Importance of the lipid-related pathways in the association between statins, mortality and cardiovascular disease risk: The Multi-Ethnic Study of Atherosclerosis

Denis Talbot[1,2], Joseph A. (Chris) Delaney[3], Veit Sandfort[4], David M. Herrington[5], Robyn L McClelland[6]

[1] Département de médecine sociale et préventive, Université Laval, Québec, Qc, Canada
[2] Unité santé des populations et pratiques optimales en santé, CHU de Québec – Université Laval research center, Québec, Qc, Canada
[3] Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA
[4] National Institutes of Health, Bethesda, MD, USA
[5] Heart and Vascular Center of Excellence, Wake Forest University School of Medicine, Winston Salem, NC, USA
[6] Department of Biostatistics, University of Washington School of Public Health, Seattle, WA, USA

Correspondance to:

Denis Talbot,
Département de médecine sociale et préventive
Faculté de médecine
Université Laval
1050, avenue de la Médecine
Pavillon Ferdinand-Vandry, room 2454
Québec (Québec) G1V 0A6
Tel. 418 656-2131 ext. 5911
Email: denis.talbot@fmed.ulaval.ca

Key words: Hydroxymethylglutaryl-CoA Reductase Inhibitors, Cholesterol, Cardiovascular diseases, Coronary disease, Mortality, Mediation analysis

Key points:

1. Few previous studies have attempted quantifying the relative importance of the lipid pathways in the effect of statins on coronary heart diseases, cardiovascular diseases and mortality using appropriate causal mediation approaches.
2. In this study, the effect of statins on coronary heart disease, cardiovascular disease and mortality appeared to be independent of their effect on high-density lipoprotein cholesterol and triglycerides.
3. The preventive effect of statins on coronary heart diseases could be attributed in large part to their effect on LDL.
4. The g-formula estimator we proposed is a promising approach to elucidate intermediate pathways for other drug classes

Abstract

Purpose: Estimating how much of the impact of statins on coronary heart diseases (CHD), cardiovascular disease (CVD) and mortality risk is attributable to their effect on low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL) and triglycerides.

Methods: A semi-parametric g-formula estimator together with data from the Multi-Ethnic Study of Atherosclerosis (a prospective multi-center cohort study) were utilized to perform a mediation analysis. A total of 5,280 participants, men and women of various race/ethnicities from multiple sites across the United States, were considered in the current study.

Results: The adherence adjusted total relative reduction (RRR) estimate (95% confidence interval) of statins on CHD was 14% (-16%, 37%) and the indirect component through LDL was 23% (-4%, 58%). For CVD, the total RRR was 23% (2%, 40%) and the indirect component through LDL was 5% (-13%, 25%). The total RRR of mortality was 18% (-1%, 35%) and the indirect component through LDL was -4% (-17%, 12%). The estimated indirect components through HDL and triglycerides were close to zero with narrow confidence intervals for all three outcomes.

Conclusions: The estimated effect of statins on mortality, CVD and CHD appeared to be independent of their estimated effect on HDL and triglycerides. Our study provides evidence that the preventive effect of statins on CHD could be attributed in large part to their effect on LDL. Our g-formula estimator is a promising approach to elucidate pathways, even if it is hard to make firm conclusions for the LDL-pathway on mortality and CVD.

INTRODUCTION

The effects of statins have been widely investigated, revealing their ability to reduce low-density lipoprotein cholesterol (LDL), triglycerides, the risks of coronary heart diseases (CHD), cardiovascular disease (CVD), and all-cause mortality, as well as to increase high-density lipoprotein cholesterol (HDL) [1, 2, 3, 4, 5, 6]. Statins have also been shown to have few adverse effects, relative to the size of their benefits [4, 5, 7].

Although statins were initially developed as a lipid-lowering drug class [8], it has become evident that they also have pleiotropic effects [9, 10, 11, 12, 13, 14, 15, 16, 17]. Garnering a better understanding of the mechanisms relating statin treatment to improved outcomes has been recognized as important and might "help to elucidate the full therapeutic benefits of these agents" [14].

Numerous randomized trials have reported that a larger reduction in cholesterol following statin treatment correlates with a greater reduction of CHD, CVD and mortality risk [6, 18, 19, 20, 21, 22, 23], thus suggesting a mediation effect. However, typical trials where only the treatment is randomized are susceptible to yield biased estimates of the importance of intermediate pathways, unless appropriate control is made for potential confounders and causal mediation analyses are conducted [24]. We are aware of only one study that used such causal mediation methods to quantify the importance of the cholesterol-pathway in the effect of statins [25].

The aim of the current study is to provide further evidence concerning the importance of lipid-related pathways in the effect of statins on CHD, CVD and mortality risks using causal mediation methods. More specifically, we investigated how much of the total effect of statins on the five-year risk of CHD is attributable to their effect on 1) LDL, 2) HDL and 3) triglycerides, and similarly for five-year risk of CVD and all-cause mortality. This five-year period was chosen for comparability with previous studies whose average follow-up time is often approximately five years. To perform this mediation analysis, we propose a g-formula estimator that directly accounts for censored time to events and provides results on scales that are easy to interpret.

METHODS

Data

We used data from the Multi-Ethnic Study of Atherosclerosis (MESA), a population-based prospective cohort study that enrolled 6,814 men and women aged between 45 and 84 years at baseline who were initially free of clinical cardiovascular diseases. The cohort was recruited from six Field Centers across the United-States between July 2000 and August 2002 (Exam 1). Four follow-up examinations have taken place: Exam 2 (September 2002 – February 2004), Exam 3 (March 2004 – September 2005), Exam 4 (September 2005 – May 2007) and Exam 5 (April 2010 – January 2012). MESA was approved by the institutional review board at each site, subjects provided informed consent and ethics approval for the current study has been obtained from the CHU de Québec – Université Laval research center's ethics committee. More details regarding MESA's design have already been published [26].

Study design

We analyzed the data from our observational cohort as if they arose from a sequence of four nonrandomized "trials", where each subject could participate in more than one trial [27]. This approach seeks to emulate the randomized trial gold-standard utilizing observational data in the design of the study and thus offers greater comparability to the results of randomized trials. Figure 1 depicts the study design. The entry examination for Trial 1 was MESA's Exam 2. Participants that had experienced a CVD event prior to the start of Trial 1 and those that were statin users at their pre-entry examination, at MESA's Exam 1, were not eligible for inclusion in Trial 1. The latter exclusion criterion was imposed to avoid the potential confounding due to the unmeasured untreated pre-entry cholesterol values for participants already taking statins before Trial 1. Thus, Trial 1 consisted of all the MESA's participants that didn't meet the two aforementioned exclusion criteria. The statin usage exposure and the cholesterol mediators were measured at the entry examination, whereas the potential confounders were measured at the pre-entry examination. Although the cholesterol mediator and the statin exposure were

measured simultaneously, the beginning of the statin exposure precedes the cholesterol measurement since it corresponds to medications that the subjects were taking prior to their entry examination. Follow-up for CHD, CVD and death events started at the date of entry into the trial. Trials 2, 3 and 4 were constructed similarly, respectively having their entry examination at MESA's Exams 3, 4 and 5, and having their pre-entry examination at MESA's Exams 2, 3 and 4. Note that already being a participant in a previous trial does not automatically preclude a subject from being eligible in another trial. Thus, for example, a given subject could simultaneously contribute follow-up information to all four trials if they remained CVD free and were not statin users throughout all pre-entry examinations.

