Published in final edited form as:
Stat Methods Med Res. 2013 February ; 22(1): 70–96. doi:10.1177/0962280211403603.

Observational data for comparative effectiveness research: an emulation of randomised trials to estimate the effect of statins on primary prevention of coronary heart disease

Goodarz Danaei[1], **Luis A. García Rodríguez**[2], **Oscar Fernández Cantero**[2], **Roger Logan**[1], and **Miguel A. Hernán**[1,3]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, USA

[2]Centro Español de Investigación Farmacoepidemiológica, Madrid, Spain

[3]Harvard-MIT Division of Health Sciences and Technology, Cambridge, USA

Abstract

This article reviews methods to estimate treatment effectiveness research using observational data. The basic idea is using an observational study to emulate a hypothetical randomised trial by comparing initiators vs. non-initiators of treatment. After adjustment for baseline confounders, one can estimate the analogue of the intention-to-treat effect. We also explain two approaches to adjust for imperfect adherence using the per-protocol and as-treated analyses after adjusting for measured time-varying confounding and selection bias using inverse probability weighting of marginal structural models.

As an example, we implemented these methods to estimate the effect of statins for primary prevention of coronary heart disease (CHD) using data from electronic medical records in the United Kingdom. Despite strong confounding by indication, our approach detected a potential benefit of statin therapy. The analogue of the intention-to-treat hazard ratio of CHD was 0.89 (0.73, 1.09) for statin initiators vs. noninitiators. The hazard ratio of CHD was 0.84 (0.54, 1.30) in the per-protocol analysis and 0.79 (0.41, 1.41) in the as-treated analysis for 2-years of use vs. no use. In contrast, a conventional comparison of current users vs. never users of statin therapy resulted in a hazard ratio of 1.31 (1.04, 1.66). We provide a flexible and annotated SAS program to implement the proposed analyses.

Keywords

comparative effectiveness; confounding; intention-to-treat analysis; per-protocol analysis; as-treated analysis; selection bias; inverse-probability weighting

1. Introduction

Developing clinical guidelines requires reliable evidence on the comparative effectiveness of medical interventions. Although randomised clinical trials are usually considered the ‘gold standard’ for this purpose, (1-3) conducting a relevant randomised trial is not always ethical, feasible, or timely. As a result, scientific evidence on comparative effectiveness is often obtained by analyzing observational data, which are increasingly based on electronic medical records and other large health care databases.

Corresponding author: Goodarz Danaei, Harvard School of Public Health, 677 Huntington Avenue, Building 3, Suite 820, Boston MA 02115, gdanaei@hsph.harvard.edu.

Causal inferences from observational data may be questionable because the studied treatments are not randomly allocated, which may result in confounding by indication, and because some commonly used analytic approaches (e.g., comparing current users with non-users) may lead to estimates that are biased or hard to interpret. (4, 5) One approach to solve the latter problem is using the observational data to emulate a hypothetical randomised trial. (6, 7)

Here we describe this approach along with a real world application to estimate the effect of statin therapy on the risk of coronary heart disease (CHD) using electronic medical records from the United Kingdom. Randomised clinical trials have shown that statin therapy reduces the risk of CHD, (8) and observational studies have shown that the magnitude of the confounding by indication is very large (9, 10). The combination of a well known effect and large confounding makes this a good case study to test the limitations of both the observational data and our analytic approach.

The structure of the paper is as follows. In Section 2, we describe the observational database that is used in our example. In Section 3, we discuss a hypothetical randomised trial and how to emulate it using the observational database. In Section 4, we describe methods to estimate the observational analogues of intention-to-treat, per-protocol, and as-treated analyses. In Section 5, we compare these methods with the current versus never users analysis which is commonly used for the analysis of observational data. We conclude by discussing the relative advantages and disadvantages of the various proposed effect estimates. We also provide an example of annotated software to implement these analyses.

2. The observational database: THIN

The Health Improvement Network (THIN) is a database of electronic medical records collected by general practitioners in the United Kingdom. In 2009 THIN covered 6.2 million individuals from about 350 practices. For each patient, demographic and socioeconomic characteristics as well as symptoms, signs and diagnoses, referrals, laboratory test results and some lifestyle information are recorded and updated using standardized methods. Vital status and cause of death data is also recorded by general practitioners based on information provided by coroners. The information recorded in THIN has been used in many research projects, and has been shown to be of sufficient quality for epidemiologic research. (11, 12) The subset of THIN that was available to us at the time of this study included information from January 2000 to December 2006.

We defined the treatment initiation date to be the first date of any statin prescription and calculated the discontinuation date using the number of pills provided in the prescription. We considered statin therapy to be uninterrupted if there was a gap of less than 60 days between two successive prescription periods (using a gap of 30 days did not materially change our results). We defined CHD as either a definite diagnosis of myocardial infarction (MI) or death from CHD. We identified 10,041 potential CHD diagnoses in the database. The validity of a random sample of 500 cases had been confirmed by medical records review in a previous analysis. (13) We reviewed the medical records of 4,126 cases that occurred after the end of the previous study using the same validation criteria and considered 92% of these cases as definite. Only definite cases were included in our analyses.

We imposed an a-priori range of plausible values to the recorded data for body mass index (BMI), systolic blood pressure (SBP), LDL- and HDL- cholesterol and number of visits and referrals in the past 3 months. When the recorded values were outside the plausible range, which occurred for 0.03% of eligible observations, we considered the value to be missing. The plausible ranges were based on both biological knowledge and the observed range of values in the National Health and Nutrition Examination Survey in the U.S. (1999-2002).

(14) If the value of a covariate was missing in a particular month, we carried the last observed value for up to 12 months for SBP, LDL- and HDL- cholesterol levels and for up to 24 months for BMI, alcohol and smoking.

This study was approved by the Institutional Review Board at the Harvard School of Public Health and the Multicenter Research Ethics Committee in the UK. All analyses were conducted using SAS version 9.2 (SAS Institute Inc., Cary, NC, USA).

3. Design of a comparative effectiveness study

In this section we first describe the design of a hypothetical randomised trial to estimate the effect of statin therapy on the risk of CHD, and then how to emulate such randomised trial using our database of electronic medical records. Since statin therapy was not randomly assigned to subjects in the database, we will refer to our observational study as a non-randomised 'trial'. We end this section by reviewing the differences between the hypothetical randomised trial and the observational 'trial'.

3.1 A hypothetical randomised clinical trial

3.1.1 Eligibility criteria and washout period—The trial will enroll participants aged 55 to 84 who have no prior history of CHD, stroke, peripheral vascular disease, heart failure, cancer, schizophrenia or dementia, and no signs of subclinical CHD, as determined by a physical examination and laboratory tests such as exercise electrocardiography. Individuals must not have used statins for at least 2 years prior to enrollment (the washout period). The duration of the enrollment period will be determined based on the required sample size and pace of recruitment.

3.1.2 Treatment groups and follow-up—Each eligible participant will be randomly assigned to either the treatment or the control group. Those in the treatment group will receive a pre-defined daily dose of a selected statin for the duration of the trial. Those in the placebo group will receive matching placebo pills if the trial has a blind design, or no pills (i.e. usual care) if the trial has an open-label design, as is common in "pragmatic" comparative effectiveness trials. CHD diagnoses will be confirmed by a physician who is also unaware of the treatment assignment if the trial is double-blind. Each participant will be followed until the first occurrence of CHD, death, loss to follow-up, or administrative end of the study, whichever occurs first.

3.2 Using observational data to emulate a non-randomised 'trial'

3.2.1 Eligibility criteria and washout period—To mimic the design of the above trial using the observational data, we identified all individuals in the THIN database who were between 55 and 84 years of age in January 2000, and who had at least 2 years of continuous recording in the database (operationalized as at least 2 years of information since the first recorded prescription for any drug), at least one health contact and no prescriptions of statins within those 2 years (i.e., the washout period). We then excluded individuals who, according to the information recorded in the database, had a prior history of CHD, angina, revascularization, stroke, peripheral vascular disease, heart failure, cancer, schizophrenia or dementia, had used digoxin or antipsychotics, or had an incomplete data for all potential confounders before January 2000. A total of 3,178 subjects in THIN met all the eligibility criteria for this January 2000 "trial".

3.2.2 Treatment groups and follow-up—Eligible individuals were classified into two groups: the initiators who initiated statin treatment during the month of January 2000 (treatment group) and the non-initiators who did not initiate statin treatment during that

month (control group). Of the 3,178 subjects in the 'trial', 18 were initiators. Each individual was followed until the first occurrence of CHD, death, loss to follow-up, or administrative end of the study, whichever occurred first. Of the 3,178 subjects, 91 developed CHD during the follow-up (1 among the initiators).

