/b. Proportion of the final cohort with observable history in the database of A) statin use and B) cancer for the 1-year and 3-year look-back approaches.

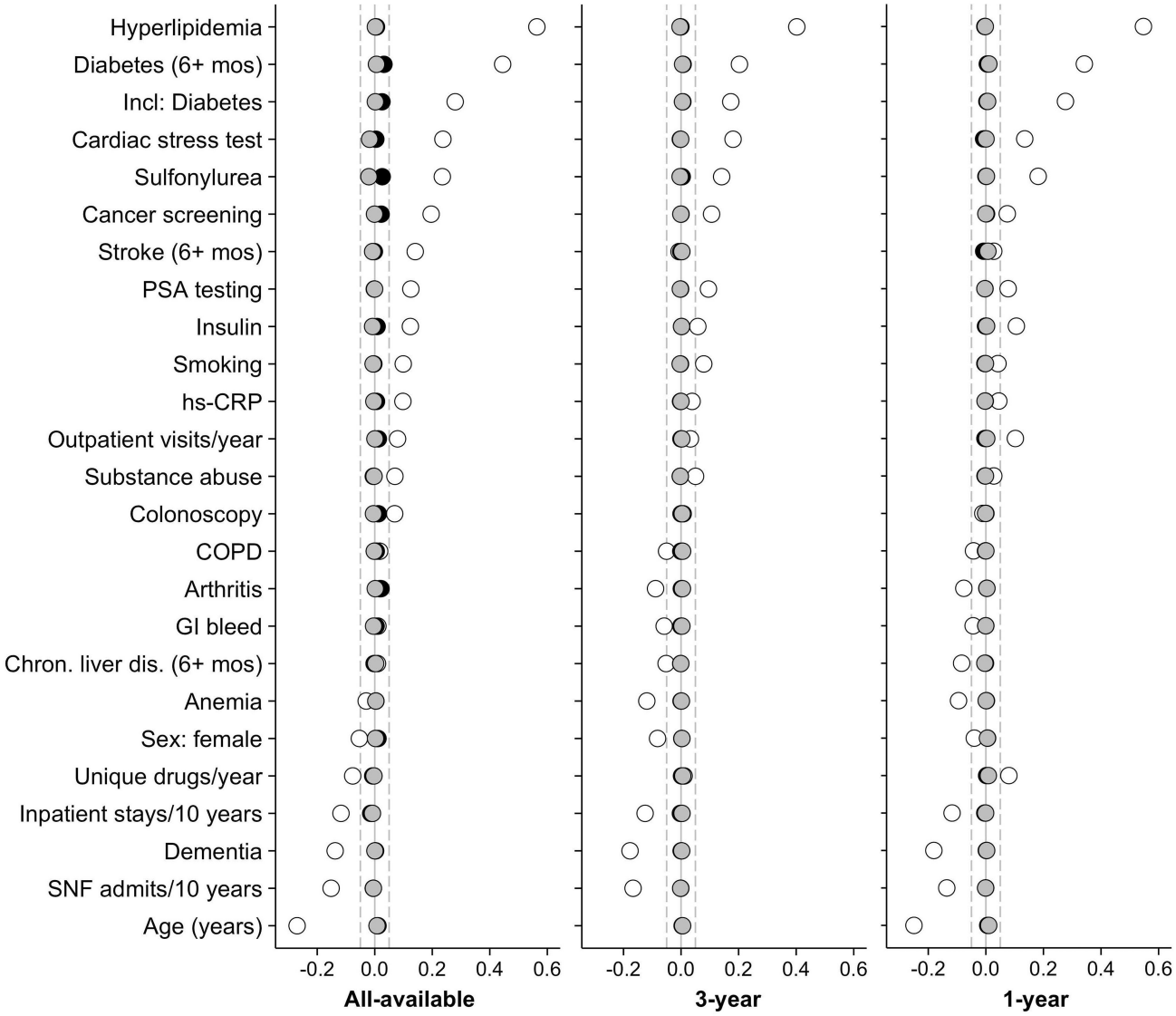

Figure 4.

Average standardized mean difference for selected variables in the analysis of six-month cancer, for the crude (white) analysis and SMRW analysis before (black) and after (grey) 1% asymmetric trimming.