CHD, CVD and All-Cause Mortality

All-cause mortality as well as CHD and CVD events have been monitored for all participants from cohort entry until the end of 2013 utilizing multiple sources of information. CHD included myocardial infarction, resuscitated cardiac arrest, definite angina, probable angina followed by revascularization and CHD death. In addition to any CHD event, CVD also included stroke, stroke death, other atherosclerotic death, and other CVD death. For more information, see the Appendix of reference [28].

LDL, HDL and Triglycerides

Blood samples were obtained at all examinations and were assayed for HDL and triglycerides. LDL was estimated using Friedewald equation [29].

Statins

The use of statins was determined by questionnaire. The participants were also asked to bring containers for all medications used in the two weeks preceding each examination [30].

Potential confounders

A rich set of 16 potential confounders was selected *a priori* (see Table 1). Fasting glucose was determined from blood samples and diabetes as fasting glucose >6.94 mmol/L. Two computed tomographies scanning of the chest of each participant were performed, the Agatston score was computed for each and the results were averaged [31]. A BMI<25 kg/m^2 was categorized as normal, 25≤BMI<30 kg/m^2 as overweight and BMI≥30 kg/m^2 as obesity [32]. Systolic and diastolic blood pressures were calculated as the average of the last two measures, of a series of three. Hypertension was defined as either having a systolic blood pressure ≥140 mmHg, a diastolic blood pressure ≥90 mmHg or taking antihypertensive medication [33].

Education, gender and race were only assessed at Exam 1. Family history of a premature cardiovascular event was determined once at Exam 2. Agatston score was assessed for all participants at Exam 1 and then up to three more times (average=1.5 times). All other covariates were measured at every examination.

Statistical analyses

We describe in more detail the analyses that were performed to decompose the estimated effect of statins on five-year CHD risk that is attributable to the effect of statins on LDL. The method used to decompose the estimated effect of statins on CHD associated with HDL or triglycerides, or to decompose the estimated effect of statins on CVD and all-cause mortality attributable to LDL, HDL or triglycerides, are analogous.

Let $A_{i,t}$ denote statin use of subject $i = 1, \ldots, n_t$ at entry in Trial $t = 1,\ 2,\ 3,\ 4$, ($A_{i,t} = 1$ if subject *i* was taking statins, and $A_{i,t} = 0$ otherwise), where n_t is the number of subjects included in Trial *t.* Let $M_{i,t}$ be the LDL of subject *i* at entry in Trial *t,* and $Y_{i,t} = 1$ if subject *i* had their first CHD in the five years following entry and $Y_{i,t} = 0$ otherwise.

We consider the potential outcome framework to causal inference to define the quantities we targeted [34]. To simplify the presentation, we henceforth drop the subscripts *i* and t. We denote by Y_{am} the value that Y would have taken if $A = a$ and $M = m$, by M_a the value that M would have taken if $A = a$, and by $Y_{aM_{a^*}}$ the value that Y would have taken if $A = a$ and $M = M_{a^*}$. For instance, Y_{1M_0} represents the counterfactual five-year incidence of CHD for a given subject if they had taken statins, but their LDL was the same as if they had not taken statins.

The average total effect of statins on five-year risk of CHD is $E\left[Y_{1M_1}\right] - E\left[Y_{0M_0}\right]$, which can be interpreted as the difference between the proportion of individuals who would have had their first CHD in the following five years if everyone had received statins and the same proportion if no one had received statins. This effect can be decomposed into a direct component, not due to the effect of statins on LDL, and an indirect component, due to the effect of statins on LDL [35]. The direct component is $E\left[Y_{1M_0}\right] - E\left[Y_{0M_0}\right]$, that is the difference between the proportion of subjects that would have had their first CHD in the following five years if everyone had received statins, but their LDL-level had been unaffected, and the same proportion if no one had received statins. The indirect component is the difference between the total effect and the direct effect.

We also considered the previous quantities on a relative risk reduction (RRR) scale, dividing by the negative of the baseline risk, -$E\left[Y_{0M_0}\right]$.

Intention to treat estimation. We first describe the analyses we conducted to obtain an observational study equivalent of intention to treat estimates where statins initiators are compared to non-initiators, regardless of whether they remained adherent to their initial treatment (taking statins or not taking statins) in follow-up visits.

We propose a semi-parametric g-formula estimator of the total effect and its components. Our estimator shares similarities with the approach of Imai et al [36]. Details of the development of our estimator are provided in Appendix 1 and simulation studies investigating its performance are presented in Appendix 2.

Letting $\boldsymbol{C}$ be the potential confounders in Table 1 and $\boldsymbol{c}_{i,t}$ be the observed values for subject *i* in trial *t*, we propose the following estimators of $E\left[Y_{aM_{a^*}}\right]$:

$$\hat{E}\left[Y_{aM_a}\right] = \frac{1}{N}\sum_{t=2}^{5}\sum_{i=1}^{n_t} \hat{\mathrm{E}}\left[Y \middle| A = a, \boldsymbol{C} = \boldsymbol{c}_{i,t}\right],$$

when $a = a^*$ and

$$\hat{E}\left[Y_{aM_{a^*}}\right] = \frac{1}{N}\sum_{t=2}^{5}\sum_{i=1}^{n_t} \int_M \hat{\mathrm{E}}\left[Y \middle| A = a, M = m, \boldsymbol{C} = \boldsymbol{c}_{i,t}\right] \hat{f}\left(m \middle| A = a^*, \boldsymbol{C} = \boldsymbol{c}_{i,t}\right) dM,$$

when $a \neq a^*$, where $N = n_2 + n_3 + n_4 + n_5$.

To compute $\hat{\mathrm{E}}\left[Y \middle| A = a, \boldsymbol{C} = \boldsymbol{c}_{i,t}\right]$, we first fitted a Cox model for the time until a CHD event that included the statin use at entry and the pre-entry confounders. We have then computed the predicted 5-year risk employing Equation (5.5) from reference [37]. This approach directly accounts for censored time-to-events due to loss to follow up or competing events through the use of the Cox model. Age was entered utilizing restricted cubic splines. Agatston score was introduced as two variables: a dummy variable equal to one if the score is greater than zero, and equal to zero otherwise, as well as an interaction between this variable and the log of the Agatston score. Time to death (CVD/CHD) was considered as right censored at the end of follow-up. Time to CHD (CVD) was further considered as right censored if death from a non-CHD (non-CVD) cause occurred before the first CHD (CVD) [38]. We did not censor follow-up time after five years of follow-up, since events occurring after this period could be informative in estimating the parameters of the Cox model. $\hat{E}\left[Y \middle| A = a, M = m, \boldsymbol{C} = \boldsymbol{c}_{i,t}\right]$ was estimated similarly, but also including LDL at entry and an interaction between statin use and LDL at entry (to simplify the methods, each mediator was investigated in a separate model).

To obtain $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\boldsymbol{c}_{i,t}\right)$, we have first fitted a linear regression where the outcome was $\log(M_t)$. This model included statin use at entry and all pre-entry potential confounders. The pre-entry values of LDL, HDL and triglycerides were all log-transformed. Age and Agatston score were included in the same manner as in the Cox model. The model further included an interaction between statin use at entry and pre-entry log(LDL). We then took $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\ \boldsymbol{c}_{i,t}\right)$ as a log-normal distribution whose mean on the log-scale was the linear predictor from the model and whose variance was the residual variance. The integral over the *M* values was performed using the trapezoidal approximation.