3.2.3 Creating a sequence of non-randomised 'trials'—The small number of initiators (18) and CHD cases among them (1) makes it impossible to conduct a meaningful analysis of our 'trial'. To increase the number of initiators and cases, we applied the above approach to every month between January 2000 and November 2006. Thus we emulated 83 'trials', each of them with a one-month enrollment period. Table A1 in the Appendix shows the numbers of participants, initiators, and CHD events by 'trial'. The eligibility criteria for at least one of these 83 'trials' were met by 74,806 subjects. Of these, 635 had CHD, 1,667 died from other causes, 307 were lost to follow-up (e.g., transferred to a non-THIN practice), and 72,197 reached the end of follow-up (December 2006 or age 85) alive and without CHD.

On average, each of the 74,806 eligible subjects participated in 11 'trials'. For example, many of the non-initiators in the January 2000 'trial' still met all eligibility criteria on February 1st and thus were included in the February 2000 'trial' too. In contrast, all initiators in the January 2000 'trial' were ineligible for the February 2000 'trial' because they received treatment during the washout period for the February 2000 'trial'. When pooling the participants across all 'trials', there were 844,800 person-'trials', of which 13,599 were initiators and 6,335 developed CHD (117 among initiators). The average duration of follow-up was 25.8 months for initiators and 29.6 months for non-initiators. See Figure 1 for a flow-chart of the person-'trials' included in the study.

3.3. Differences between the hypothetical randomised trial and the non-randomised 'trials'

Each one of our non-randomised 'trials' emulates the eligibility criteria, washout period, treatment groups, and follow-up period of a hypothetical randomised trial. However, substantial differences remain between our 'trials' and the hypothetical randomised trial described in Section 3.1.

The most prominent difference is that treatment was not assigned at random in our observational 'trials'. Rather, those with worse prognostic factors (older age, higher LDL- and lower HDL- cholesterol, higher SBP, smoking, hypertension, diabetes, and more medical encounters) were more likely to initiate statin therapy (Table 1), i.e., there is confounding by indication. These differences between initiators and non-initiators are not expected in a randomised trial, and thus no confounding adjustment is usually performed in randomised trials. In contrast, estimating causal effects from our 'trials' requires confounding adjustment, as described in Section 4. We selected the variables in Table 1 as potential confounders based on a priori knowledge and a review of observational studies on the effectiveness of statins. (15)

Another difference is that the hypothetical randomised trial would exclude individuals who have undiagnosed CHD at baseline via a systematic physical exam or laboratory tests. In contrast, many of these individuals will be included in our 'trials' and may have a higher probability of initiating statin therapy if their doctors suspected that they had undiagnosed CHD. Because clinical suspicions may not be recorded in the database, the initiators and the non-initiators in our 'trials' may differ with respect to unmeasured prognostic factors that cannot be adjusted for in the analysis (i.e., there may be unmeasured or residual confounding). We will return to this point in Section 4.5.

Finally, unlike in a hypothetical double-blind randomised trial, in our 'trials' both patients and doctors are aware of treatment assignment. This awareness may affect the behavior of patients (e.g., treated patients may become lenient on following a healthy diet) and doctors (e.g., doctors may monitor treated patients more closely than untreated ones, which might result in earlier diagnosis). Like in open-labeled randomised trials, the effect of treatment initiation in our non-randomised 'trials' will be a combination of the pharmacological effects of statin therapy and of the behavioral changes on CHD risk. In contrast, a double-blinded randomised trial is designed to estimate the pharmacological effects only. The choice between open-labeled and double-blinded designs depends on the question of interest. For instance, if we are interested in the overall effect of treatment in clinical practice, such effect includes the effect of behavioral changes. In that case an open-labeled trial may be more appropriate than a double-blinded trial. Indeed, 'pragmatic' randomised trials with less restrictive eligibility criteria and an open-labeled design are increasingly being advocated as a more reliable source of evidence on comparative effectiveness in clinical settings than double-blinded placebo-controlled randomised trials. (16)

We now describe how to estimate the overall effect of statin therapy on CHD risk using our non-randomised 'trials'.

4. Analysis of a comparative effectiveness study

4.1. Intention-to-treat analysis

In randomised trials the intention-to-treat effect is estimated by comparing the risk (or the incidence rate) of the outcome in the treated group with that of the control group. This comparison is often carried out by fitting a Cox proportional hazards model, with "time since start of follow-up" as the time scale and a non time-varying indicator for treatment assignment as the only covariate. This model estimates an average hazard (rate) ratio over the duration of the follow-up for the treated vs. the control group, which is commonly referred to as the intention-to-treat hazard ratio. Alternatively, one can approximately estimate this hazard ratio via the odds ratio from a pooled logistic model (17) that includes a flexible functional form of the continuous covariate 'time since start of follow-up' (e.g., polynomials, splines).

We estimated an observational analogue of the intention-to-treat hazard ratio in our non-randomised 'trials' by fitting a pooled logistic model like the one described above except that (i) the indicator for experimental treatment assignment was replaced by an indicator for observed therapy initiation, and (ii) potential confounders (measured at the baseline of each 'trial') were added to the model as covariates. We pooled the 844,800 person-'trials' from all 83 'trials' data into a single model rather than fitting a separate model for each 'trial' and then pooling the 83 hazard ratio estimates. The model included "month at the trial's baseline" (taking values from 1 to 83) and month of follow-up in each "trial" and their squared terms as continuous covariates. The covariate values for each person-'trial' were based on the most recently recorded data by the start of that 'trial'. Both the model and the structure of the dataset required to fit it are described in more detail in the Appendix. Because many subjects participated in more than one 'trial', we used a robust variance estimator to estimate conservative 95% confidence intervals (18).

The average hazard ratio (95% CI) of CHD for statin therapy initiation vs. no initiation was 1.37 (1.14, 1.66) when the model did not include any baseline covariates, 1.29 (1.06, 1.56) after adding age (in 5-year age groups) and sex, and 0.89 (0.73, 1.09) after adding all potential confounders in Table 1. If we assume that all important confounders are included in the model, then the hazard ratio of 0.89 can be interpreted as the intention-to-treat average hazard ratio of CHD conditional on the confounders.

To assess the potential heterogeneity of the hazard ratio across 'trials', we fit a separate model that included all of the above covariates plus a product term between the indicator for therapy initiation and the month of the 'trial'. We tested whether the product term was statistically different from 0 (Wald test, chi-square with one degree of freedom) and did not find a strong indication of heterogeneity across 'trials' (P value 0.65).

Another approach to estimating an intention-to-treat hazard ratio is to use propensity score methods. The propensity score is the predicted probability of receiving treatment (i.e., initiating statin therapy at the start of the 'trial') conditional on the confounders. (19) We estimated the propensity score for each person-'trial' by fitting a logistic regression model for the probability of treatment that included the potential confounders as covariates. We then fit a model to the 844,800 person-'trials' with indicators for therapy initiation and for quantiles of the estimated propensity score (20 categories). The average CHD hazard ratio was 0.88 (95% CI 0.72, 1.08) which is almost identical to the hazard ratio of the model with covariates described above.

4.2 Per-protocol analysis

The magnitude of the intention-to-treat effect depends on how closely participants adhere to the treatment assigned at the start of the 'trial'. In our study, 17% of initiators stopped taking statins within 1 year of initiation (30% within 5 years), and 10% of non-initiators started taking statins within 1 year (39% within 5 years) (Figure 2). In a population with a different degree of adherence to treatment ---or one in which different factors reduce adherence---, the intention-to-treat hazard ratio is guaranteed to differ from our estimate of 0.89. Furthermore, in studies comparing therapy to no therapy, imperfect adherence moves the intention-to-treat effect towards the null, which may result in missing small beneficial or hazardous effects.

In randomised trials, a common attempt to adjust for imperfect adherence is restricting the analysis to participants who adhered to the treatment protocol. This is referred to as a per-protocol (or on-treatment) analysis. In one version of the per-protocol analysis, the analysis would be restricted to individuals who adhered to their assigned treatment throughout the entire follow-up (the per protocol population). In another version, the analysis is restricted to individuals who adhered to their assigned treatment at baseline and individuals are censored if/when they deviate from the assigned treatment (the per protocol person-time). In our example, the goal of a per-protocol analysis is to estimate the effect of 'continuous treatment' (20, 21) defined as the CHD risk in initiators if all individuals had stayed on statin therapy compared with the risk in non-initiators if all individuals had refrained from taking statin therapy during the study period.

An analogueous per-protocol analysis can be conducted in observational studies. In our example, we stop following each person-'trial' if and when they deviated from their initial treatment. That is, initiators are censored when they stop statin therapy, and non-initiators when they start. In such artificially censored population, all initiators are continuous users and all non-initiators are never-users. If we now fit the intention-to-treat model to the artificially censored population, the CHD hazard ratio estimated by the model compares continuous treatment vs. no treatment (conditional on the measured covariates at baseline). (21)

After applying this artificial censoring to our 'trials', the average duration of follow-up was 21.6 months for initiators and 25.0 months for non-initiators, The hazard ratio for continuous treatment vs. no treatment was 0.86 (0.59, 1.27). For this hazard ratio to unbiasedly estimate the effect of continuous treatment we need to adjust for both confounding for treatment initiation and time-varying selection bias because the artificial

censoring, which depends on treatment history, may share common causes with the outcome, i.e., may be informative censoring. (22) We attempted to adjust for baseline confounding by including the baseline covariates in the model. Section 4.4 describes how to use inverse probability (IP) weighting to adjust for time-varying selection bias. Note that the need to adjust for both time-varying covariates applies to per protocol analyses of both randomised trials and observational studies.