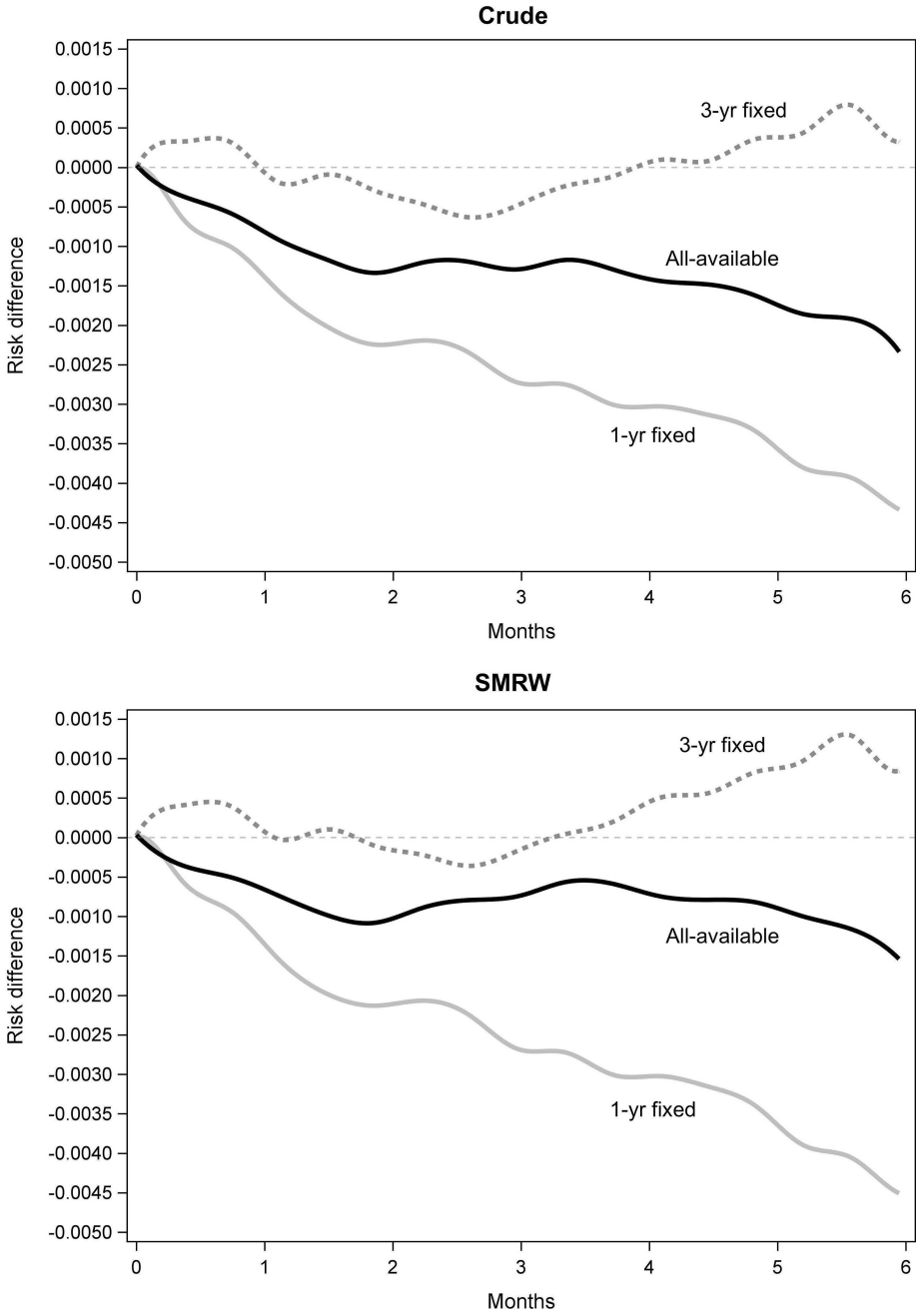

Figure 5.

Crude and SMRW-adjusted cumulative risk differences in the 6-month cancer analysis using the all-available, 3-year, and 1-year look-back approaches.