Assumptions of the models were visually verified and appeared reasonable for most of the range of the data. Confidence intervals (CI) were obtained through the percentile method by performing non-parametric bootstrap with 2000 resamples [39]. The bootstrap samples were taken from the original dataset to account for the within-subject correlation arising from the fact that each subject could participate in more than one trial.

Adherence-adjusted estimation. Intention to treat estimates can be problematic because they depend on the proportion of the subjects who adhere to their initial treatment [27]. We thus consider an adherence-adjusted analysis as our primary analysis. Our approach for obtaining adherence-adjusted estimates essentially consisted in discarding information on follow-up visits once a subject discontinued their initial treatment and using inverse probability weighting to redistribute the weights of such subjects onto similar subjects that pursued their initial treatment. More details are provided in Appendix 3.

Sensitivity analyses. We conducted an adherence-adjusted analysis comparing Atorvastatin (the most commonly used statin in our data) users to non-statin users to investigate if the decomposition of the effect might vary according to type of statins. We have also performed two sensitivity analyses where alcohol consumption (yes or no) and self-reported walking pace (in 5 categories ranging from very slow to brisk), as a measure of physical activity [40], were considered as additional confounders. These are considered as sensitivity rather than main analyses because of the amount of missing data on these variables (≈10% for alcohol, walking

pace was not collected at Exam 4). Sensitivity analyses yielded similar results to those presented below (see Appendix 4).

Missing data. A single Expectation-Maximization imputation of missing Agatston scores was performed (correlations between 0.86 and 0.99 were observed between measures taken at Exam 1 and measures taken at other Exams, not presented). Otherwise, only rows without any missing data were considered. Available data is described in Figure 2.

RESULTS

Pre-entry characteristics of the participants in the trials are reported in Table 1 (characteristics by trial are reported in Appendix 5). A total of 5,280 subjects participated in at least one of our trials, of which 336 experienced a CHD event, 509 a CVD event and 638 died before the end of follow-up. A total of 4402 subjects participated in multiple trials; the average number of trials to which subjects participated was 2.9. The average follow-up times for trials 1, 2, 3 and 4 were respectively 8.3, 7.5, 6.8 and 2.9 years, after censoring observations that discontinued their initial treatment. The adherence proportion after 1, 2 and 3 follow-up visits were respectively of 86%, 85% and 81%.

Below, we summarize the adherence-adjusted estimates for the decomposition of the effect of statins on all-cause mortality, CVD and CHD. The complete results are presented in Table 2. The intention to treat estimates are reported in Appendix 6.

The total adherence-adjusted association between statin use and CHD was -0.4% (95% CI = -1.1%, 0.4%) on the risk difference (RD) scale and 14% (-16%, 37%) on the RRR scale. A large portion, if not all, of this association might be due to the LDL-lowering properties of statins (RD = -0.7%, 95% CI = -1.7%, 0.1%; RRR = 23%, 95% CI = -4%, 58%). The total association between statins and CVD was -1.0% (-1.8%, -0.1%) on the RD scale and 23% (2%, 40%) on the RRR scale. A small portion of this association was estimated to be attributable to the LDL pathway, but wide confidence intervals were obtained (RD = -0.2%, 95% CI = -1.1%, 0.5%;

RRR = 5%, 95% CI = -13%, 25%). The total association between statin use and mortality was -0.9% (95% CI = -1.8%, 0.0%) on a RD scale and 18% (-1%, 35%) on the RRR scale. Results were inconclusive regarding the importance of the LDL-pathway (RD = 0.2%, 95% CI = -0.6%, 0.8%; RRR = -4%, 95% CI = -17%, 12%). For all three outcomes, the portion attributable to the HDL or triglycerides pathways was estimated to be negligible.

DISCUSSION

Our results provide evidence of the importance of the LDL-lowering properties of statins on reducing the risk of CHD. In fact, our indirect effect estimate is larger than the total effect estimate. This suggests that statins decrease the risk of CHD because of their LDL-lowering properties, but slightly increase the risk through other pathways; the net effect remaining beneficial. Our study also provides evidence that the HDL- and triglycerides-pathways play a minor role in the effect of statins on CHD, CVD and mortality. Regarding the importance of the LDL-pathway on CVD or mortality, our study is somewhat inconclusive: small indirect associations were observed, but wide confidence intervals were obtained. An expanded discussion of the substantive results is provided in Appendix 7.

It is important to take into account a number of potential limitations when interpreting our results. First, because we considered observational data, any causal inference made rests upon the assumptions of no unmeasured confounding. Since we considered a rich set of potential confounders, which was built based on substantive knowledge, it appears reasonable to assume that the most important confounding factors were accounted for. Moreover, our total effect estimates are consistent with those produced by large meta-analyses of randomized controlled trials [2, 5].

The validity of our results also relies on the assumptions of the models that were used to construct our g-formula estimator, such as the proportional hazard assumption of the Cox model. These assumptions were visually verified and appeared to be reasonable.

Some bias could have been introduced by our treatment of missing data. A better alternative might have been to perform multiple imputations [41]. However, this option was thought to be impractical to implement, due to the computational burden of combining bootstrapping with multiple imputations.

Although we know that the initiation of our statin exposure precedes the measure of the cholesterol mediators, we do not know the precise time when initiation occurred. Some bias could arise if the delay between initiation and measurement of the cholesterol mediator is too short in some cases for the full impact of statin treatment on cholesterol to be realized. However, the mean LDL-reduction associated with statin treatment we observed was very close to the one reported in a large meta-analysis of randomized trials (0.93 mmol/L vs 1.00 mmol/L – analysis not presented) [5], which suggests that this bias might be small.

Our adherence-adjusted analysis attempted to correct for the misclassification that occurs when a user becomes a non-user or vice-versa. However, the exact time where switching occurs is unknown. Therefore, some individuals have contributed follow-up time for outcome events in the wrong exposure group before being censored, which is likely to have biased our effect estimates towards the null. Because the adherence proportion after one visit is high (84%) and because our total effect point estimates are consistent with previous findings, we hypothesize that this misclassification bias may be small. Moreover, our analysis only investigated the contribution of statins through short-term changes in cholesterol, not accounting for how cholesterol varies over time. This is similar to the approach used in many clinical trials where the reduction in cardiovascular events is correlated with the reduction in cholesterol after one or two years [6, 20, 23]. It is nonetheless possible that the indirect effect of statins on cardiovascular outcomes through cholesterol depends both on short- and long-term effects on cholesterol. Recently, a parametric and a semi-parametric g-formula approach have been proposed to perform mediation analysis with time-varying exposures and mediators and could be considered to provide further insights on the mechanism relating statin to better cardiovascular outcomes [42, 43]. However, accounting for the time-varying nature of the variables comes at the cost of adding further complexities to the analysis. For instance, the parametric g-formula approach of Lin et al [42] depends on more stringent assumptions than ours, such as the

correct parametric specification of the joint distribution of the exposure and confounders at each time-point whereas our method uses a non-parametric estimator for the joint distribution of the confounders. While the semi-parametric method of VanderWeele and Tchetgen Tchetgen [43] avoids making such assumptions, it is potentially unstable in certain circumstances [42].

Our study also has a number of noteworthy strengths. The generalizability of the results to the American population is improved by the fact that we used data from a multi-ethnic population-based cohort with participants from multiple centers across the United States. The large number of participants, the length of the follow-up period and the relatively small number of losses-to-follow-up for a study of that length should also be noted. From a methodological perspective, we believe that our semi-parametric g-formula estimator is a promising approach to help elucidate intermediate pathways, since it has the ability of directly accounting for censored time to events through the use of the Cox model and avoids relying on stringent parametric assumptions. It also provides estimates on risk difference or relative risk reduction scales that are easy to interpret. Our approach is of particular interest for investigating exposure effects attributable to intermediate pathways in a relatively specific time-frame, such as short-term effects on the mediator.