4.3 As-treated analysis

An 'as-treated' analysis classifies subjects according to the treatment that they received during the follow-up rather than according to the treatment that they initiated at baseline. In randomised trials, an as-treated analysis ignores that the data come from a randomised trial and rather treats them as coming from an observational study. As a result, an as-treated estimate will be confounded if there are time-varying prognostic factors that moved participants to change treatment and these factors are not appropriately adjusted for.

There are many versions of as-treated analyses. In our example, suppose we fit the intention-to-treat model with one modification: we replace the indicator for therapy initiation by a function of the treatment history received. Specifically, we used 'total duration of treatment' for our primary as-treated analysis. For example, if a participants enrolled in a 'trial' as an initiator in month 10 and continued her treatment until month 20, the total duration of treatment would be 1 for month 10; 2 for month 11 and so on, until it reached 11 in month 20. Our model replaced the indicator for treatment initiation by the time-varying covariates 'total duration of treatment' and its squared term, and thus estimated the average hazard ratio of CHD for each additional month of treatment (conditional on the measured baseline covariates).

We used the above dose-response model to estimate the average hazard ratio estimate for 2 years of treatment vs. no treatment. The hazard ratio estimate was 0.73 (95% CI 0.50, 1.06). This estimate may not have a causal interpretation because of residual baseline confounding, because adjustment for time-varying confounding may be necessary (see Section 4.4) and because our choice of function of treatment history ('total duration of treatment') may not accurately represent the dose-response relation between statin treatment and CHD risk. For example, 'total duration of treatment' does not differentiate between different patterns of treatment that result in the same duration. That is, a patient who took the drug for 2 non-consecutive periods of 5 months has the same total duration as a patient who used the drug continuously for 10 months, regardless of how recently treatment was used. If recency and continuity of use are important determinants of the effect of statin therapy, then other functions of treatment history that capture these factors should be used (e.g., duration of use in the last year, or average duration of treatment as defined in the Appendix). (21)

When the goal is estimating the effect of continuous treatment, per-protocol and as-treated analyses have relative advantages and disadvantages. A per-protocol analysis with artificial censoring does not require a correctly specified dose-response function, but censoring may result in a wider 95% confidence interval for the hazard ratio. Conversely, modeling the dose-response function in an as-treated analysis may yield a narrower 95% confidence interval, but it does so at the risk of model misspecification. One way to assess the sensitivity of the effect estimate to potential model misspecification is to explore different dose-response functions that are consistent with a-priori knowledge and available data. For example, as a sensitivity analysis, we fit a separate model that replaced total duration of treatment by average duration of treatment. See Appendix for technical details.

4.4 Inverse probability weighting

We used IP weighting to adjust for the time-varying selection bias introduced by artificial censoring in the per-protocol analysis and the time-varying confounding in the as-treated analysis. Informally, the IP weight for each eligible subject at a given time is the inverse (reciprocal) of her probability of having received the treatment history that she actually received. (23, 24) For example, suppose a subject received statins during the first month of follow-up but not during the second. Then, her IP weight at the second month is the inverse of the product of her probability of receiving treatment in the first month and her probability of not receiving treatment in the second month. See Appendix for details.

By assigning time-varying IP weights to each subject, we effectively create a pseudo-population in which there is no association between the measured confounders and treatment at any time during the follow-up. (23, 24) Thus, IP weighting eliminates bias due to both measured confounding (because treatment is independent of the confounders in the pseudo-population) and selection bias due to artificial censoring (because censoring is a function of treatment). (23, 24)

To compute the time-varying IP weights, we need each subject's probability of receiving treatment at each time conditional on all time-varying confounders. We estimated these probabilities in the THIN dataset by fitting a pooled logistic model for the probability of treatment use (1: yes, 0: no) that included all the variables in Table 1 (both the value at each subject's baseline and the most recently measured value) in addition to baseline calendar year, duration of last treatment episode and its squared term, calendar month of follow-up and its squared term, and months since last measurement of SBP, LDL- and HDL-cholesterol (in 6 categories). Covariate values were carried forward indefinitely until a new measurement was available. We allowed the association between confounders and treatment to vary according to treatment status in the previous month by fitting separate models for those who were on treatment in the previous month (users) and those who were not (nonusers).

Table 2 shows the association between selected time-varying characteristics and the probability of receiving treatment. Among non-users, strong determinants of starting statin therapy were being a male, being 65 to 74 years of age, using antihypertensive drugs, having high LDL-cholesterol, high SBP, diabetes and more doctor visits. Among users, strong determinants of continuing therapy were having low LDL-cholesterol, using antihypertensive drugs, not using other lipid-lowering drugs, and having more doctor visits. The probability of initiating treatment increased in later 'trials'.

The IP weights we have described so far are often referred to as unstabilized weights. Because unstabilized weights generally yield wide confidence intervals for the hazard ratio, one needs to stabilize the weights in most applications. Informally, stabilization is accomplished by multiplying a subject's unstabilized IP weight at each time by his probability of having received his treatment history conditional on baseline (but not time-varying) covariates. We estimated these probabilities by fitting a second pair of logistic models for treatment (one for users and one for non-users) that were identical to those described above except that they did not include any time-varying covariates. We truncated the stabilized weights at a maximum value of 10, which is larger than the 99th percentile of IP weights in both per-protocol and as-treated analyses, to prevent unduly influence by outliers with very large weights. (25) The mean (standard deviation) of the estimated IP weights after truncation was 0.97 (0.47) before censoring non-adherent person-time (as-treated analysis) and 1.02 (0.21) after censoring (per-protocol analysis).

In our analysis of THIN, the IP weighted hazard ratio (95% confidence interval) was 0.84 (0.54, 1.30) in the per-protocol analysis, and 0.79 (0.41, 1.41) for 2 years of treatment in the as-treated analysis. These estimates did not change materially when the weight models included product terms between pairs of a-priori selected variables (i.e., age and sex, age and calendar year, LDL-cholesterol and calendar year). In the absence of unmeasured confounding, the parameters of our weighted pooled logistic model consistently estimate the parameters of a marginal structural model (26, 27) and represent the effect of continuous treatment vs. no treatment (per-protocol) or the effect of 2 years of treatment vs. no treatment (as-treated).

The as-treated effect estimates are generally sensitive to the dose-response function used in the analysis, and thus careful thought needs to be given to this issue. For example, if we had used average duration of treatment instead of total duration of treatment, the IP weighted hazard ratio for 2 years of treatment vs. no treatment would have been 0.88 (0.50, 1.57).

Finally, IP weights can also be estimated to adjust for informative censoring due to loss to follow-up, which may arise in all types of analyses (intention-to-treat, per-protocol, as-treated). We examined the effect of censoring due to loss to follow-up in the THIN database by using IP weights for censoring. As expected, the results were almost identical to those without IP weights for censoring because only 0.4% of subjects in this database were lost to follow-up.

4.5 Unmeasured confounding by undiagnosed disease

All of the analyses presented in the previous Sections assume no unmeasured confounding. This assumption would be incorrect if patients with undiagnosed disease at the start of the 'trials' were more likely to receive treatment due to reasons (e.g., their physician's clinical judgment) that are not recorded in the patient's profile. A proposed *ad hoc* method to adjust for this unmeasured confounding is excluding the early period of follow-up. (10, 28, 29) The length of the excluded follow-up period is chosen to be at least equal to the duration of asymptomatic (latent) period for the disease of interest.

The rational for this exclusion is as follows. Suppose one knows that at least 6 months of treatment are required to affect the risk, i.e., that the Kaplan-Meier curves for initiators and non-initiators of statin therapy overlap ---or, equivalently, that the hazard ratio is exactly 1--- during the first 6 months of follow-up. In this setting, estimating an average hazard ratio greater than 1 during the first 6 months would indicate that there is unmeasured confounding by indication. Thus eliminating the first 6 months from the analysis would prevent this confounding from affecting the average hazard ratio over the rest of the follow-up.