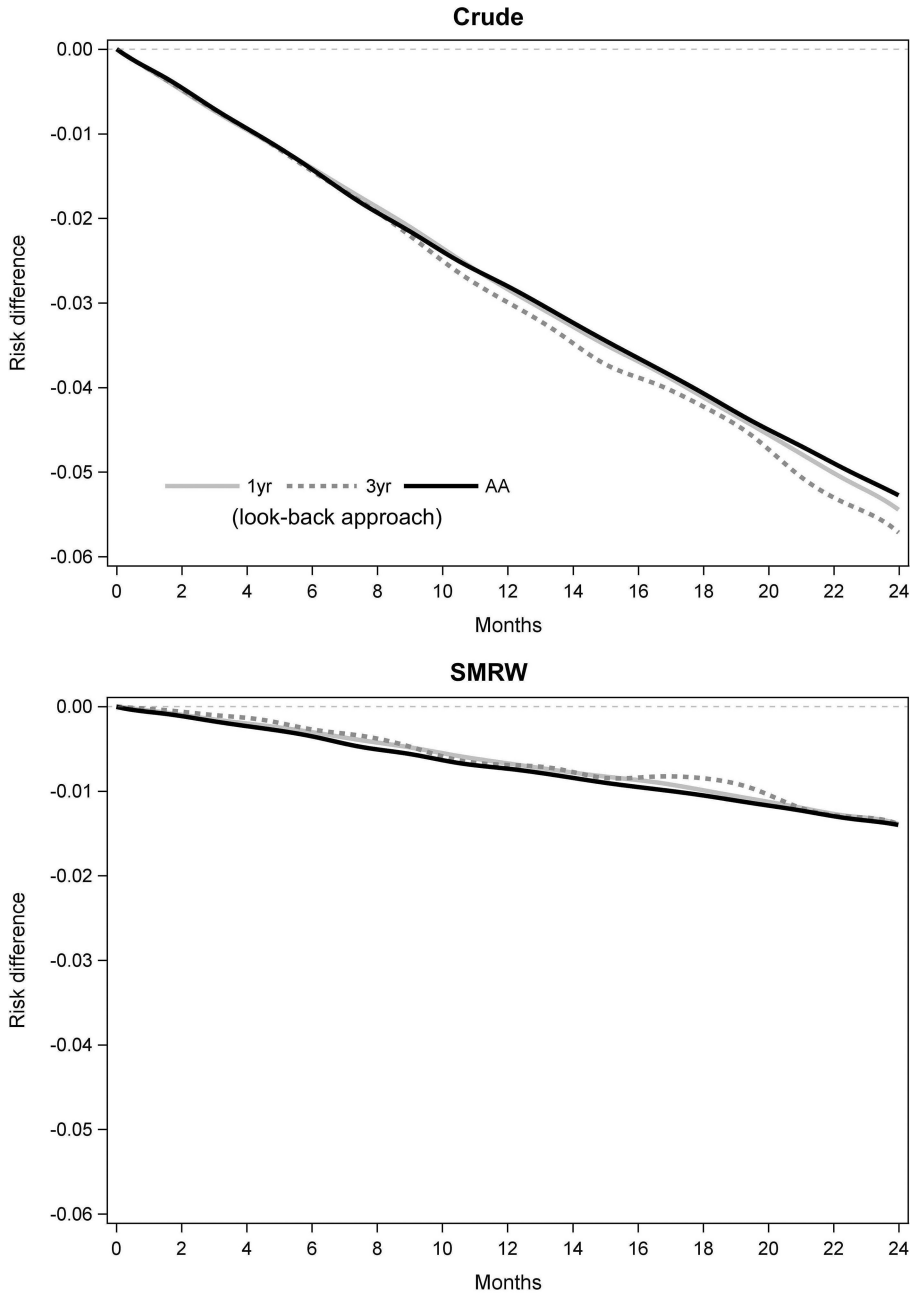

Figure 6.

Crude and SMRW-adjusted cumulative risk differences in the 2-year mortality analysis using the all-available, 3-year and 1-year look-back approaches.

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2019 July 01.

Table 1

Cohort sizes, outcome frequencies, and hazard ratios (crude and SMRW-adjusted) for primary analyses uniformly applying the same window for all look-back components and sub-analyses varying each look-back component individually.

