Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
22(1) 70–96

DOI: 10.1177/0962280211403603
smm.sagepub.com
SAGE

Observational data for comparative effectiveness research: An emulation of randomised trials of statins and 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]

Abstract

This article reviews methods for comparative effectiveness research using observational data. The basic idea is using an observational study to emulate a hypothetical randomised trial by comparing initiators versus non-initiators of treatment. After adjustment for measured baseline confounders, one can then conduct the observational analogue of an intention-to-treat analysis. We also explain two approaches to conduct the analogues of per-protocol and as-treated analyses after further adjusting for measured time-varying confounding and selection bias using inverse-probability weighting. 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 UK. Despite strong confounding by indication, our approach detected a potential benefit of statin therapy. The analogue of the intention-to-treat hazard ratio (HR) of CHD was 0.89 (0.73, 1.09) for statin initiators versus non-initiators. The HR 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 versus no use. In contrast, a conventional comparison of current users versus never users of statin therapy resulted in a HR of 1.31 (1.04, 1.66). We provide a flexible and annotated SAS program to implement the proposed analyses.

Keywords

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

[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

Corresponding author:
Goodarz Danaei, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
Email: gdanaei@hsph.harvard.edu

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 analysing observational data, which are increasingly based on electronic medical records and other large healthcare databases.

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 or never 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 (UK). 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 this article 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 to analyze 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: The Health Improvement Network

The Health Improvement Network (THIN) is a database of electronic medical records collected by general practitioners in the UK. In 2009, THIN covered 6.2 million individuals from about 350 practices. For each individual, demographic and socioeconomic characteristics as well as symptoms, signs and diagnoses, referrals, laboratory test results and some lifestyle information are recorded and updated using standardised methods. Vital status and cause of death data are 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 4126 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 on the recorded data for body mass index (BMI), systolic blood pressure (SBP), LDL- and HDL-cholesterol levels 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 USA (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 individuals in the database, we will refer to our observational study as a non-randomised observational '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 enrol participants aged 55–84 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. Participants must not have used statins for at least 2 years prior to enrolment (the washout period). The duration of the enrolment 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 (operationalised 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, revascularisation, 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 3178 individuals 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 3178 individuals 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 3178 individuals, 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 1-month enrolment period. Table A1 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 individuals. Of these, 635 had CHD, 1667 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 individuals 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 18 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 6335 developed CHD (117 among initiators). The average duration of follow-up was 25.8 months for initiators and 29.6 months for non-initiators. Figure 1 shows a flowchart 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 levels, 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 contrast, estimating causal effects from our 'trials' requires confounding adjustment, as described in

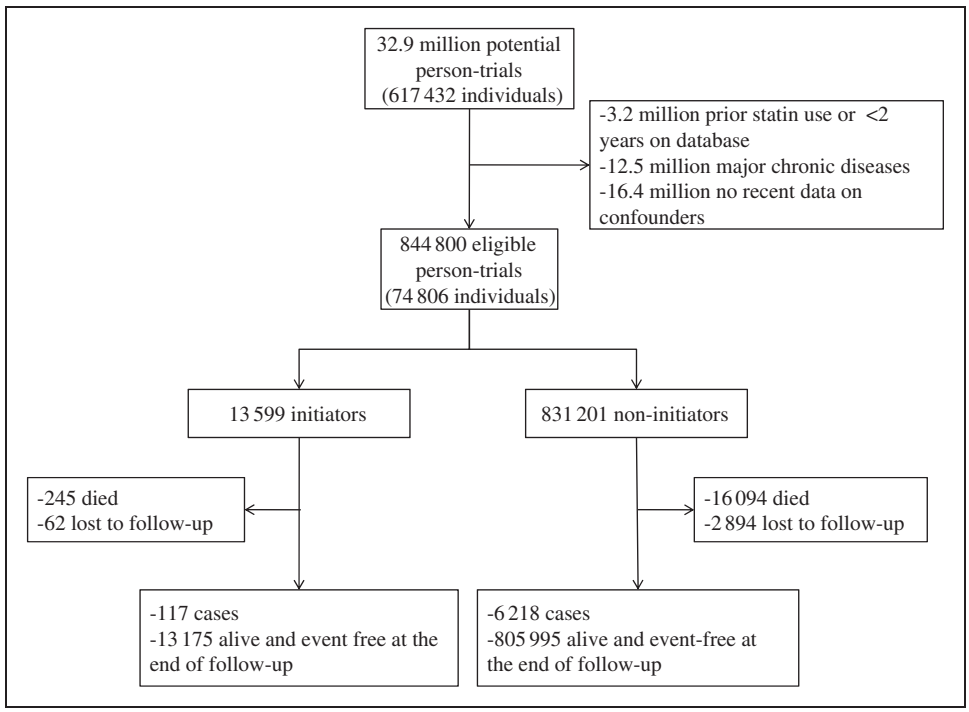

Figure 1. Flowchart of person-trials in the analysis (THIN database 2000–2006).

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.

Another difference between our non-randomised 'trials' and the hypothetical randomised trial is that the hypothetical randomised trial would exclude patients who have undiagnosed CHD at baseline *via* a systematic physical exam or laboratory tests. In contrast, many of these patients 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 behaviour 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-labelled 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 behavioural changes. In contrast, a double-blinded randomised trial is designed to estimate the pharmacological effects only. The choice between open-labelled 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 behavioural changes. In that case, an open-labelled trial may be more appropriate than a double-blinded trial.