In our 'trials', the average hazard ratio was 1.61 (1.17, 2.21) during the first 6 months, and 1.17 (0.89, 1.53) during the first 12 months. However, randomised trials indicate that the hazard ratio should be 1 because statin therapy does not affect the risk of CHD during the first 6 to 12 months of use. (30) It is therefore likely that our effect estimates are affected by unmeasured confounding during the early follow-up. Figure 3 shows the changes in the overall average hazard ratio after excluding the early follow-up by increments of 3 months up to 24 months. The intention-to-treat hazard ratio was 0.71 (0.53, 0.94) after excluding 12 months and 0.66 (0.45, 0.97) after excluding 24 months. The hazard ratio for continuous treatment in the per-protocol analysis was 0.53 (0.27, 1.02) after excluding 12 months and 0.48 (0.19, 1.20) after excluding 24 months.

Unfortunately, the interpretation of the effect estimates after excluding the early follow-up is not straightforward. For example, consider the downward change in the intention-to-treat hazard ratio from 0.89 to 0.71 when the first 12 months of follow-up were excluded. In

general, one cannot interpret this change as an indication of better adjustment for confounding for at least two reasons.

First, a change in the value of the hazard ratio is arithmetically expected even in the absence of unmeasured confounding because the overall hazard ratio is a weighted average of the time-varying hazard ratios. Thus, if the hazard ratio were exactly 1 during the first few months and less than 1 afterwards, eliminating the early months from the calculation would result in a downward change in the weighted average compared with the overall weighted average Similarly, if the hazard ratios were increasing in time, excluding early follow-up would increase the average hazard ratio.

Second, a downward change may be due to introduced selection bias. This bias will arise when treatment affects outcome during the early follow-up. Consider, for example, the postmenopausal estrogen plus progestin therapy commonly used in the U.S. A large randomised trial showed that this hormone therapy increases the risk of CHD during at least the first two years of use, (6, 31) i.e., the average hazard ratio is greater than 1 during the first two years. Excluding the first few years of follow-up in the trial yielded a hazard ratio of less than 1. This result could be due to either a reversal of the treatment effect over time (from harmful to beneficial) or, more likely, selection bias: the women included in the calculation of the average hazard ratio after excluding the first few years of follow-up are those resilient women who had survived for all those years without developing CHD despite being exposed to hormone therapy. This selected group of women is expected to have a lower CHD incidence than women who were not exposed to hormone therapy. That is, even in the absence of a long-term protective effect of hormone therapy on the absolute risk of CHD, the hazard ratio will be less than 1 after excluding the first few years. (32)

In our statin analysis, one could argue that exclusion of the early follow-up does not introduce selection bias because randomised trials suggest that statin therapy does not affect the risk of CHD during the first 6 to 12 months of use. But we cannot generally expect to have definite evidence on the timing of treatment effects from randomised trials. (If such evidence were available, the benefits of using observational data for comparative effectiveness would be questionable.) In most cases, it will be unclear whether excluding early follow-up decreases the net bias.

4.6 Summary of results

The main results in Sections 4.1 to 4.5 are summarized in Table 3. The intention-to-treat effect after adjustment for all baseline covariates was 0.89 on the average hazard ratio scale. The large difference between the unadjusted (1.37) and adjusted (0.89) hazard ratios indicates the large magnitude of confounding by indication for the initiation of statin therapy. The effect of continuous treatment in the per-protocol analysis was very similar to the intention-to-treat effect whereas the as-treated analysis showed a slightly stronger protective effect (HR of 0.79). To obtain these two estimates, we used two forms of multivariate adjustment: including all baseline covariates in the model for CHD and using time-varying covariates to estimate IP weights. The unweighted estimates were 0.86 in the per-protocol, and 0.73 in the as-treated analyses, which indicates that either the time-varying confounding was of moderate magnitude, or that we had insufficient information on the determinants of treatment across time.

Excluding the first year of follow-up resulted in hazard ratios that were further from the null: 0.71 for intention-to-treat effect, 0.53 for the per-protocol effect and 0.68 for 2 years of treatment in the as-treated analysis. These estimates can be presented as an *ad hoc* sensitivity analysis for unmeasured confounding, with the caveats discussed in Section 4.5. A more formal --- but also harder to implement--- approach would be to conduct a g-estimation-

based analysis that incorporates the assumptions about the timing of treatment effects into a structural model. (33)

5. The ‘current vs. never users’ comparison

A common version of as-treated analysis in observational studies is the comparison of current users vs. non-users or never users of a particular treatment. In this contrast the analyst simply compares the incidence of the outcome in those who are using the treatment (the current users) with that in individuals who are not using the treatment (the non-users) at a particular time, after adjusting for time-varying covariates in a multivariate survival model. Often the non-users are separated into two groups: those who had never used treatment (the never users) and those who have used it in the past but are not currently using it (the past users). The current-user vs. non-user or never user average hazard ratio is the main result reported in many pharmacoepidemiologic studies.

Of the 2,479,873 person-months in our study (74,806 unique individuals times an average follow-up of 33.2 months), 475,489 were classified as current users and 1,953,878 as never users. Table 4 shows the ‘current vs. never users’ average hazard ratio in our study using different methods for handling missing data and exclusion criteria. The multivariate adjusted hazard ratio of CHD for ‘current vs. never users’ of statins ranged between 1.31 and 1.42. These estimates might be naïvely interpreted as indicating a harmful effect of statins. However, the ‘current vs. never users’ average hazard ratio may be surprisingly difficult to interpret because of at least two reasons.

First, the ‘current vs. never users’ analysis may not appropriately adjust for measured confounders that are affected by prior treatment (e.g., LDL-cholesterol) (5). In our study, we have shown that there is much measured confounding by indication for statin therapy. However, only a small difference between the unadjusted (1.65) and the fully-adjusted (1.42) hazard ratios was found in the ‘current vs. never users’ analyses. This small difference suggests that either this attempt to adjust for confounding is unsuccessful or that it introduces selection bias. See Appendix for a more detailed description of this form of selection bias.

Second, as discussed in Section 4.5, the person-time closer to the start of statin therapy has on average a higher hazard ratio because the hazard ratio is greater than 1 during the early follow-up, possibly due to unmeasured confounding. One could try to eliminate this unmeasured confounding by excluding all early follow-up (see Section 4.5.) or by excluding the early follow-up among current users only. A common implementation of the latter approach is replacing “current users” by “long-term current users” (also referred to as “persistent users”). For example, one could compare the risk among (persistent) users who have received treatment for at least 1 year with that among those who never received treatment.

Under this definition of persistent use, Table 4 shows that the fully-adjusted hazard ratio of CHD for ‘long-term current vs. never users’ comparison ranged between 0.95 and 1.05 in our study, compared with 1.31 to 1.42 for ‘current vs. never users’. The ‘improvement’ from 1.42 to 1.05 (or from 1.31 to 0.95) may be incorrectly interpreted as better adjustment for confounding in the ‘long-term current users vs. never users’ contrast compared with the ‘current vs. never users’. However, this explanation is not fully satisfactory because the age- and sex- adjusted hazard ratios for ‘long-term current vs. never users’ were in fact more ‘improved’ (ranged between 0.88 and 1.01) than the fully-adjusted hazard ratios. That is, additional adjustment for measured confounders made it harder to find an apparently beneficial effect of statins. This counterintuitive result may be explained by selection bias because the persistent users are, by definition, selected initiators who survived a year

without CHD. This form of selection bias is partly corrected by including more variables in the model.

Therefore, using contrasts based on current users is a gamble in which the investigators bet that the amount of selection bias introduced is less than the amount of confounding eliminated. It is especially problematic that one can vary the period to define persistent use, and thus the magnitude of selection bias, to obtain a desired value for the hazard ratio. For example, in our study the hazard ratio of CHD was 0.77 (0.51, 1.18) when comparing long-term current users for at least 24 months vs. never users (compared with 1.05 for current users of 1 year or more in the third column of Table 4). Note than the presence of a truly time-varying effect (rather than selection bias) is always a potential alternative, or complementary, explanation that cannot be ruled out empirically.

Without further information on how the true effects change over time, it is impossible to know whether a 'current vs. never users' hazard ratio is an appropriate method to adjust for imperfect adherence. For example, in re-analyses of observational studies of postmenopausal hormone therapy (6, 34) the 'initiators vs. non-initiators' hazard ratio of CHD did not suggest that treatment was beneficial even though a 'long-term current vs. never users' did. The latter was effectively excluding the early follow-up when hormone therapy increased the risk of CHD and restricted analysis to selected users who had survived without CHD.