Funding: This research was supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute and by grants UL1-TR-000040 and UL1-TR-001079 from NCRR. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Sharing: The R code utilized for conducting the analyses is available from the corresponding author upon request.

Acknowledgments: DT thanks Université Laval for sponsoring his visit at the University of Washington. This manuscript has been reviewed by MESA for scientific content and consistency of data interpretation with previous MESA publications.

References

1. Simes J, Furberg C, Barunwald E, Davis B, Ford I, Tonkin A, Sheperd J. Effects of pravastatin on mortality in patients with and without coronary heart disease across a broad range of cholesterol levels. *Eur Heart J*. 2002;23(3):207-215.

2. LaRosa J, He J, Vupputuri S. Effect of statins on risk of coronary disease: Meta-analysis of randomized controlled trials. *JAMA*. 1999;282(24):2340-2346.

3. Law M, Wald N, Rudnicka A. Quantifying effect of statins on low density lipoprotein cholesterol, ischeamic heart disease, and stroke: systematic review and meta-analysis. *BMJ*. 2003;326(7404):1423-1427.

4. Maron D, Fazio S, Linton M. Current perspectives on statins. *Circulation*. 2000;101(2):207-213.

5. Taylor F, Huffman MD, Macedo AF, Moore THM, Burke M, Smith D, Ward K, Ebrahim S. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013(1).

6. Cholesterol Treatment Trialists' (CTT) Collaborators. The effects of lowering LDL cholesterol with statin therapy in people at low riks of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet*. 2012;380(9841):581-590.

7. Smeeth L, Douglas I, Hall A, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. *Br J Clin Pharmacol*. 2009;67(1):99-109.

8. Tobert JA. Lovastatin and beyond: The history of HMG-CoA reductase inhibitos. *Nat Rev Drug Discov*. 2003;2:517-526.

9. Wheeler DC. Are there potential non-lipid-lowering uses of statins? *Drugs*. 1998;56(4):517-522.

10. Jain MK, Ridker PM. Anti-inflammatory effects of statins: Clinical evidence and basic mechanisms. *Nature reviews drug discovery*. 2005;4:977-987.

11. Williams JK, Sukhova GK, Herrington DM, Libby P. Pravastatin has cholesterol-lowering independent effects on the artery wall of atherosclerotic monkeys. *J Am Coll Cardiol*. 1998;31(3):684-691.

12. Bonetti PO, Wilson SH, Rodriguez-Porcel M, Holmes DR, Lerman LO, Lerman A. Simvastatin preserves myocardial permeability in experimental hypercholesterolemia independent of lipid lowering. *Journal of the American college of cardiology*. 2002;40(3):546-554.

13. Wilson SH, Simari RD, Best PJ, Peterson TE, Lerman LO, Aviram M, Nath KA, Holmes DR, Amir L. Simvastatin preserves coronary endothelial function in hypercholesterolemia in the absence of lipid lowering. *Arterioscler Thromb Vasc Biol*. 2001;21:122-128.

14. Liao JK. Beyond lipid lowering: the role of statins in vascular protection. *Int J Cardiol*. 2002;86(1):5-18.

15. Liao JK. Isoprenoids as mediators of the biological effects of statins. *J Clin Invest*. 2002;110(3):285-288.

16. Davignon J, Laaksonen R. Low-density lipoprotein-independent effects of statins. *Curr Opin Lipidol*. 1999;10(6):543-559.

17. McFarlane SI, Muniyappa R, Francisco R, Sowers JR. Pleiotropic effects of statins: lipid reduction and beyond. *J Clin Endocrinol Metab*. 2002;87(4):1451-1458.

18. Gould AL, Rossouw JE, Santanello NC, Heyse JF, Furberg CD. Cholesterol Reduction Yields Clinical Benefit. Impact of Statin Trials. *Circulation*. 1998;97:946-952.

19. Everette BM, Glynn RJ, MacFadyen JG, Ridker PM. Rosuvastatin in the prevention of stroke among men and women with elevated levels of C-reactive protein. Justification for the use of statins in prevention: An intervention trial evaluating rosuvastatin (JUPITER). *Circulation*. 2010;121:143-150.

20. Cholesterol Treatment Trialists' Collaborators. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *Lancet*. 2005;366(9493):11267-1278.

21. Ridker PM, Cannon CP, Morrow D, Rifai N, Rose LM, McCabe CH, Pfeffer MA, Braunwald E. C-reactive protein levels and outcomes after statin therapy. *N Engl J Med*. 2005;352(1):20-28.

22. Nissen SE, Tuzcu EM, Schoenhagen P, Crowe T, Sasiela WJ, Tsai J, Orazem J, Magorien RD, O'Shaughnessy C, Ganz P. Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *N Engl J Med*. 2005;352(1):29-38.

23. Robinson JG, Smith B, Maheshwari N, Schrott H. Pleiotropic effects of statins: benefit beyond cholesterol reduction? A meta-regression analysis. *J Am Coll Cardiol*. 2005;46(10):1855-1862.

24. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137-150.

25. Shinozaki T, Matsuyama Y, Ohashi Y. Estimation of controlled direct effects in time-varying treatments using structural nested mean models: application to a primary prevention trial for coronary events with pravastatin. *Stat Med*. 2014;33(18):3214-3228.

26. Bild DE, Bluemeke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacobs DRJ, Kronmal R, Liu K, Clark Nelson J, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol*. 2002;156(9):871-881.

27. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Stampfer MJ, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. November 2008;19(6):766-779.

28. Bluemke DA, Kronmal RA, Lima JA, Liu K, Olson J, Burke GL, Folsom AR. The relationship of left ventricular mass and geometry to incident cardiovascular events. *J Am Coll Cardiol*. 2008;52(25):2148-2155.

29. Fridewald WT, Levy RI, Frederickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, whitout use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.

30. Smith NL, Psaty BM, Heckbert SR, Tracy RP, Cornell ES. The reliability of medication inventory methods compared to serum levels of cardiovascular drugs in the elderly. *J Clin Epidemiol*. 1999;52(2):143-146.

31. Carr JJ, Nelson JC, McNitt-Gray M, Arad Y, Jacobs DRJ, Sidney S, Bild DE, Williams OD, Detrano RC. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: Standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Radiology*. 2005;234(1):35-43.

32. World Health Organization. Physical status: the use and interpretation of anthropometry: report of a WHO Expert Committee. *World Health Organ Tech Ser Rep*. 1995;854:1-452.

33. Joint National Committee on Detection. The sixth report of the joint national committee on detection, evaluation, and treatment of high blood pressure. *Arch Intern Med*. 1997;157:2413-2446.

34. Rubin DB. Causal inference using potential outcomes: Design, Modeling, Decisions. *J am Stat Associ*. 2005;100(469):322-331.

35. VanderWeele. A unification of mediation and interaction: A 4-way decomposition. *Epidemiology*. 2014;25(5):749-761.

36. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):209-334.

37. Efron B. The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc*. 1977;72(359):557-565.

38. Bakoyannis G, Touloumi G. Practical methods for competing risks data: A review. *Stat Methods Med Res*. 2010;21(3):257-272.

39. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC; 1993.