Table 1. Characteristics of initiators and non-initiators of statin therapy at the start of the trial's follow-up, THIN 'trials' 2000–2006

Characteristics[a]	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 (7.7)	65.5 (7.3)
Low-density lipoprotein cholesterol, mmol/L	3.53 (0.9)	4.11 (1.0)
High-density lipoprotein cholesterol, mmol/L	1.49 (0.4)	1.43 (0.4)
Body mass index, kg/m^2	27.5 (4.9)	28.4 (5.0)
Systolic blood pressure, mmHg	141 (17)	145 (18)
Alcohol use, units/day	0.07 (0.3)	0.07 (0.3)
Doctor visits in the past 3 months	2.7 (2.5)	5.0 (2.7)
Referrals in the past 3 months	0.45 (0.9)	0.80 (1.2)
Hospitalisations in the past 3 months	0.02 (0.2)	0.03 (0.2)
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
Calendar year 2005, %	22.9	22.6
Calendar year 2006, %	24.7	29.8

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

Indeed, 'pragmatic' randomised trials with less restrictive eligibility criteria and an open-labelled 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.[15]

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 versus the control group, which is commonly referred to as the intention-to-treat hazard ratio (HR). Alternatively, one can approximately estimate this HR *via* the odds ratio from a pooled logistic model[16] 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 HR in our non-randomised 'trials' by fitting a pooled logistic model like the one described above except that (1) the indicator for experimental treatment assignment was replaced by an indicator for observed therapy initiation and (2) potential confounders (measured at the baseline of each 'trial') were added to the model as covariates. We pooled data from the 844 800 person-'trials' from all 83 'trials' into a single model rather than fitting a separate model for each 'trial' and then pooling the 83 HR 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 individuals participated in more than one 'trial', we used a robust variance estimator to estimate conservative 95% confidence intervals (CI).[17]

The average HR (95% CI) of CHD for statin therapy initiation versus 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 HR of 0.89 can be interpreted as the intention-to-treat average HR of CHD conditional on the confounders.

To assess the potential heterogeneity of the HR 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 HR 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.[18] 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 HR was 0.88 (95% CI 0.72, 1.08) which is almost identical to the HR 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

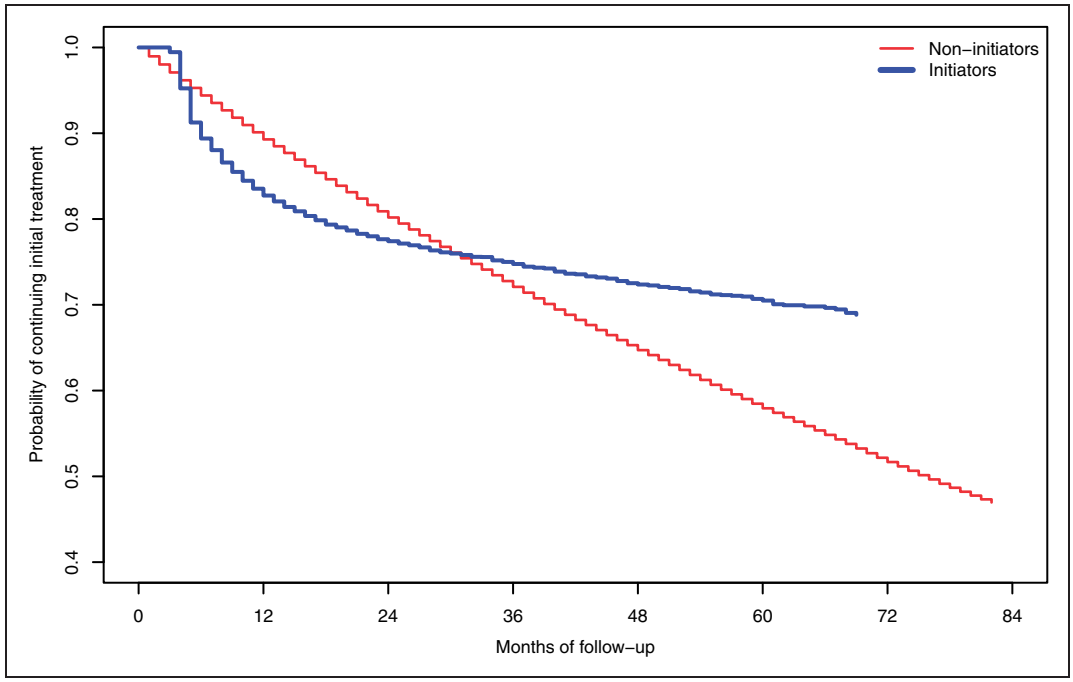

Figure 2. Adherence to statin treatment by initiation status (THIN 'trials' 2000–2006).

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 HR is guaranteed to differ from our estimate of 0.89. Furthermore, in studies comparing therapy with no therapy, imperfect adherence moves the intention-to-treat effect towards the null, which may result in missing small beneficial or harmful 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 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',[19,20] 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 analogous 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 HR estimated by the model compares continuous treatment versus no treatment (conditional on the measured covariates at baseline).[20]

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 HR for continuous treatment versus no treatment was 0.86 (0.59, 1.27). For this HR 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. censoring may be informative.[21] 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 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 individuals 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 prognostic factors that caused 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 enroled 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 HR 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 HR estimate for 2 years of treatment versus no treatment. The HR 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 (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 two 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 past year, or average duration of treatment as defined in the Appendix).[20]

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% CI for the HR. Conversely, modelling the dose–response function in an as-treated analysis may yield a narrower 95% CI, 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 IP 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 individual at a given time is the inverse (reciprocal) of her probability of having received the treatment history that she actually received.[22,23] For example, suppose an individual 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 individual, 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.[22,23] 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).[22,23]

To compute the time-varying IP weights, we need each individual'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 individual'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 levels (in six 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 (non-users).