Besides not being subject to the above problems, the analytic approach proposed in this paper can be easily extended to estimate standardized survival curves that compare the cumulative risks under different treatment scenarios. (21)

6. Discussion

Using observational data from electronic medical records, we estimated a beneficial effect of statin therapy on CHD risk using intention-to-treat, per-protocol, and as-treated analyses. This was a surprising finding. Given the large magnitude of confounding by indication for the effect of statin therapy, we expected an observational analysis to be hopeless, and the hazard ratio estimate to be greater than 1. This paper was supposed to demonstrate the limitations of observational data to estimate the intended effects of a medical intervention (as opposed to unintended effects such as toxicity). (35) Rather, this paper demonstrated that the combination of subject-matter knowledge (to choose the potential confounders), high-quality data (electronic medical records with careful validation of the outcome of interest), and a sound methodology (emulation of 'trials') may overcome much confounding and selection bias. An observational analysis of the comparative effectiveness of two drugs (rather than comparing a drug with no drugs, as we did) will generally be even less affected by confounding, especially if the two drugs are prescribed to patients with relatively similar baseline characteristics. (36, 37)

Our success, however, was not complete. Our estimate of the intention-to-treat hazard ratio of CHD for statin therapy (0.89, 95% CI: 0.73, 1.09) had a wide 95% confidence interval and underestimated the benefit suggested by randomised clinical trials (0.72, 95% CI: 0.65, 0.79). (8) This underestimation may be explained by residual confounding by indication (35) and higher adherence to treatment in the randomised trials compared with our observational 'trials'. In fact, our estimates of the effect of continuous treatment, which adjust for imperfect adherence, suggested a greater benefit than the intention-to-treat estimate.

Two previous observational studies found a beneficial effect of statin therapy initiation for primary prevention of CHD. (9, 10) Smeeth and colleagues used the THIN database and reported the effect of statin initiation on risk of myocardial infarction only after excluding

the first year of follow-up. The reported hazard ratio was 0.86 (0.76, 0.97). (10) Seeger and colleagues used a database from a community Health Management Organization in the U.S. and conducted their analysis using propensity score matching. They excluded 30% of initiators who could not be matched with a non-initiator and estimated an average hazard ratio of 0.69 (0.52, 0.93). (9) Like our study, both of these studies compared initiators vs. non-initiators. As showed above, the use of the 'current vs. never users' comparison would have incorrectly suggested that statins have a null or harmful effect.

We restricted the study population to individuals without a history of CHD and other chronic diseases to mimic the eligibility criteria of a primary prevention randomised trial. Restriction may also result in a more homogeneous population and thus limit confounding in observational analyses of pharmacological interventions. (38, 39) We further restricted our analysis to individuals with complete information on all potential confounders (after carrying forward the last observed values for 12 or 24 months). The age-sex adjusted intention-to-treat hazard ratio before applying the restriction to individuals with non-missing values was 1.19 (95% CI 1.08, 1.30) compared with 1.29 (1.06, 1.56) after restriction, which suggests that restriction did not contribute much to limit confounding in this context.

Besides residual confounding, another concern in observational studies is that there may not be both initiators and non-initiators for each combination of values of the measured confounders, i.e., the positivity condition may not hold. (25) In the THIN data positivity approximately holds as shown by the extent of the overlap between the distribution of the estimated propensity scores for initiators and non-initiators (Figure 4). To guarantee positivity, we also conducted a matched analysis where we matched each initiator person-'trial' with up to 5 non-initiators in 0.01 calipers of their propensity score (scaled from 1 to 100). This process resulted in a matched population of 75,540 person-'trials' (36,634 distinct individuals). We then fit a Cox model to the matched population of person-'trials' with the initiation indicator as the only covariate. The CHD hazard ratio estimate was 0.83 (95% CI 0.67, 1.03) which is almost equal to our intention-to-treat hazard ratio estimated using either multivariate or propensity score adjustment.

Our approach can be easily modified to estimate the effect of discontinuing a treatment (see Appendix of reference 6). The study population would be restricted to those who have been on treatment for a pre-specified period of time. Then stoppers and non-stoppers in each 'trial' would be identified. In the THIN database, we defined eligible individuals as those who have been on treatment for at least 6 months and we identified 3,561 stoppers and 202,749 non-stoppers. The intention-to-treat (or rather 'intention-to-stop') hazard ratio was 1.49 (0.96, 2.33) before adjustment for baseline covariates and 1.30 (0.83, 2.03) after adjustment. This hazard ratio suggests that cessation of statin therapy increases the risk of CHD.

Another analytic method that might be have been used to adjust for confounding by indication is instrumental variable (IV) estimation. (20, 40) If an instrument is available and certain assumptions hold, one can adjust for both measured and unmeasured confounding. In double-blind randomised trials, the randomisation indicator can be used as an instrument. In observational studies, however, it is not possible to demonstrate that any particular variable is an instrument. Further, the assumptions required for valid IV estimation of the effect of continuous treatment are often implausible. In our example, we were not able to identify any potential instrument in the database.

Finally, we designed our per-protocol analysis to estimate the effect of continuous treatment. However, this would be a poor choice of effect measure if the treatment had frequent toxic effects, i.e., there is little interest in estimating the effect of continuous treatment if many

subjects will not be able to take treatment continuously because of toxicity. In settings with toxicity, the protocol of the randomised trial would specify that treatment has to be discontinued in case of toxicity. We have described how to conduct this per-protocol analysis using observational data elsewhere. (21)

In summary, we showed that the analysis of observational studies for comparative effectiveness research requires background knowledge, high-quality information and appropriate analytical methods. We propose, as a first step, to emulate the observational analogue of the intention-to-treat effect of a (hypothetical) randomised trial, comparing initiators and non-initiators after adjusting for as many potential confounders as possible. As a second step, appropriately adjusted per-protocol and as-treated analyses can be conducted to remove the impact of imperfect adherence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Tyler VanderWeele for his valuable comments on the discussion regarding the direction of bias in the Appendix.

References

1. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr. et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA. 2003; 289(19):2560–72. [PubMed: 12748199]
2. Detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). National Institute of Heart Lung and Blood; Bethesda: 2002.
3. Statins for the prevention of cardiovascular events. National Institute for Health and Clinical Excellence; London: 2006.
4. Gibbons RJ, Gardner TJ, Anderson JL, Goldstein LB, Meltzer N, Weintraub WS, et al. The American Heart Association's principles for comparative effectiveness research: a policy statement from the American Heart Association. Circulation. Jun 9; 2009 119(22):2955–62. [PubMed: 19433753]
5. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. Nov 1; 2003 158(9):915–20. [PubMed: 14585769]
6. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomised experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008; 19(6):766–79. [PubMed: 18854702]
7. Hernán MA, Robins JM. Authors' response, part I: Observational studies analyzed like randomised experiments: best of both worlds. Epidemiology. 2008; 19(6):789–92.
8. Taylor F, Ward K, Moore TH, Burke M, Davey Smith G, Casas JP, et al. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev. 2011; 1:CD004816. [PubMed: 21249663]
9. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. Pharmacoepidemiol Drug Saf. Jul; 2005 14(7):465–76. [PubMed: 15651087]
10. Smeeth L, Douglas I, Hall AJ, Hubbard R, Evans S, Seeger JD, et al. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomised trials, an application of propensity score matching using claims data. Br J Clin Pharmacol. Jan-Jul;2009 67(1):99–109. [PubMed: 19006546]
11. Jick H, Zornberg GL, Jick SS, Seshadri S, Drachman DA. Statins and the risk of dementia. Lancet. 2000; 356(9242):1627–31. [PubMed: 11089820]

12. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. Apr; 2007 16(4):393–401. [PubMed: 17066486]
13. Garcia Rodriguez LA, Tacconelli S, Patrignani P. Role of dose potency in the prediction of risk of myocardial infarction associated with nonsteroidal anti-inflammatory drugs in the general population. J Am Coll Cardiol. Nov 11; 2008 52(20):1628–36. [PubMed: 18992652]
14. National Health and Nutrition Examination Survey data sets and related documentation (Survey Questionnaire, Examination and Laboratory Protocols, 1988-1994 and 1999-2002): Centers for Disease Control and Prevention; 2002.
15. Danaei, G.; Tavakkoli, M.; Hernan, MA. The case of statins. 2010. Can we trust observational studies for comparative effectiveness research?. submitted
16. Mullins CD, Whicher D, Reese ES, Tunis S. Generating evidence for comparative effectiveness research using more pragmatic randomised controlled trials. Pharmacoeconomics. Oct 1; 2010 28(10):969–76. [PubMed: 20831305]
17. Thompson WA Jr. On the treatment of grouped observations in life studies. Biometrics. Sep; 1977 33(3):463–70. [PubMed: 911970]
18. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. Mar; 1986 42(1):121–30. [PubMed: 3719049]
19. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomised control group. Stat Med. Oct 15; 1998 17(19):2265–81. [PubMed: 9802183]
20. Toh S, Hernán MA. Causal Inference from Longitudinal Studies with Baseline Randomisation. The International Journal of Biostatistics. 2008; 4(1) Article 22.
21. Toh S, Hernandez-Diaz S, Logan R, Robins JM, Hernan MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomisation. Epidemiology. Jul; 2010 21(4):528–39. [PubMed: 20526200]
22. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. Sep; 2004 15(5):615–25. [PubMed: 15308962]
23. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. Jul; 2006 60(7):578–86. [PubMed: 16790829]
24. Robins, JM.; Hernan, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Monenberghs, G., editors. Longitudinal data analysis. Chapman & Hall/CRC; Boca Raton: 2009.
25. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. Sep 15; 2008 168(6):656–64. [PubMed: 18682488]
26. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. Sep; 2000 11(5):550–60. [PubMed: 10955408]
27. Hernan MA, Brumback B, Robins JM. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomised Treatments. Journal of the American Statistical Association. 2001; 96(454):440–8.
28. Jee SH, Sull JW, Park J, Lee SY, Ohrr H, Guallar E, et al. Body-mass index and mortality in Korean men and women. NEnglJMed. 2006; 355(8):779–87.
29. Allison DB, Faith MS, Heo M, Townsend-Butterworth D, Williamson DF. Meta-analysis of the effect of excluding early deaths on the estimated relationship between body mass index and mortality. ObesRes. 1999; 7(4):342–54.
30. Simes J, Furberg CD, Braunwald E, Davis BR, Ford I, Tonkin A, et al. Effects of pravastatin on mortality in patients with and without coronary heart disease across a broad range of cholesterol levels. The Prospective Pravastatin Pooling project. EurHeart J. 2002; 23(3):207–15.
31. Manson JE, Hsia J, Johnson KC, Rossouw JE, Assaf AR, Lasser NL, et al. Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. Aug 7; 2003 349(6):523–34. [PubMed: 12904517]
32. Hernán MA. The hazards of hazard ratios. Epidemiology. Jan; 2010 21(1):13–5. [PubMed: 20010207]
33. Robins JM. Causal models for estimating the effects of weight gain on mortality. Int J Obes (Lond). Aug; 2008 32(Suppl 3):S15–41. [PubMed: 18695650]