	Look-back parameters				Cohort size (N)		Outcome frequency		Hazard ratio (95% CI)			
	Eligibility criteria			Model					6-month cancer		2-year mortality	
	Cont Enr.	BL statin	Can Hist	PS vars†	N_{Total}‡	N_{Statin}‡	$\%_{Can}$	$\%_{Death}$	Crude	SMRW	Crude	SMRW
Primary results												
1-year fixed	1yr	1yr	1yr	1yr	646,394	86,923	1.8%	8.4%	0.78 (0.73, 0.84)	0.79 (0.73, 0.84)	0.48 (0.47, 0.50)	0.79 (0.76, 0.82)
3-year fixed	3yr	3yr	3yr	3yr	223,167	18,918	1.4%	8.5%	1.00 (0.87, 1.16)	1.05 (0.90, 1.21)	0.50 (0.46, 0.53)	0.82 (0.76, 0.88)
All-available	AA	AA	AA	AA	548,179	71,347	1.5%	8.0%	0.85 (0.79, 0.92)	0.90 (0.83, 0.98)	0.47 (0.45, 0.49)	0.77 (0.74, 0.80)
Varying look-back component												
Continuous enrollment requirement	6mo	6mo	6mo	6mo	952,296	163,184	2.7%	7.8%	0.68 (0.66, 0.71)	0.64 (0.61, 0.67)	0.49 (0.48, 0.50)	0.80 (0.78, 0.82)
	1yr	6mo	6mo	6mo	817,987	137,984	2.7%	7.9%	0.69 (0.66, 0.72)	0.65 (0.62, 0.68)	0.50 (0.48, 0.51)	0.80 (0.78, 0.82)
	3yr	6mo	6mo	6mo	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.69 (0.65, 0.74)	0.51 (0.49, 0.53)	0.82 (0.79, 0.85)
Baseline statin use	3yr	6mo	6mo	6mo	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.69 (0.65, 0.74)	0.51 (0.49, 0.53)	0.82 (0.79, 0.85)
	3yr	1yr	6mo	6mo	372,173	40,687	2.9%	8.3%	0.70 (0.64, 0.76)	0.67 (0.61, 0.72)	0.51 (0.48, 0.53)	0.83 (0.79, 0.87)
	3yr	3yr	6mo	6mo	288,687	23,293	3.1%	8.8%	0.73 (0.66, 0.81)	0.70 (0.63, 0.78)	0.51 (0.48, 0.54)	0.85 (0.79, 0.90)
	3yr	AA	6mo	6mo	255,267	19,779	3.1%	9.1%	0.72 (0.65, 0.80)	0.67 (0.60, 0.75)	0.51 (0.48, 0.54)	0.86 (0.80, 0.92)
Cancer history	3yr	6mo	6mo	6mo	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.69 (0.65, 0.74)	0.51 (0.49, 0.53)	0.82 (0.79, 0.85)
	3yr	6mo	1yr	6mo	404,030	59,987	1.8%	7.8%	0.80 (0.74, 0.86)	0.82 (0.76, 0.89)	0.51 (0.49, 0.53)	0.83 (0.79, 0.86)
	3yr	6mo	3yr	6mo	340,814	51,929	1.3%	7.5%	0.91 (0.83, 1.01)	0.93 (0.84, 1.03)	0.51 (0.48, 0.53)	0.83 (0.79, 0.86)
	3yr	6mo	AA	6mo	300,628	47,042	1.2%	7.5%	0.92 (0.83, 1.02)	0.94 (0.84, 1.04)	0.51 (0.49, 0.54)	0.83 (0.79, 0.87)
Propensity-score variables	3yr	6mo	6mo	6mo	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.69 (0.65, 0.74)	0.51 (0.49, 0.53)	0.82 (0.79, 0.85)
	3yr	6mo	6mo	1yr	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.68 (0.63, 0.72)	0.51 (0.49, 0.53)	0.83 (0.79, 0.86)
	3yr	6mo	6mo	3yr	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.67 (0.63, 0.72)	0.51 (0.49, 0.53)	0.80 (0.77, 0.83)
	3yr	6mo	6mo	AA	440,427	64,604	2.8%	7.8%	0.72 (0.67, 0.77)	0.66 (0.62, 0.71)	0.51 (0.49, 0.53)	0.79 (0.76, 0.83)

† *Variables included in propensity score (PS) models for both the 6-month cancer analysis and the 2-year mortality analysis:* sex, age (as a continuous linear term, continuous squared term, categorical term with 5-year categories), calendar year, race, inpatient stays/month (continuous linear term and categorical term divided by quintile), outpatient visits/month, skilled nursing facility admissions/month, unique drugs/month, smoking, substance abuse, anemia, COPD, dementia, hyperlipidemia, venous thromboembolism, cancer screening, cardiac stress test, colonoscopy, hs-CRP, sulfonylurea, insulin, home oxygen. *Variables only included in PS models for the 6-month cancer analysis:* inclusion for diabetes (≤6-months), diabetes (>6-months), stroke (>6-months), chronic liver disease (>six months), arthritis, rheumatoid arthritis, gastrointestinal bleed, PSA testing, creatinine. *Variables only included in PS models for the 2-year mortality analysis:* inclusion for stroke (≤6-months), chronic kidney disease (> 6-months), obesity, angiography, pulmonary circulation disorders, peripheral vascular disease, osteoarthritis, asthma, atrial fibrillation, psychiatric disorder, inflammatory bowel, paralysis, sepsis, vertigo, lipid panel, echocardiograph, fecal occult blood testing, ARB, diuretics, thiazide, ambulatory life support, weakness, wheelchair.

‡ These counts denote unique observations in the dataset. Patients who enter the cohort twice for eligible initiations and non-initiations are counted twice in the N_{total} statistic (one for each exposure). Since they cannot appear twice in the same exposure group, the N_{statin} statistic denotes counts of unique patients.