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 65–74 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 unstabilised weights. Because unstabilised weights generally yield wide CIs for the HR, one needs to stabilise the weights in most applications. Informally, stabilisation is accomplished by multiplying a individual's unstabilised 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 stabilised 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.[24] 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 HR (95% CI) 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*

Table 2. Odds ratio (95% CI) 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)
BMI (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)
SBP (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)
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)

(continued)

Table 2. Continued

Factor	Non-users	Current users
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)
2–4	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)
2–4	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)
Hospitalisations 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 not censor non-adherent individuals (i.e. as-treated). Estimates obtained after artificially censoring non-adherent individuals (per-protocol) 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 ($\leq$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).

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[25,26] and represent the effect of continuous treatment versus no treatment (per-protocol) or the effect of 2 years of treatment versus 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 HR for 2 years of treatment versus 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 and as-treated). We examined the effect of censoring due to loss to follow-up in the THIN database using IP weights for censoring. As expected, the results were almost identical to those without IP weights for censoring because only 0.4% of individuals in this database were lost to follow-up.

4.5 Unmeasured confounding by undiagnosed disease

All 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 judgement) 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.[10,27,28] 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 HR is exactly 1 – during the first 6 months of follow-up. In this setting, estimating an average HR 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 HR over the rest of the follow-up.

In our 'trials', the average HR 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 HR should be 1 because statin therapy does not affect the risk of CHD during the first 6–12 months of use.[29] 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 HR after excluding the early follow-up by increments of 3 months up to 24 months. The intention-to-treat HR was 0.71 (0.53, 0.94) after excluding 12 months and 0.66 (0.45, 0.97) after excluding 24 months. The HR 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 HR 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 HR is arithmetically expected even in the absence of unmeasured confounding because the overall HR is a weighted average of the time-varying HRs. Thus, if the HR 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 HRs were increasing in time, excluding early follow-up would increase the average HR.

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 oestrogen plus progestin therapy commonly used in the USA. A large randomised trial showed that this hormone therapy increases the risk of CHD during at least the first 2 years of use,[6,30] i.e. the average HR is greater than 1 during the first 2 years. Excluding the first few years of follow-up in the trial yielded a HR 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 HR 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 HR will be less than 1 after excluding the first few years.[31]

In our statin analysis, one could argue that excluding 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 6–12 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.

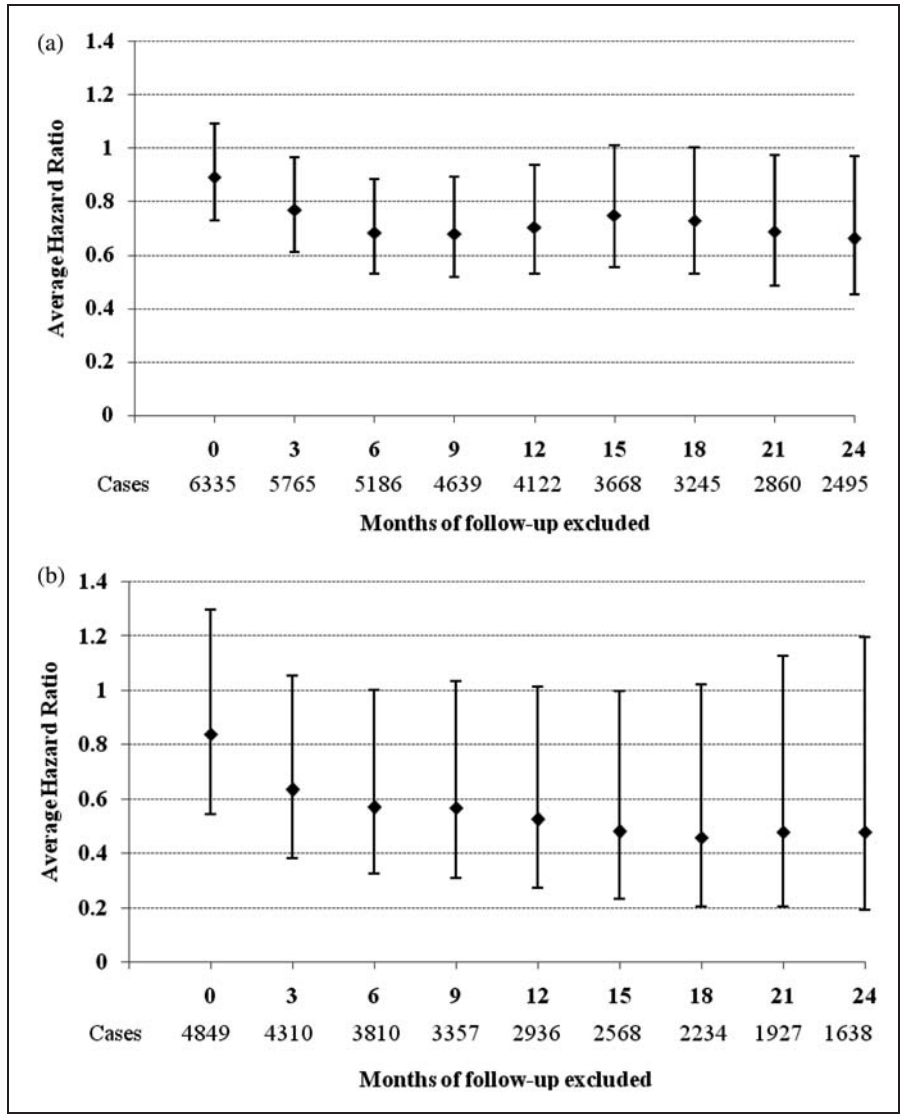

Figure 3. Impact of excluding varying periods of early follow-up (from 0 to 24 months) on the intention-to-treat HR (a), and the HR for continuous use *vs.* no use (b), THIN 'trials' 2000–2006.
Notes: Vertical bars represent 95% CIs. Early follow-up was excluded from the outcome model but not from the model used to estimate the IP weights.