34. Hernán MA, Robins JM. Discussion on “Statistical Issues Arising in the Women’s Health Initiative”. Biometrics. 2005; 61(4):922–30.

35. Miettinen OS. The need for randomisation in the study of intended effects. Stat Med. Apr-Jun;1983 2(2):267–71. [PubMed: 6648141]

36. Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, Ali F, Sholley C, Worrall C, et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. JAMA. Jul 28; 2010 304(4):411–8. [PubMed: 20584880]

37. Haag MD, Bos MJ, Hofman A, Koudstaal PJ, Breteler MM, Stricker BH. Cyclooxygenase selectivity of nonsteroidal anti-inflammatory drugs and risk of stroke. Arch Intern Med. Jun 9; 2008 168(11):1219–24. [PubMed: 18541831]

38. Schneeweiss S, Patrick AR, Sturmer T, Brookhart MA, Avorn J, Maclure M, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomised trial results. Med Care. Oct; 2007 45(10 Supl 2):S131–42. [PubMed: 17909372]

39. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. JAMA. Aug 25; 304(8):897–8. [PubMed: 20736474]

40. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist’s dream? Epidemiology. Jul; 2006 17(4):360–72. [PubMed: 16755261]

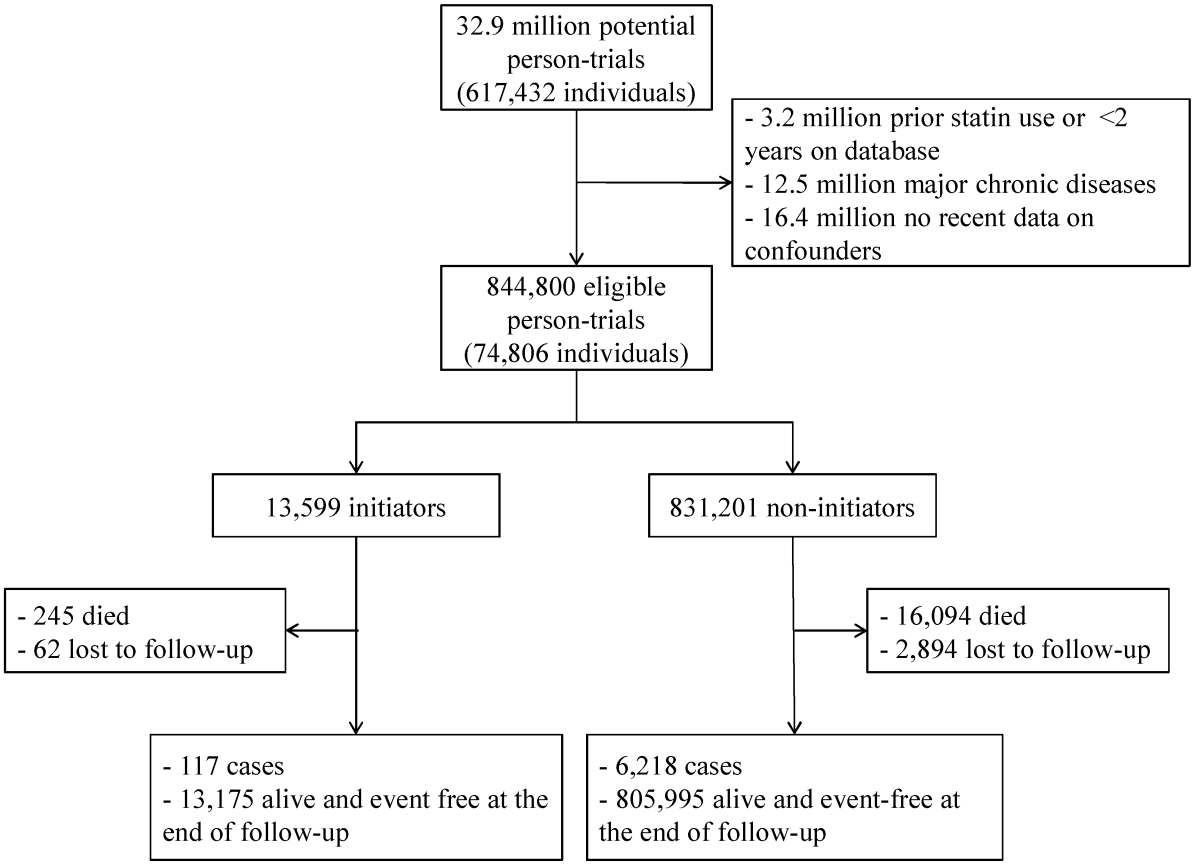

Figure 1.
Flowchart of person-trials in the analysis, THIN 'trials' 2000-2006

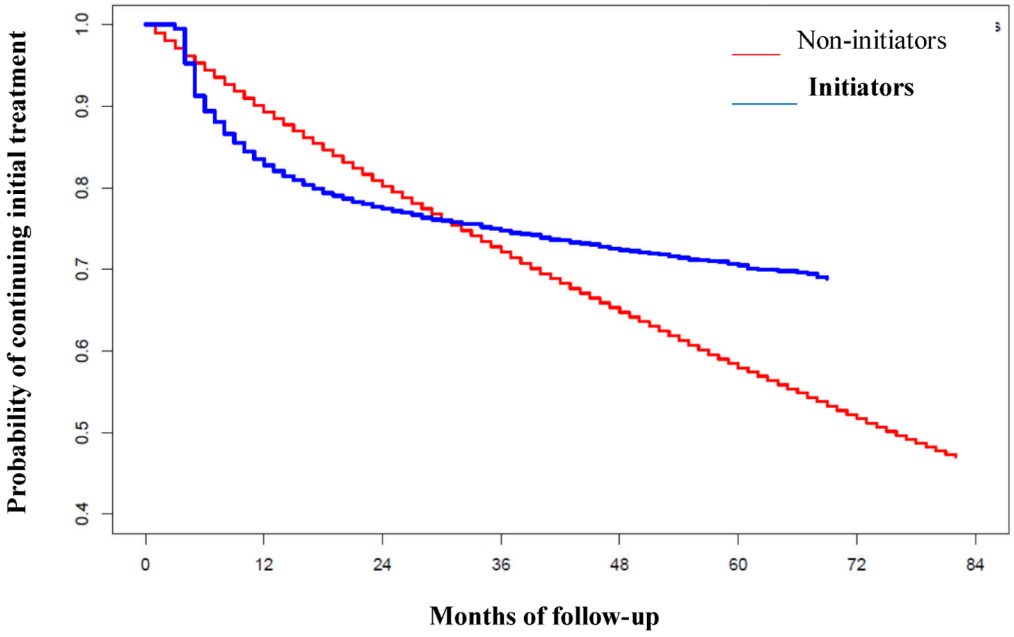

Figure 2.
Adherence to statin treatment by initiation status, THIN 'trials' 2000-2006