40. Bertoni AG, Whitt-Glover MC, Chung H, Le KY, Barr RG, Mahesh M, Jenny NS, Burke GL, Jacobs DR. The association between physical activity and subclinical atherosclerosis: The Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol*. 2008;169(4):444-454.

41. Little RJ, Rubin DB. *Statistical analysis with missing data, 2nd Edition*. New Jersey: John Wiley & Sons; 2002.

42. Lin SH, Young J, Logan R, Tchetgen Tchetgen EJ, VanderWeele TJ. Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. *Epidemiol*. 2017;28(2):266-274.

43. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *J R Stat Soc Series B Stat Methodol*. 2017;79(3):917-938.

44. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the national cholesterol education program expert panel on detection, evaluation, and treatment of high blood cholesterol in adults. *JAMA*. 2001;285(19):2486-2497.

45. Schaefer EJ, McNamara JR, Tayler T, Daly JA, Gleason JL, Seman LJ, Ferrari A, Rubenstein JJ. Comparisons of effects of statins (atorvastatin, fluvastatin, lovastatin, pravastatin, and simvastatin) on fasting and postprandial lipoproteins in patiens with coronary heart disease versus control subjects. *Am J Cardiol*. 2004;93(1):31-39.

Table 1. Pre-entry Characteristics of Participants in the Trials According to Statin Usage at the Entry Examination

	Non-Users (n = 13,702)	Users (n = 1,651)
Age, mean (SD)	62.5 (10.1)	64.8 (9.3)
Education (highest level reached)		
- Less than high school	15.3	18.1
- High school	16.6	19.1
- More than high school	67.8	62.8
Gender female	53.8	54.7
Health insurance	92.2	95.1
Race/ethnicity		
- White, Caucasian	38.4	38.2
- Chinese American	12.4	10.1
- Black, African-American	27.2	28.1
- Hispanic	21.9	23.6
Site		
- WFU	15.3	16.6
- COL	17.2	16.5
- JHU	13.6	15.4
- UMN	15.9	16.1
- NWU	19.3	15.5
- UCLA	19.3	15.5
LDL in mmol/L, mean (SD)	3.0 (0.8)	3.5 (0.9)
HDL in mmol/L, mean (SD)	1.4 (0.4)	1.3 (0.4)
Triglycerides in mmoL/L, mean (SD)	1.3 (0.7)	1.6 (0.8)
Agatston score > 0	41.3	53.5
Ln of Agatston score when > 0, mean (SD)	4.1 (1.8)	4.4 (1.8)
BMI		
- Normal	32.0	24.5
- Overweight	38.0	38.2
- Obesity	30.0	37.3
Cigarette smoking		
- Never	48.3	48.0
- Former	40.0	42.9
- Current	11.7	9.2
Diabetes	9.0	22.4
Family history of CVD	25.9	28.9
Hypertension	39.9	57.1
Number of medications, mean (SD)	4.0 (3.6)	4.9 (3.6)

Results are expressed as percentage unless otherwise indicated

Table 2. Decomposition of the Adherence-adjusted Effect Estimate of Statins on Coronary Heart Disease, Cardiovascular Disease and Mortality (95% confidence intervals in parenthesis)

	Measure	Total effect of statins	Direct effect of statins	Indirect effect of statins (via the mediator)
CHD				
LDL	RD	-0.4% (-1.1%, 0.4%)	0.3% (-0.9%, 1.7%)	-0.7% (-1.7%, 0.1%)
	RRR	14% (-16%, 37%)	-9% (-59%, 29%)	23% (-4%, 58%)
HDL	RD	-0.4% (-1.1%, 0.4%)	-0.4% (-1.1%, 0.4%)	0.0% (-0.1%, 0.0%)
	RRR	14% (-16%, 37%)	13% (-16%, 36%)	1% (-1%, 3%)
Trig.	RD	-0.4% (-1.1%, 0.4%)	-0.4% (-1.2%, 0.4%)	0.0% (-0.1%, 0.1%)
	RRR	14% (-16%, 37%)	14% (-15%, 37%)	0% (-4%, 3%)
CVD				
LDL	RD	-1.0% (-1.8%, -0.1%)	-0.7% (-1.9%, 0.6%)	-0.2% (-1.1%, 0.5%)
	RRR	23% (2%, 40%)	17% (-14%, 42%)	5% (-13%, 25%)
HDL	RD	-1.0% (-1.8%, -0.1%)	-0.9% (-1.8%, 0.0%)	0.0% (-0.1%, 0.0%)
	RRR	23% (2%, 40%)	22% (0%, 40%)	1% (0%, 3%)
Trig.	RD	-1.0% (-1.8%, -0.1%)	-1.0% (-1.8%, 0.0%)	0.0% (-0.1%, 0.1%)
	RRR	23% (2%, 40%)	22% (1%, 40%)	0% (-2%, 2%)
Mortality				
LDL	RD	-0.9% (-1.8%, 0.0%)	-1.1% (-2.1%, 0.0%)	0.2% (-0.6%, 0.8%)
	RRR	18% (1%, 35%)	23% (0%, 44%)	-4% (-17%, 12%)
HDL	RD	-0.9% (-1.8%, 0.0%)	-0.8% (-1.7%, 0.1%)	-0.1% (-0.2%, 0.0%)
	RRR	18% (-1%, 35%)	17% (-3%, 34%)	2% (0%, 4%)
Trig.	RD	-0.9% (-1.8%, 0.0%)	-0.9% (-1.8%, 0.0%)	0.0% (-0.1%, 0.1%)
	RRR	18% (-1%, 35%)	19% (0%, 35%)	-1% (-3%, 1%)

The total estimated effect is decomposed in an indirect effect, due to the effect of statins on a given mediator, and a direct component, not due to the effect of statins on that given mediator. LDL = low-density lipoprotein, HDL = high-density lipoprotein, Trig. = triglycerides, RD = risk difference, RRR = relative risk reduction, CHD = coronary heart diseases, CVD = cardiovascular diseases.

Estimates are adjusted for age, education, gender, health insurance, race/ethnicity, Field Center, previous LDL, previous HDL, previous triglycerides, Agatston score, BMI, cigarette smoking, diabetes, family history of CVD, hypertension and total number of medication.