4.6 Summary of results

The main results in Sections 4.1–4.5 are summarised in Table 3. The intention-to-treat HR after adjustment for all baseline covariates was 0.89. The large difference between the unadjusted (1.37) and adjusted (0.89) HRs 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

Table 3. HR (95% CI) estimates of the effect of statin therapy on risk of CHD, THIN 'trials' 2000–2006

	Intention-to-treat analysis Initiator vs. non-initiators	Per-protocol analysis Continuous users *vs.* never users	As-treated analysis 2 years of use *vs.* no use
Entire follow-up			
Unique cases	635	488	635
Unique persons	74 806	74 806	74 806
Cases	6335	4849	6335
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	4122	2936	4122
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.

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 the per-protocol and as-treated estimates, we used two forms of multivariate adjustment: including 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 time-varying confounders.

Excluding the first year of follow-up resulted in HRs 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 are often interpreted as an *ad hoc* sensitivity analysis for unmeasured confounding, but remember the caveats discussed in Section 4.5. A more formal – but also harder to implement – approach for sensitivity analysis would be to incorporate the assumptions about the timing of treatment effects into g-estimation of a structural model.[32]

5 The 'current versus never users' comparison

A common version of as-treated analysis in observational studies is the comparison of current users versus non-users or never users of a particular treatment. In this contrast, the analyst simply

Table 4. HRs (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 database 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 individual.
[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.

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 HR 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 HR in our study using different methods for handling missing data and exclusion criteria. The multivariate adjusted HR 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 HR may be 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) HRs 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 HR because the HR 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 (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 HR of CHD for 'long-term current *vs*. never users' comparison ranged between 0.95 and 1.05 in our study, compared with 1.31–1.42 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 HRs for 'long-term current *vs*. never users' were in fact more 'improved' (ranged between 0.88 and 1.01) than the fully adjusted HRs. 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 HR. For example, in our study, the HR 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 that 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' HR is an appropriate method to adjust for imperfect adherence. For example, in re-analyses of observational studies of postmenopausal hormone therapy[6,33], the 'initiators versus non-initiators' HR 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 article can be easily extended to estimate standardised survival curves that compare the cumulative risks under different treatment scenarios.[20]

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 HR estimate to be greater than 1. This article 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).[34] Rather, this article 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.[35,36]

Our success, however, was not complete. Our estimate of the intention-to-treat HR of CHD for statin therapy (0.89, 95% CI: 0.73, 1.09) had a wide 95% CI 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[34] 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 MI only after excluding the first year of follow-up. The reported HR was 0.86 (0.76, 0.97).[10] Seeger and colleagues used a database from a community Health Management Organisation in the USA 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 HR of 0.69 (0.52, 0.93).[9] Like our study, both of these studies compared initiators versus non-initiators. As showed above, the use of the 'current versus 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.[37,38] 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 HR 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.[24] 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 five non-initiators in 0.01 callipers 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 HR estimate was 0.83 (95% CI 0.67, 1.03) which is similar to our intention-to-treat HR 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 Hernán et al.[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') HR was 1.49 (0.96, 2.33) before adjustment for baseline covariates and 1.30 (0.83, 2.03) after adjustment. This HR suggests that cessation of statin therapy increases the risk of CHD.

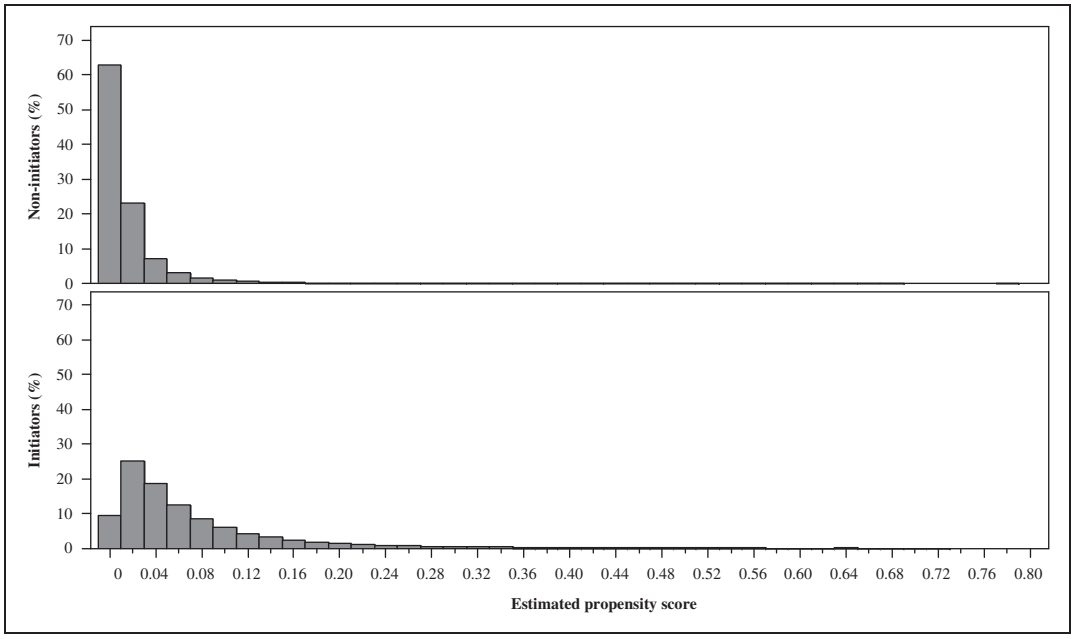

Figure 4. Distribution of estimated propensity for treatment initiation by treatment status at the start of the THIN 'trials' 2000–2006.