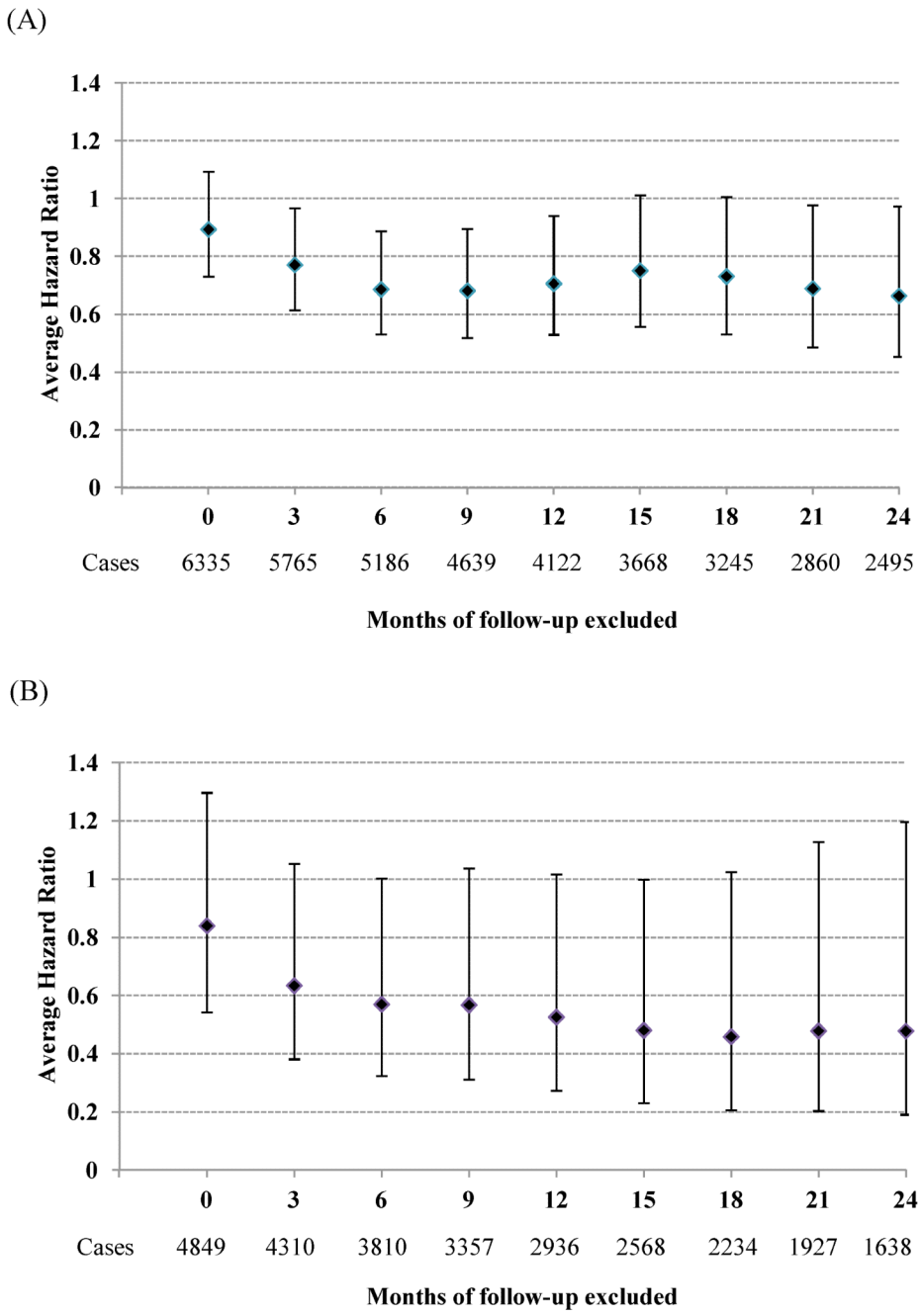

Figure 3.

Impact of excluding varying periods of early follow-up (from 0 to 24 months) on the intention-to-treat hazard ratio (A), and the hazard ratio for continuous use vs. no use (B), THIN 'trials' 2000-2006 [a]

[a]Vertical bars represent 95% confidence intervals. Early follow-up was excluded from the outcome model but not from the model used to estimate the IP weights.

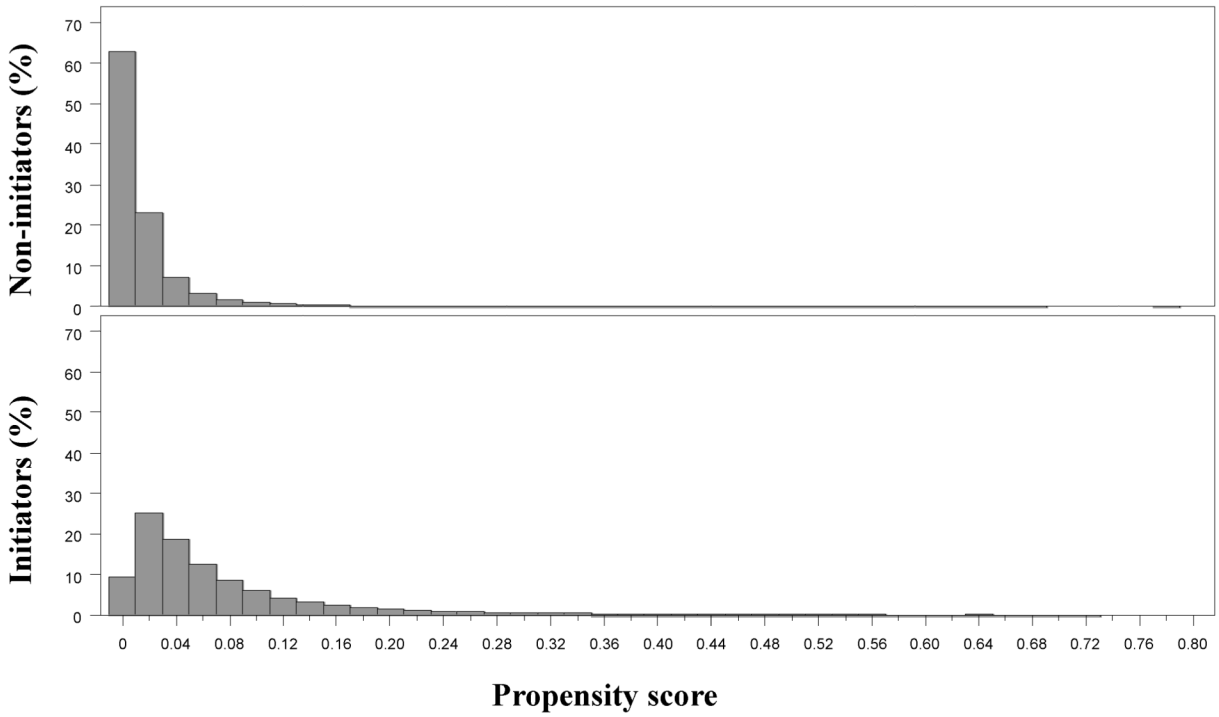

Figure 4.
Distribution of estimated propensity for treatment initiation by treatment status at the start of the THIN 'trials' 2000-2006

Table 1

Characteristics of initiators and noninitiators of statin therapy at the start of the trial's follow-up, THIN 'trials' 2000-2006

Characteristic*	Non-initiators 831,201 person-'trials'	Initiators 13,599 person-'trials'
Male, %	48.9	52.6
Townsend deprivation score of 3 or more, %	46.9	50.5
Age, years	64.0 (0.01)	65.5 (0.06)
Low-density lipoprotein cholesterol, mmol/L	3.53 (0.001)	4.11 (0.009)
High-density lipoprotein cholesterol, mmol/L	1.49 (0.0005)	1.43 (0.003)
Body mass index, kg/m^2	27.5 (0.005)	28.4 (0.04)
Systolic blood pressure, mmHg	141 (0.02)	145 (0.16)
Alcohol use, units/day	0.07 (0.0003)	0.07 (0.002)
Doctor visits in the past 3 months	2.7 (0.003)	5.0 (0.02)
Referrals in the past 3 months	0.45 (0.001)	0.80 (0.01)
Hospitalizations in the past 3 months	0.02 (0.0002)	0.03 (0.002)
Alcoholism, %	0.07	0.04
Smoking prevalence (current or former), %	43.3	49.0
Hypertension, %	59.5	70.2
Antihypertensive use, %	53.6	70.0
Diabetes, %	12.7	30.6
Insulin use, %	1.2	2.3
NSAIDs use, %	10.9	12.6
Aspirin use, %	9.3	27.3
Other lipid-lowering drugs use, %	0.9	1.0
Beta-blockers use, %	19.5	24.3
Hormone replacement therapy, % of women	9.8	5.8
Chronic obstructive pulmonary disease, %	3.4	3.9
Oral steroids use, %	1.6	1.6
Inhaled steroids use, %	5.7	6.0
Atrial fibrillation, %	1.3	1.8
Depression, %	15.1	15.0
Antidepressant use, %	6.3	7.1
Hypothyroidism, %	5.5	5.6
Osteoporosis, %	2.9	2.9
Calendar year 2000, %	6.1	2.6
Calendar year 2001, %	7.1	5.2
Calendar year 2002, %	8.2	6.8
Calendar year 2003, %	12.6	12.9
Calendar year 2004, %	18.3	20.2