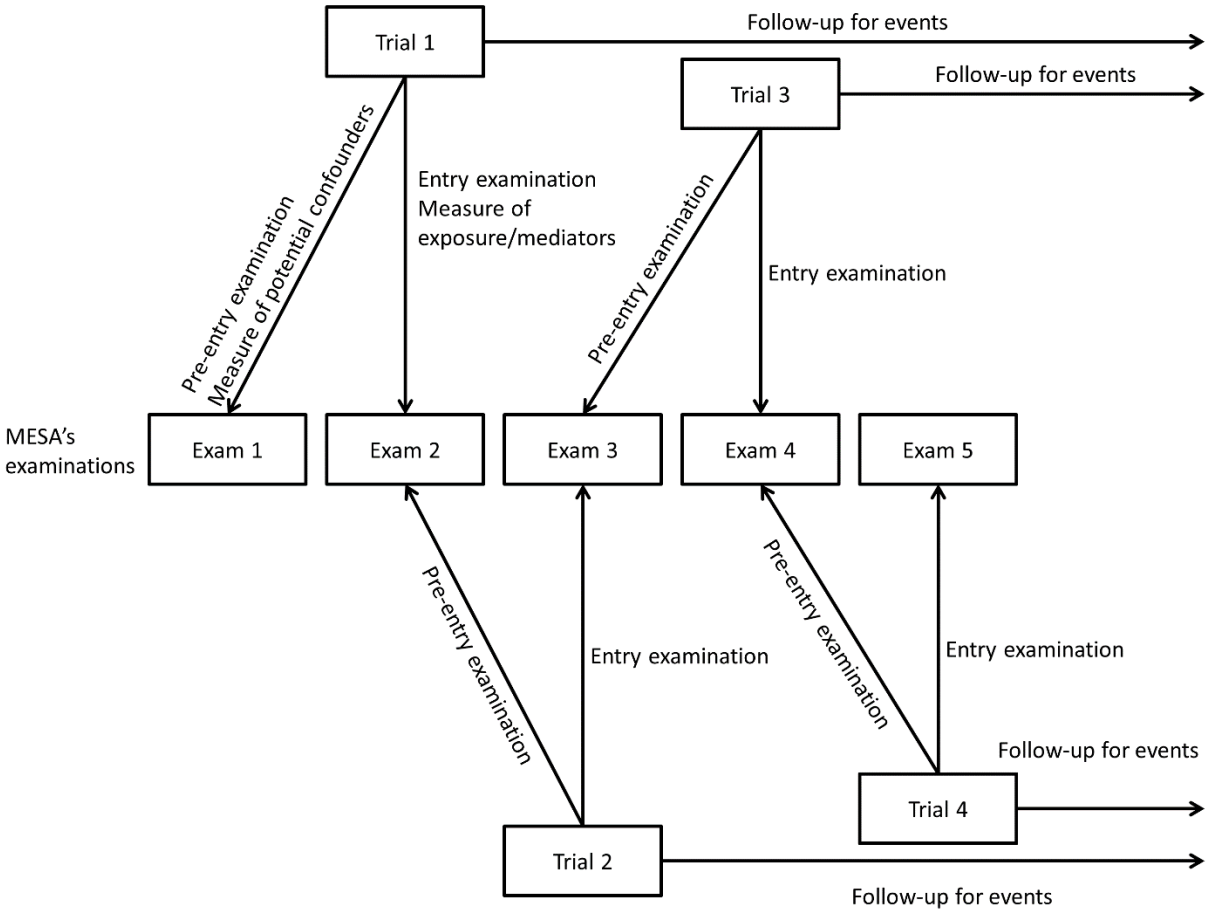

Figure 1: Schematic representation of the study design

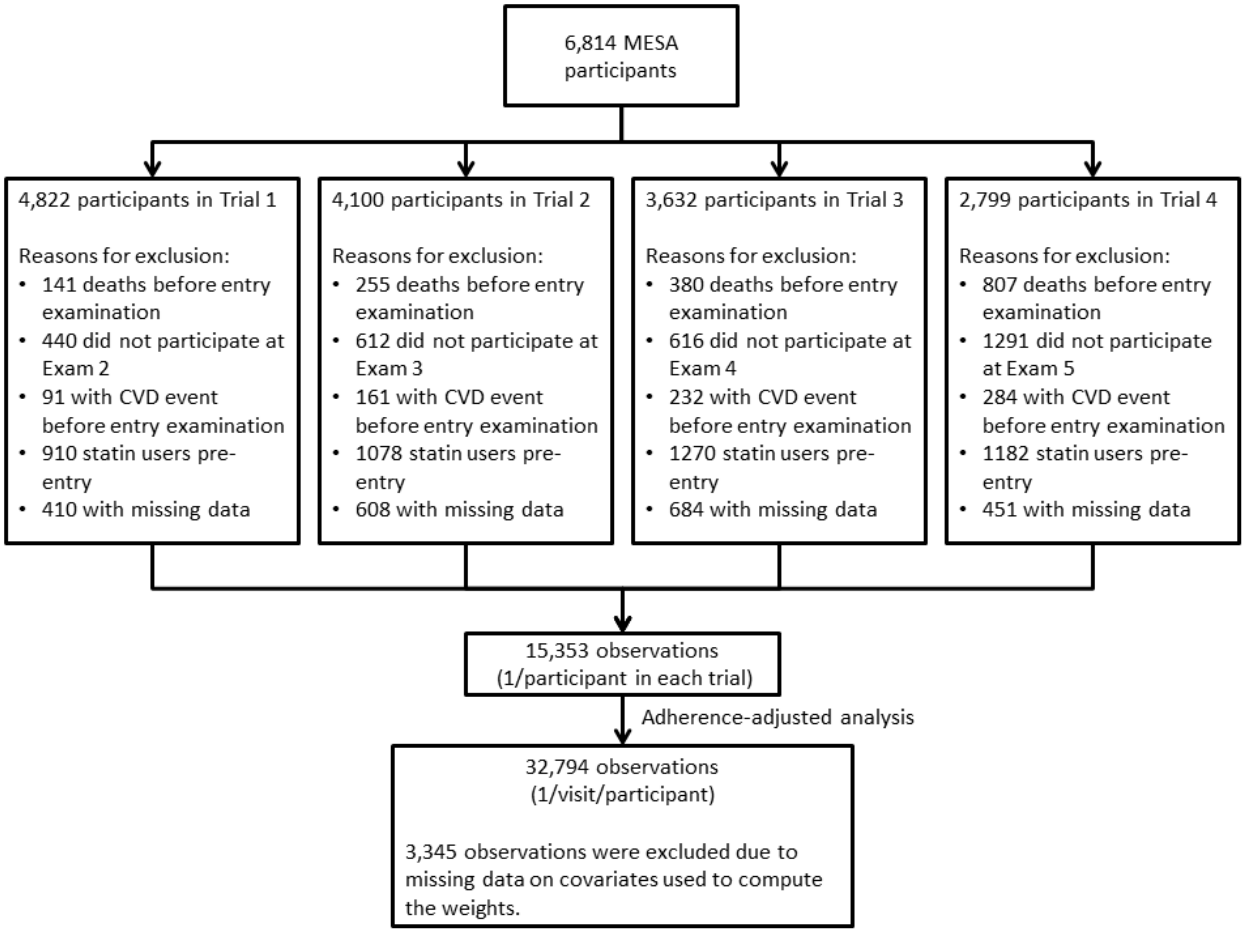

Figure 2: Flowchart of the available data

Appendices

Appendix 1: Development of the estimators of $E\left[Y_{aM_{a^*}}\right]$

We explain how the counterfactuals $E\left[Y_{aM_{a^*}}\right]$ can be estimated from the observed data in order to estimate the total effect and its direct and indirect components. The development of the estimators presented in our paper is based on the 6 following assumptions (where ⫫ denotes statistical independence):

1) $\boldsymbol{C}$ contains all confounders of the effect of statins on CHD ($Y_{am} \perp\!\!\!\perp A|\boldsymbol{C}$);
2) Statins and $\boldsymbol{C}$ contains all confounders of the effect of LDL on CHD ($Y_{am} \perp\!\!\!\perp M|A, \boldsymbol{C}$);
3) $\boldsymbol{C}$ contains all confounders of the effect of statins on LDL ($M_a \perp\!\!\!\perp A|\boldsymbol{C}$);
4) None of the LDL-CHD confounders are affected by statins ($Y_{am} \perp\!\!\!\perp M_a^*|\boldsymbol{C}$);
5) When statin exposure takes the value *a* and the LDL level is *m*, the observed five-year survival *Y* corresponds to the counterfactual five-year survival Y_{am} ($Y_{am} = Y$ if $A = a$ and $M = m$);
6) When statin exposure is *a,* the observed LDL level *M* corresponds to the counterfactual LDL level M_a ($M_a = M$ if $A = a$)