Another analytic method that might be have been used to adjust for confounding by indication is instrumental variable (IV) estimation.[19,39] 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 because there is little interest in estimating the effect of continuous treatment when many individuals cannot take treatment continuously. In settings with toxicity, the protocol of the randomised trial would specify that treatment has to be discontinued in case of toxicity. In this setting, a simple extension of our censoring procedure can be used to conduct a per-protocol analysis using observational data, as described elsewhere.[19]

In summary, 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.

Acknowledgements

The authors 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, 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): 2560–2572.
2. NIHLB. *Detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III)*. Bethesda: National Institute of Heart Lung and Blood, 2002.
3. NIHCE. *Statins for the prevention of cardiovascular events*. London: National Institute for Health and Clinical Excellence, 2006.
4. Gibbons RJ, Gardner TJ, Anderson JL, et al. The American Heart Association's principles for comparative effectiveness research: a policy statement from the American Heart Association. *Circulation* 2009; **119**(22): 2955–2962.
5. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–920.
6. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008; **19**(6): 766–779.
7. Hernán MA and Robins JM. Authors' response, part I: Observational studies analyzed like randomized experiments: best of both worlds. *Epidemiology* 2008; **19**(6): 789–792.
8. Taylor F, Ward K, Moore TH, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2011; **1**: CD004816.
9. Seeger JD, Williams PL and Walker AM. An application of propensity score matching using claims data. *Pharmacoepidemiol Drug Saf* 2005; **14**(7): 465–476.
10. Smeeth L, Douglas I, Hall AJ, et al. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials, an application of propensity score matching using claims data. *Br J Clin Pharmacol* 2009; **67**(1): 99–109.
11. Jick H, Zornberg GL, Jick SS, Seshadri S and Drachman DA. Statins and the risk of dementia. *Lancet* 2000; **356**(9242): 1627–1631.
12. Lewis JD, Schinnar R, Bilker WB, Wang X and Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2007; **16**(4): 393–401.
13. Garcia Rodriguez LA, Tacconelli S and 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* 2008; **52**(20): 1628–1636.
14. NHANES. *National Health and Nutrition Examination Survey data sets and related documentation (Survey Questionnaire, Examination and Laboratory Protocols, 1988-1994 and 1999-2002)*. MD: Centers for Disease Control and Prevention, 2002.
15. Mullins CD, Whicher D, Reese ES and Tunis S. Generating evidence for comparative effectiveness research using more pragmatic randomized controlled trials. *Pharmacoeconomics* 2010; **28**(10): 969–976.
16. Thompson WA Jr. On the treatment of grouped observations in life studies. *Biometrics* 1977; **33**(3): 463–470.
17. Zeger SL and Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**(1): 121–130.
18. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17**(19): 2265–2281.
19. Toh S and Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat* 2008; **4**(1): Article 22.
20. Toh S, Hernandez-Diaz S, Logan R, Robins JM and Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology* 2010; **21**(4): 528–539.
21. Hernán MA, Hernandez-Diaz S and Robins JM. A structural approach to selection bias. *Epidemiology* 2004; **15**(5): 615–625.
22. Hernán MA and Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006; **60**(7): 578–586.
23. Robins JM and Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G and Monenberghs G (eds) *Longitudinal data analysis*. Boca Raton: Chapman & Hall/CRC, 2009, pp.553–599.
24. Cole SR and Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; **168**(6): 656–664.
25. Robins JM, Hernán MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**(5): 550–560.
26. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 2001; **96**(454): 440–448.
27. Jee SH, Sull JW, Park J, et al. Body-mass index and mortality in Korean men and women. *N Engl J Med* 2006; **355**(8): 779–787.
28. Allison DB, Faith MS, Heo M, Townsend-Butterworth D and Williamson DF. Meta-analysis of the effect of excluding early deaths on the estimated relationship between body mass index and mortality. *Obes Res* 1999; **7**(4): 342–354.
29. Simes J, Furberg CD, Braunwald E, 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. *Eur Heart J* 2002; **23**(3): 207–215.
30. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; **349**(6): 523–534.
31. Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010; **21**(1): 13–15.
32. Robins JM. Causal models for estimating the effects of weight gain on mortality. *Int J Obes* 2008; **32**(Suppl 3): S15–S41.

33. Hernán MA and Robins JM. Discussion on "Statistical Issues Arising in the Women's Health Initiative". *Biometrics* 2005; **61**(4): 922–930.
34. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983; **2**(2): 267–271.
35. Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. *JAMA* 2010; **304**(4): 411–418.
36. Haag MD, Bos MJ, Hofman A, Koudstaal PJ, Breteler MM and Stricker BH. Cyclooxygenase selectivity of nonsteroidal anti-inflammatory drugs and risk of stroke. *Arch Intern Med* 2008; **168**(11): 1219–1224.
37. Schneeweiss S, Patrick AR, Sturmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–S142.
38. Psaty BM and Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* 2010; **304**(8): 897–898.
39. Hernán MA and Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**(4): 360–372.
40. Robins JM, Hernán MA and Rotnitzky A. Effect modification by time-varying covariates. *Am J Epidemiol* 2007; **166**(9): 994–1002; discussion 3–4.
41. Petersen ML, Deeks SG, Martin JN and van der Laan MJ. History-adjusted marginal structural models for estimating time-varying effect modification. *Am J Epidemiol* 2007; **166**(9): 985–993.
42. VanderWeele TJ and Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol* 2007; **166**(9): 1096–1104.
43. VanderWeele TJ, Hernán MA and Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* 2008; **19**(5): 720–728.

Appendix

Constructing nested non-randomised 'trials' from observational data

Section 3.2 describes the emulation of a sequence of 83 non-randomised 'trials' in the THIN database. Table A1 presents the number of participants, initiators and cases for each 'trial'. As explained in Section 4, we pooled these 'trials' to estimate the analogue of the intention-to-treat effect, which requires data on the baseline confounders. Table A2 presents the data for three hypothetical individuals. We restricted the confounders to LDL-cholesterol to simplify the table.