Characteristic*	**Non-initiators 831,201 person-'trials'**	**Initiators 13,599 person-'trials'**
Calendar year 2005, %	22.9	22.6
Calendar year 2006, %	24.7	29.8

* Numbers are mean (standard error), unless otherwise specified Standard deviations reported here do not take into account that each individual may contribute to more than one trial

Table 2

Odds ratio (95% confidence interval) for probability of receiving statin therapy by previous treatment status,[a,b] THIN 'trials' 2000-2006

Factor	Non-users	Current users
Male	1.45 (1.41, 1.50)	1.12 (1.05, 1.20)
Age at baseline		
55-59	1.00	1.00
60-64	1.32 (1.24, 1.41)	1.09 (0.97, 1.22)
65-69	1.49 (1.36, 1.64)	1.10 (0.92, 1.31)
70-74	1.46 (1.29, 1.65)	1.06 (0.84, 1.33)
75-79	1.06 (0.91, 1.25)	0.91 (0.68, 1.20)
80-84	0.94 (0.77, 1.15)	0.88 (0.61, 1.26)
Townsend deprivation score		
1 (least deprived)	1.00	1.00
2	0.98 (0.94, 1.02)	0.96 (0.88, 1.05)
3	1.01 (0.97, 1.05)	0.87 (0.80, 0.94)
4	1.06 (1.02, 1.10)	0.86 (0.79, 0.94)
5 (most deprived)	1.12 (1.07, 1.18)	0.81 (0.73, 0.90)
LDL-cholesterol (mmol/L)		
< 2.6	1.00	1.00
2.6 - 3.35	1.38 (1.29, 1.49)	0.50 (0.46, 0.54)
3.36 - 4.13	2.65 (2.45, 2.86)	0.31 (0.28, 0.34)
4.14 - 4.89	4.29 (3.93, 4.67)	0.28 (0.25, 0.31)
≥ 4.9	7.90 (7.18, 8.69)	0.27 (0.23, 0.30)
HDL-cholesterol (mmol/L)		
< 1.0	1.01 (0.94, 1.08)	1.06 (0.94, 1.21)
1.0 - 1.39	1.00	1.00
1.4 - 1.69	1.01 (0.96, 1.06)	1.01 (0.92, 1.09)
≥ 1.7	1.06 (1.00, 1.13)	0.98 (0.88, 1.09)
Body mass index (kg/m^2)		
< 20	0.87 (0.78, 0.96)	0.85 (0.70, 1.04)
20 - 24.9	1.00	1.00
25 - 29.9	1.05 (1.02, 1.09)	1.03 (0.96, 1.11)
30 - 34.9	1.05 (1.01, 1.09)	1.05 (0.96, 1.15)
≥ 35	1.06 (1.01, 1.12)	0.92 (0.82, 1.03)
Systolic blood pressure (mmHg)		
< 130	1.00	1.00
130 – 139	1.16 (1.08, 1.24)	1.02 (0.90, 1.15)
140 – 159	1.28 (1.19, 1.37)	0.92 (0.81, 1.04)
≥ 160	1.61 (1.49, 1.74)	0.79 (0.68, 0.91)

Factor	Non-users	Current users
Diabetes	3.30 (3.16, 3.45)	1.09 (1.00, 1.19)
Hypertension	0.89 (0.85, 0.93)	0.75 (0.69, 0.81)
Atrial fibrillation	1.24 (1.14, 1.34)	0.91 (0.78, 1.07)
Antihypertensive use	2.12 (2.01, 2.22)	3.85 (3.52, 4.21)
Insulin	0.98 (0.84, 1.15)	0.98 (0.76, 1.26)
Other lipid-lowering drugs	0.54 (0.46, 0.62)	0.11 (0.09, 0.13)
Beta-blockers	1.30 (1.24, 1.36)	1.04 (0.95, 1.15)
Doctor visits in the past 3 months		
None	1.00	1.00
1	4.01 (3.48, 4.62)	1.52 (1.38, 1.67)
2 to 4	15.50 (13.61, 17.66)	1.70 (1.56, 1.85)
5 or more	27.28 (23.91, 31.12)	1.63 (1.47, 1.80)
Referrals in the past 3 months		
None	1.00	1.00
1	1.14 (1.10, 1.18)	0.99 (0.91, 1.06)
2 to 4	1.28 (1.23, 1.33)	0.87 (0.79, 0.95)
5 or more	1.15 (1.04, 1.26)	0.64 (0.53, 0.77)
Hospitalizations in the past 3 months		
None	1.00	1.00
1	1.23 (1.14, 1.33)	0.93 (0.78, 1.11)
2 or more	1.66 (1.48, 1.87)	0.63 (0.50, 0.81)
Month of the ‘trial’	1.06 (1.06, 1.07)	0.99 (0.98, 1.00)
Month of the ‘trial’ squared	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)

[a]These estimates correspond to an analysis that did censor non-adherent subjects (i.e., approach ii). Estimates obtained after artificially censoring non-adherent patients (approach i) were similar (data not shown).

[b]All models also included baseline levels of all of the covariates, four sets of indicators for time since last measurement of SBP, LDL- and HDL-cholesterol (≤6, 7-12, 13-18, 19-24 and >24 months), indicator for baseline year for each individual, and duration of last treatment episode in months (linear and quadratic).

Table 3

Hazard ratio (95% CI) estimates of the intention-to-treat effect and the effect of continuous statin therapy on risk of CHD, THIN 'trials' 2000-2006

	Intention-to-treat analysis	Per-protocol analysis	As-treated analysis
	Initiator vs. non-initiators	Continuous users vs. never users	2 years of use vs. no use
Entire follow-up			
Unique cases	635	488	635
Unique persons	74,806	74,806	74,806
Cases	6,335	4,849	6,335
Person-'trials'	844,800	844,800	844,800
Unadjusted	1.37 (1.14, 1.66)	1.70 (1.20, 2.41)	1.24 (0.88, 1.74)
Age-sex adjusted	1.29 (1.06, 1.56)	1.54 (1.09, 2.18)	1.11 (0.79, 1.56)
Adjusted for baseline covariates [a]	0.89 (0.73, 1.09)	0.86 (0.59, 1.27)	0.73 (0.50, 1.06)
Adjusted for baseline [a] and time-varying covariates [b]	n/a	0.84 (0.54, 1.30)	0.79 (0.41, 1.41)
Excluding first year of follow-up			
Unique cases	425	304	425
Unique persons	58,602	58,602	58,602
Cases	4,122	2,936	4,122
Person-'trials'	626,006	626,006	626,006
Unadjusted	1.06 (0.81, 1.39)	1.14 (0.66, 1.96)	1.17 (0.83, 1.67)
Age-sex adjusted	1.00 (0.77, 1.31)	1.02 (0.59, 1.76)	1.06 (0.74, 1.52)
Adjusted for baseline covariates [a]	0.71 (0.53, 0.94)	0.56 (0.31, 1.01)	0.72 (0.49, 1.06)
Adjusted for baseline [a] and time-varying covariates [b]	n/a	0.53 (0.27, 1.02)	0.68 (0.35, 1.31)

[a] Baseline variables in Table 1 were included as covariates in model for CHD risk.

[b] Time-varying variables were used to estimate IP weights. The model also included baseline variables measured at the start of follow-up for each individual.

Table 4

Hazard ratios (95% CI) of CHD for current users vs. never-users of statin therapy by duration of use and using different analytical methods for handing missing data, THIN 'trials' 2000-2006 [a]

Analysis	Current users vs. never-users [b]	Selected current vs. never-users [b,c]	Selected current vs. never-users, LOCF [c]
All current users			
Unique cases	635	513	513
Unique persons	74,806	61,490	61,490
Unadjusted	1.65 (1.38, 1.98)	1.53 (1.24, 1.89)	1.53 (1.24, 1.89)
Age-sex adjusted	1.48 (1.23, 1.77)	1.38 (1.11, 1.71)	1.38 (1.11, 1.71)
Adjusted for covariates [d]	1.42 (1.16, 1.73)	1.31 (1.04, 1.66)	1.41 (1.11, 1.78)
Current users with at least 1 year of statin use			
Unadjusted	1.16 (0.89, 1.51)	1.00 (0.72, 1.40)	1.00 (0.72, 1.40)
Age-sex adjusted	1.01 (0.77, 1.31)	0.88 (0.62, 1.23)	0.88 (0.62, 1.23)
Adjusted for covariates [d]	1.05 (0.79, 1.40)	0.95 (0.66, 1.35)	1.04 (0.72, 1.49)

LOCF: last observation carried forward indefinitely for each patient

[a] All models included month of follow-up and its squared term.

[b] Using missing indicators for variables with missing values for more than 12 or 24 months

[c] Restricted to those who did not develop selected chronic diseases (i.e., cancer, schizophrenia, psychosis, renal diseases, liver diseases and dementia) during follow-up

[d] Adjusted for the variables presented in Table 2.