Using these 6 assumptions, if $a = a^*$:

$$E\left[Y_{aM_a}\right] = \int_{\boldsymbol{c}} E\left[Y_{aM_a}|\boldsymbol{C}\right] f(\boldsymbol{c}) dF_{\boldsymbol{c}}$$

Using the previous assumptions we get:

$$E[Y|A=a,\boldsymbol{C}] = E\left[Y_{aM_a}\middle|A=a,\boldsymbol{C}\right] \qquad \text{(assumption 5)}$$

$$= E\left[Y_{aM_a}\middle|\boldsymbol{C}\right] \qquad \text{(assumption 1)}$$

Thus, $E\left[Y_{aM_a}\right] = \int_{\boldsymbol{c}} E[Y|A=a,\boldsymbol{C}]f(\boldsymbol{c})dF_{\boldsymbol{c}}$

If $a \neq a^*$:

$E\left[Y_{aM_{a^*}}\right] = \int_{\boldsymbol{c}} E\left[Y_{aM_{a^*}}\middle|\boldsymbol{C}\right]f(\boldsymbol{c})dF_{\boldsymbol{c}}$, where

$$E\left[Y_{aM_{a^*}}\middle|\boldsymbol{C}\right] = \int_{M_{a^*}} E\left[Y_{aM_{a^*}}\middle|\boldsymbol{C},M_{a^*}=m\right] f_{M_{a^*}|\boldsymbol{C}}(m|\boldsymbol{C})dF_{M_{a^*}}$$

$$= \int_{M_{a^*}} E[Y_{am}|\boldsymbol{C},M_{a^*}=m]\, f_{M_{a^*}|\boldsymbol{C}}(m|\boldsymbol{C})dF_{M_{a^*}}$$

$$= \int_{M_{a^*}} E[Y_{am}|\boldsymbol{C}]\, f_{M_{a^*}|C}(m|\boldsymbol{C})dF_{M_{a^*}}. \qquad \text{(assumption 4)}$$

Using the previous assumptions we get:

$$E[Y|A=a,M=m,\boldsymbol{C}] = E[Y_{am}|A=a,M=m,\boldsymbol{C}] \qquad \text{(assumption 5)}$$

$$= E[Y_{am}|A=a,\boldsymbol{C}] \qquad \text{(assumption 2)}$$

$$= E[Y_{am}|\boldsymbol{C}], \qquad \text{(assumption 1)}$$

and

$$f_{M|A,\boldsymbol{C}}(m|A=a^*,\boldsymbol{C}) = f_{M_{a^*}|A,\boldsymbol{C}}(m|A=a^*,\boldsymbol{C}) \qquad \text{(assumption 6)}$$

$$= f_{M_{a^*}|\boldsymbol{C}}(m|\,\boldsymbol{C}). \qquad \text{(assumption 3)}$$

Thus, we have

$$\mathrm{E}\left[Y_{aM_{a^*}}\right] = \int_{\boldsymbol{c}} \int_{M} E[Y|A=a, M=m, \boldsymbol{C}]\, f_{M|A,\boldsymbol{C}}(m|A=a^*, \boldsymbol{C})\, f(\boldsymbol{c}) dF_M dF_{\boldsymbol{c}}.$$

To complete the development of the estimators, we plug in estimators of $E[Y|A=a, \boldsymbol{C}]$ and $f(\boldsymbol{c})$ in $\int_{\boldsymbol{c}} E[Y|A=a, \boldsymbol{C}] f(\boldsymbol{c}) dF_{\boldsymbol{c}}$, as well as estimators of $E[Y|A=a, M=m, \boldsymbol{C}]$, $f_{M|A,\boldsymbol{C}}(m|A=a^*, \boldsymbol{C})$ and $f(\boldsymbol{c})$ in $\int_{\boldsymbol{c}} \int_{M} E[Y|A=a, M=m, \boldsymbol{C}]\, f_{M|A,C}(m|A=a^*, \boldsymbol{C})\, f(\boldsymbol{c}) dF_M dF_{\boldsymbol{c}}$. The estimators of $E[Y|A=a, \boldsymbol{C}]$, $E[Y|A=a, M=m, \boldsymbol{C}]$ and $f_{M|A,\boldsymbol{C}}(m|A=a^*, \boldsymbol{C})$ are presented in our paper. To estimate $f(\boldsymbol{c})$ we utilized a non-parametric estimator that consists in calculating the proportion of subjects for which $\boldsymbol{C} = \boldsymbol{c}$. Computing the integral over $\boldsymbol{C}$ is then mathematically equivalent to computing the mean of the quantity inside the integral over all participants in all trials, that is, the integral over $\boldsymbol{C}$ can be replaced by a sum over subjects divided by the number of subjects.

Appendix 2: Empirical investigation of the performance of our estimator

Simulation studies were performed to investigate the empirical performance of our semi-parametric g-formula estimator. Four different simulation scenarios were devised to investigate the performance under different circumstances. We first present the general framework of our simulation study, then provide specific details and finally present the results of our simulation study. Scenarios 1-3 feature a single time-point, whereas Scenario 4 features two time-points. The latter scenario thus allows for investigating the performance of the approach we have used for producing adherence-adjusted inferences (see Appendix 3).

General framework:

Scenarios 1-3: In each of these scenarios, we have first simulated a population of $n = 1{,}000{,}000$ individuals. For each subject in this population, we simulated a continuous confounder C and a binary exposure A. We then simulated two counterfactual mediators M_1 and M_0 that correspond to the value that the mediator would have taken had A taken the value 1 or 0, respectively. Next, we simulated four counterfactual time to event $T_{aM_{a^*}}$ variables, from which we constructed four binary counterfactual outcome variables indicating whether the event would have happened inside a five-year window ($Y_{aM_{a^*}} = 1$ if $T_{aM_{a^*}} < 5$ and $Y_{aM_{a^*}} = 0$ if $T_{aM_{a^*}} \geq 5$). Since, by design, we had access to all four counterfactual outcome variables for each subject, the total effect, its direct component and its indirect component were directly computed, both on the risk difference scale and the RRR scale.

We then simulated a sample of size $n = 500$ using the same equations as for the population simulation. However, we additionally simulated a random censoring time (e.g., due to lost to follow-up or competing events). We then estimated the total effect and its components using the observed version (not the counterfactual) of the mediator and the time to event variables with the estimator presented in our paper.

Scenario 4: The general framework for Scenario 4 is similar to the one for Scenarios 1-3, but features a second time-point that is set exactly three years after the first time-point. As such, the design of this simulation is more complicated. As in Scenarios 1-3, we have first simulated a population of $n = 1{,}000{,}000$ individuals. At the first time-point, we simulated C, A, M_1, M_0 for each subject, similarly to Scenarios 1-3. We then simulated four provisional counterfactual time to event $T_{aM_{a^*}}$ variables. These provisional times to event represent the time between the first time-point and either the second time-point or the event of interest, whichever came first, and thus took a maximum value of three years for subjects who did not experience the event before the second time-point. To distinguish variables at the second time-point from variables at the first time-point, we add an apostrophe to the former. At the second time-point, we considered four counterfactual confounder covariates $C'_{aM_{a^*}}$ and four counterfactual exposure level $A'_{aM_a^*}$ corresponding to what would have happened if the subject had experienced exposure level a and had their mediator taken the value M_{a^*} at the first time-point. We also simulated a new observed exposure level, A', at the second time-point. Finally, we simulated eight counterfactual mediator variables, $M'_{aM_{a^*}a}$, and eight counterfactual residual time to event, $T'_{aM_{a^*a}}$, corresponding to what would have happened if the subject had experienced exposure level a at the first time-point, their mediator taken the value M_{a^*} at the

first time-point, and experienced the exposure level a' at the second time-point. The total counterfactual time to event was equal to the sum of the provisional and the residual time to event when the counterfactual provisional time to event allowed participating at the second time-point, and was equal to the counterfactual provisional survival time otherwise. As in Scenarios 1-3, we then constructed eight binary counterfactual outcome variables indicating whether the event would have happened inside a five-year window ($Y_{aM_{a^*}a'}$). The total effect, its direct component and its indirect component were directly computed, both on the risk difference scale and the RRR scale utilizing the simulated counterfactuals. Note that in this scenario, the effects of interest compared counterfactual situations where the subjects remained adherent to their initial treatment (where $a = a'$).