To fit the pooled logistic regression model described in Section 4, we expanded the data set so that each observation corresponds to 1 month of follow-up for a particular individual in an eligible person-'trial'. Table A3 presents the expanded dataset where m denotes the baseline month of the 'trial', t the follow-up month, E_m the eligibility for 'trial' m (1: eligible, 0: ineligible), A_m and A_{m+t} the values of treatment at time m and $m + t$ of 'trial' m, respectively (1: on treatment, 0: not on treatment), L_{m+t} the value of LDL at month t of 'trial' m and finally D_{m+t+1} the occurrence of CHD in month $t + 1$ of 'trial' m.

In this particular example, the first and second individuals each contributed to two 'trials' and the third individual to three 'trials'. Notice that we only expanded the follow-up starting from the first month of each eligible person-'trial' but all subsequent months for each 'trial' were included in the follow-up, even if they were not eligible for expansion themselves. For brevity, we have assumed that follow-up ended at month 70.

Disease models

We used pooled logistic models to estimate HRs in all analyses suggested in Section 4. Here, we present the models using the same notation as described above for Table A3 except that from here onwards L_m denotes a vector of covariates used in the model. We have also added overbars to indicate the previous history of a covariate from the start of follow-up and used a superscript T to denote a transpose of a vector of parameters.

The model for the intention-to-treat analysis is:

$$\text{logit}\left[pr\left(D_{m+t+1} = 1 | A_m, L_m, \bar{D}_{m+t} = 0, E_m = 1\right)\right] = \alpha_{0,m+t} + \alpha_1 A_m + \alpha_2^T L_m$$

Table A1. Number of participants, initiators and cases in each non-randomised THIN 'trial'

Trial number	Calendar month	Participants	CHD cases	Initiators	CHD cases in initiators
1	Jan 2000	3178	91	18	1
2	Feb 2000	3931	102	25	0
3	Mar 2000	4067	101	36	0
4	Apr 2000	4209	109	35	2
5	May 2000	4365	107	28	2
6	Jun 2000	4573	107	34	0
7	Jul 2000	4961	113	25	0
8	Aug 2000	5177	117	33	0
9	Sep 2000	5386	120	39	1
10	Oct 2000	5571	120	34	0
11	Nov 2000	6000	129	41	2
12	Dec 2000	6319	134	50	1
13[a]	Jan 2001	3615	72	37	1
14	Feb 2001	3757	74	43	1
15	Mar 2001	3933	75	42	1
16	Apr 2001	4280	76	43	0
17	May 2001	4501	80	63	3
18	Jun 2001	4717	81	54	2
19	Jul 2001	5237	89	72	3
20	Aug 2001	5573	91	66	2
21	Sep 2001	5775	91	58	1
22	Oct 2001	6064	89	83	1
23	Nov 2001	6225	89	100	2
24	Dec 2001	6568	91	95	3
25[a]	Jan 2002	4280	63	50	1
26	Feb 2002	4525	63	57	0
27	Mar 2002	4818	75	81	0
28	Apr 2002	4885	79	60	0
29	May 2002	5256	85	80	3
30	Jun 2002	5404	75	72	0
31	Jul 2002	6026	83	92	2
32	Aug 2002	6433	80	62	0
33	Sep 2002	6734	84	81	1
34	Oct 2002	7018	95	101	0
35	Nov 2002	7208	92	94	2
36	Dec 2002	7405	94	86	5
37	Jan 2003	7488	86	96	0
38	Feb 2003	7722	83	111	1
39	Mar 2003	7953	79	138	2
40	Apr 2003	8282	78	136	2
41	May 2003	8627	83	127	4
42	Jun 2003	8841	86	167	0
43	Jul 2003	9204	93	196	6
44	Aug 2003	9612	90	138	1
45	Sep 2003	9988	89	165	3

(continued)

Table A1. Continued

Trial number	Calendar month	Participants	CHD cases	Initiators	CHD cases in initiators
46	Oct 2003	10 449	92	195	3
47	Nov 2003	10 921	102	196	4
48	Dec 2003	11 169	93	168	4
49	Jan 2004	11 182	91	193	4
50	Feb 2004	11 564	94	213	3
51	Mar 2004	12 118	97	254	2
52	Apr 2004	12 532	89	247	0
53	May 2004	12 745	92	229	4
54	Jun 2004	12 982	82	205	0
55	Jul 2004	13 460	86	237	3
56	Aug 2004	13 759	80	179	1
57	Sep 2004	14 110	80	256	1
58	Oct 2004	14 536	74	259	1
59	Nov 2004	14 843	77	307	3
60	Dec 2004	15 128	73	227	2
61	Jan 2005	15 557	70	234	1
62	Feb 2005	15 874	65	283	1
63	Mar 2005	15 824	67	219	2
64	Apr 2005	15 822	62	236	1
65	May 2005	15 919	65	247	1
66	Jun 2005	15 927	61	268	3
67	Jul 2005	16 207	53	232	0
68	Aug 2005	16 573	53	233	1
69	Sep 2005	16 602	53	262	2
70	Oct 2005	16 677	48	312	1
71	Nov 2005	16 954	45	316	0
72	Dec 2005	17 162	41	299	1
73	Jan 2006	16 875	40	315	0
74	Feb 2006	17 075	39	386	0
75	Mar 2006	17 156	35	427	0
76	Apr 2006	16 907	30	334	0
77	May 2006	17 012	25	320	1
78	Jun 2006	16 934	26	345	1
79	Jul 2006	17 397	21	289	0
80	Aug 2006	18 081	17	303	0
81	Sep 2006	18 310	14	313	1
82	Oct 2006	18 357	13	373	2
83	Nov 2006	18 409	7	344	1

[a]The number of eligible participants drops after 12 and 24 months because missing data on covariates renders a person-trial ineligible and the value of covariates is only carried forward 12 or 24 months.

Table A2. Data for three hypothetical individuals

Individual	Trial	Eligible	Initiator	Current user	Baseline LDL(mmol/L)	LDL(mmol/L)	Month CHD	Month dead
1	12	1	0	0	2.47	2.47	0	14
1	13	1	0	0	2.47	2.47	0	14
2	24	1	0	0	2.77	2.77	26	26
2	25	1	1	1	2.77	2.77	26	26
2	26	0	0	1	2.77	2.84	26	26
3	43	1	1	1	2.88	2.88	0	0
3	44	0	0	1	2.88	2.88	0	0
3	45	0	0	0	2.88	2.77	0	0
3	⋮	⋮	⋮	⋮	2.88	⋮	0	0
3	67	0	0	0	2.88	2.71	0	0
3	68	1	0	0	2.88	2.71	0	0
3	69	1	1	1	2.88	2.73	0	0
3	70	0	0	1	2.88	2.73	0	0

Table A3. The expanded dataset for the three hypothetical individuals in Table A2

Individual	Trial (m)	Follow-up month (t)	Eligible (E_m)	Initiator (A_m)	Current user (A_{m+t})	Baseline LDL (L_m)	Time-varying LDL (L_{m+t})	Event (D_{m+t})
1	12	0	1	0	0	2.47	2.47	0
1	12	1	1	0	0	2.47	2.47	0
1	13	0	1	0	0	2.47	2.47	0
2	24	0	1	0	0	2.77	2.77	0
2	24	1	1	0	1	2.77	2.77	0
2	24	2	1	0	1	2.77	2.84	1
2	25	0	1	1	1	2.77	2.77	0
2	25	1	1	1	1	2.77	2.84	1
3	43	0	1	1	1	2.88	2.88	0
3	43	1	1	1	1	2.88	2.88	0
3	43	2	1	1	0	2.88	2.77	0
3	43	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	43	26	1	1	1	2.88	2.73	0
3	43	27	1	1	1	2.88	2.73	.
3	68	0	1	0	0	2.71	2.71	0
3	68	1	1	0	1	2.71	2.73	0
3	68	2	1	0	1	2.71	2.73	.
3	69	0	1	1	1	2.73	2.73	0
3	69	1	1	1	1	2.73	2.73	.

where $\alpha_{0,m+t}$ is a time-varying intercept (estimated as a constant plus linear and quadratic terms for both m and t) and α_1 approximately equals[16] the log HR for treatment initiation averaged over the duration of follow-up.

In the per-protocol analysis, we censored the person-time when someone discontinued their initial treatment assignment. That is, we fit the model:

$$\text{logit}\left[pr\left(D_{m+t+1}=1|A_m, L_m, \bar{D}_{m+t}=0, E_m=1, \bar{C}_{m+t+1}=0\right)\right]=\beta_{0,m+t}+\beta_1 A_m+\beta_2^T L_m$$

where $\bar{C}_{m+t+1}$ is the indicator for artificial censoring due to discontinuing initial treatment at time $t+1$ in 'trial' m.

Finally, for the dose–response functions in the as-treated analysis, we used the total duration of treatment and its squared term in the model as presented below:

$$\text{logit}\left[pr\left(D_{m+t+1}=1|A_m,L_m,\bar{D}_{m+t}=0,E_m=1\right)\right]=\gamma_{0,m+t}+\gamma_1\sum_{k=m}^{m+t}A_k+\gamma_2\left(\sum_{k=m}^{m+t}A_k\right)^2+\gamma_3^TL_m$$

Note that the average HR comparing x months of treatment versus no treatment is now approximately equal to $e^{\gamma_1 x+\gamma_2 x^2}$.

In the sensitivity analysis, we replaced total duration of treatment by average duration of treatment:

$$\text{logit}\left[pr\left(D_{m+t+1}=1|A_m,L_m,\bar{D}_{m+t}=0,E_m=1\right)\right]=\delta_{0,m+t}+\delta_1\left(\frac{1}{t}\sum_{k=m}^{m+t}A_k\right)+\delta_2\left(\frac{1}{t}\sum_{k=m}^{m+t}A_k\right)^2+\delta_3^TL_m$$

To make the model more flexible, one can add product ('interaction') terms between treatment history and time. For example:

$$\begin{aligned}&\text{logit}\left[pr\left(D_{m+t+1}=1|A_m,L_m,\bar{D}_{m+t}=0,E_m=1\right)\right]\\&=\theta_{0,m+t}+\theta_1\sum_{k=m}^{m+t}A_k+\theta_2\left(\sum_{k=m}^{m+t}A_k\right)^2+\theta_3t\sum_{k=m}^{m+t}A_k+\theta_4t\left(\sum_{k=m}^{m+t}A_k\right)+\theta_5^TLm\end{aligned}$$

In this case, the effect estimate cannot be summarised in a single average HR because the HR is allowed to vary over time.

Treatment models for IP weights

The stabilised weights for each patient at each time $m+t$ are defined as

$$SW_{m+t}^A=\prod_{k=m}^{m+t}\frac{f\left(A_k\middle|\,\bar{A}_{k-1},L_0,\bar{D}_{k-1}=0\right)}{f\left(A_k\middle|\,\bar{A}_{k-1},\bar{L}_k,\bar{D}_{k-1}=0\right)}$$

As described elsewhere,[25,26] we fit the logistic model

$$\text{logit}\left[pr\left(A_k=1|A_{k-1}=a,L_0,\bar{L}_k,\bar{D}_{k-1}=0\right)\right]=\varphi_0+\varphi_i^TL_0+\varphi_2^TL_k$$

to estimate $f\left(A_k\middle|\,\bar{A}_{k-1},\bar{L}_k,\bar{D}_{k-1}=0\right)$, and the logistic model

$$\text{logit}\left[pr\left(A_k=1|A_{k-1}=a,L_0,\bar{D}_{k-1}=0\right)\right]=\eth_0+\eth_1^TL_0$$

to estimate $f\left(A_k\middle|\,\bar{A}_{k-1},L_0,\bar{D}_{k-1}=0\right)$. We fit these models separately for $a=1$ and $a=0$ in the pre-expansion dataset (i.e. using data as in Table A2). Because we allowed a 60-day gap after the end of any prescription, the first two person-months after initiation of treatment had, by definition, a probability of being treated of 1. Therefore, we excluded these observations from the weight models.

Note that we could have chosen IP weights with a numerator different from that of the SW weights, and that certain choices of numerator would result in more efficient estimators (i.e. narrower 95% CIs) at the expense of complex computations. The relative inefficiency of the SW weights explains why the 95% CI of the HR was wider for the as-treated analysis compared with the per-protocol analysis when we used a weighted model.

As explained in Section 4.4, for the disease models used in the per-protocol and as-treated analyses, the contributions of each individual at each time $m + t$ were weighted by the individual-specific weights SW^{A}_{m+t} in order to adjust for time-varying confounding or selection bias. Our estimates can be considered as a special case of IP-weighted estimation of a history-adjusted marginal structural model.[40,41] Because the numerator of the weights was conditional on the baseline variables L_0, the weighted disease models included these variables as well.

Explanation for bias introduced in the 'current vs. never users' comparison

Let us assume that the first two months of our study can be represented by the causal directed acyclic graph (DAG) shown in Figure A1. The nodes A_0 and A_1 represent indicators of statin therapy use at months 0 and 1, respectively, L_1 a vector of prognostic factors (e.g. LDL), D_2 the dichotomous indicator of disease during month 2 and U an unmeasured variable (e.g. poor diet or genetic factors). For simplicity, the DAG does not include the baseline values L_0. The arrow from A_0 to L_1 represents the causal effect of statins on LDL; the arrow from L_1 to A_1 represents the effect of LDL on future initiation of statin therapy; the arrow from L_1 to D_2 represents the effect of LDL on CHD; and the arrows from U to L_1 and D_2 represent the effects of poor diet or genetic factors on LDL and CHD, respectively. The absence of an arrow from U to A_0 or A_1 indicates no unmeasured confounding. The absence of a direct arrow from A_0 to D_2 indicates that the effect of statin therapy is mediated through L_1, and the absence of an arrow from A_1 to D_2 indicated that statin therapy has no effect on CHD during the first month of use. The absence of direct arrows from A_0 and A_1 to D_2 is not necessary for our argument below but it simplifies the exposition.

The conventional 'current *vs.* never users' comparisons contrast the probability of D_2 between current users ($A_1 = 1$) and never users ($A_1 = 0$, $A_0 = 0$) at each time point after conditioning on L_1 to adjust for confounding by indication. This conditioning is usually implemented by adding indicators for past user (1 if $A_1 = 0$, $A_0 = 1$) and current user (1 if $A_1 = 1$), together with L_1, as covariates in a logistic or proportional hazards regression model. Under the DAG in Figure A1, the coefficient for A_1 in such a model should be zero if the model is correctly specified. However, the DAG shows that conditioning on L_1 opens the path $A_1A_0L_1UD_2$ because L_1 is a collider. This path creates an association between A_1 and D_2 among current users. Therefore, even before conducting the analysis, it is known that the coefficient for the indicator of current use in the model will not be zero. One may attempt to block this path by including A_0 in the outcome regression, but this precludes the detection of the effect of A_0, which is the only one that exists in our example. See the Appendix of Hernán and Robins[7] for an expanded discussion.

We can explore the direction of the bias in the 'current *vs.* never users' comparison by adding signs to the arrows in the DAG to represent the direction of the causal effects as suggested by Vanderweele et al.[42,43] Statins reduce LDL levels, thus the negative sign on the arrow from A_0 to L_1; poor diet increases LDL and risk of CHD, thus positive signs on arrows from U to L_1 and D_2; prior treatment is a strong predictor of future treatment, thus, a positive sign on the arrow from A_0 to A_1. If all variables are regarded as dichotomous, and under the assumption of no interaction between A_0 and U discussed by VanderWeele et al.,[42,43] conditioning on L_1 and thus opening the path

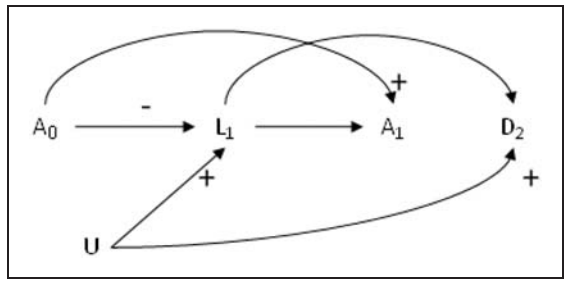

Figure A1. Causal relationships between treatment, confounders and outcome in the first 2 months of the study with the direction of the effects designated by signed arrows.

$A_1A_0L_1UD_2$ will introduce a positive bias such that under the null hypothesis of no effect of A_1 on D_2, A_1 will be associated with a higher risk of D_2. This is consistent with our results in the THIN analysis, where the 'current *vs*. never users' contrast yielded a HR of CHD greater than 1 for statin use versus no use (Section 5). A similar argument can be proposed when L_1 is replaced by D_1, and the HR of D_2 is conditional on having survived without developing CHD (i.e., $D_1 = 0$). Our proposed comparison of initiators versus non-initiators within nested non-randomised 'trials' of eligible individuals prevents this problem because it conditions on A_0 by design.