We then simulated a sample of size $n = 500$ using the same equations as for the population simulation. However, we additionally simulated provisional and residual random censoring times (e.g., due to lost to follow-up or competing events). We then estimated the total effect and its components using the adherence-adjusted approach we have proposed in this paper.

Summarizing the results: Confidence intervals were obtained through the percentile method by performing non-parametric bootstrap with 500 resamples. A total of 1,000 replications of the sample simulation were performed. Finally, we computed the bias (comparing the mean sample estimated values the true population values) and coverage probability of the 95% confidence intervals (percentage of the time the sample confidence intervals contained the true population values).

Scenarios' details

Scenario 1 aimed to investigate the performance of our approach under ideal circumstances, where the models used to decompose the total effect are correctly specified. The population data generating equations are:

$$C = \varepsilon_C$$

$$P(A = 1) = expit(1 + 0.5C)$$

$$M_0 = 0.4 \times 0 - 0.6C + \varepsilon_{M_0}$$

$$M_1 = 0.4 \times 1 - 0.6C + \varepsilon_{M_1}$$

$$\log(T_{0M_0}) = \log(6 + 0.4C + 0.3M_0 + 0.5 \times 0 + 0.1M_0 \times 0) + \log(1.5)\,\varepsilon_{T_{0M_0}}$$

$$\log(T_{1M_0}) = \log(6 + 0.4C + 0.3M_0 + 0.5 \times 1 + 0.1M_0 \times 1) + \log(1.5)\,\varepsilon_{T_{1M_0}}$$

$$\log(T_{0M_1}) = \log(6 + 0.4C + 0.3M_1 + 0.5 \times 0 + 0.1M_1 \times 0) + \log(1.5)\,\varepsilon_{T_{0M_1}}$$

$$\log(T_{1M_1}) = \log(6 + 0.4C + 0.3M_1 + 0.5 \times 1 + 0.1M_1 \times 1) + \log(1.5)\,\varepsilon_{T_{1M_1}}$$

where $\varepsilon_C, \varepsilon_{M_0}, \varepsilon_{M_1}, \varepsilon_{T_{0M_0}}, \varepsilon_{T_{1M_0}}, \varepsilon_{T_{0M_1}}, \varepsilon_{T_{1M_1}}$ are all independent $N(0,1)$ and $expit(x) = \frac{\exp(x)}{1+\exp(x)}$.

The sample data-generating equations are the same, but include the additional equations:

$$M = M_0(1 - A) + M_1 A$$

$$T = T_{0M_0}(1 - A) + T_{1M_1} A$$

$$T_C = \min(\varepsilon_{T_C}, 6)$$

where T_C is a random censoring time and $\log(\varepsilon_{T_C}) \sim N(\log(5), \log(1.5))$. Observations for which $T_C < T$ have their follow-up time censored at T_C.

To compute $\hat{\mathrm{E}}\left[Y\middle|A=a,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$, we fitted a Cox proportional hazard model that included A and C. Similarly, to compute $\hat{\mathrm{E}}\left[Y\middle|A=a,M=m,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$, we fitted a Cox model that included A, M, C and an interaction term between A and M. We computed $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\boldsymbol{c}_{i,t}\right)$ using the output of a linear regression of M according to A and C. We then took $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\ \boldsymbol{c}_{i,t}\right)$ as a normal distribution whose mean was the linear predictor from the model and whose variance was the residual variance.

Scenario 2 aimed at investigating the property of our approach under the null hypothesis of no indirect effect. This scenario uses the same data generating equations as Scenario 1, except for the equations generating the survival times that are modified as follows:

$$\log(T_{0M_0}) = \log(6 + 0.4C + 0M_0 + 0.5 \times 0 + 0M_0 \times 0) + \log(1.5)\,\varepsilon_{T_{0M_0}}$$

$$\log(T_{1M_0}) = \log(6 + 0.4C + 0M_0 + 0.5 \times 1 + 0M_0 \times 1) + \log(1.5)\,\varepsilon_{T_{1M_0}}$$

$$\log(T_{0M_1}) = \log(6 + 0.4C + 0M_1 + 0.5 \times 0 + 0M_1 \times 0) + \log(1.5)\,\varepsilon_{T_{0M_1}}$$

$$\log(T_{1M_1}) = \log(6 + 0.4C + 0M_1 + 0.5 \times 1 + 0M_1 \times 1) + \log(1.5)\,\varepsilon_{T_{1M_1}}$$

The models used to decompose the effect are also the same as in scenario 1.

Scenario 3 was devised to investigate the sensitivity of our approach to mild misspecifications of the models used to decompose the total effect. The following data-generating equations are modified as compared with Scenario 1:

$$M_0 = 0.4 \times 0 - 0.6C + 0.2C^2 + \varepsilon_{M_0}$$

$$M_1 = 0.4 \times 1 - 0.6C + 0.2C^2 + \varepsilon_{M_1}$$

$$\log(T_{0M_0}) = \log(6 + 0.4C + 0.3M_0 + 0.5 \times 0 + 0.1M_0 \times 0 - 0.2C \times M_0) + \log(1.5)\,\varepsilon_{T_{0M_0}}$$

$$\log(T_{1M_0}) = \log(6 + 0.4C + 0.3M_0 + 0.5 \times 1 + 0.1M_0 \times 1 - 0.2C \times M_0) + \log(1.5)\,\varepsilon_{T_{1M_0}}$$

$$\log(T_{0M_1}) = \log(6 + 0.4C + 0.3M_1 + 0.5 \times 0 + 0.1M_1 \times 0 - 0.2C \times M_1) + \log(1.5)\,\varepsilon_{T_{0M_1}}$$

$$\log(T_{1M_1}) = \log(6 + 0.4C + 0.3M_1 + 0.5 \times 1 + 0.1M_1 \times 1 - 0.2C \times M_1) + \log(1.5)\,\varepsilon_{T_{1M_1}}$$

The models used to decompose the effect are the same as in Scenario 1, thus resulting in misspecifications (the Cox model does not include an interaction term between C and M, and the linear regression does not include a quadratic term for C).

Scenario 4 specifically investigated the performance of our adherence-adjusted weighting approach. The population data generating equations for the first time-point are:

$$C = \varepsilon_C$$

$$P(A = 1) = expit(1 + 0.5C)$$

$$\mathrm{M_a} = -0.4\mathrm{a} + \mathrm{C} + \varepsilon_{\mathrm{M_a}}$$

$$\log(T_{aM_{a^*}}) = \min(\log(6 - 0.3C - 0.4M_{a^*} + 0.5a + 0.1M_{a^*} \times a) + \log(2)\,\varepsilon_{T_{0M_0}}, \log(3))$$

For the second time-point, the data generating equations are as follows:

$$C'_{0M_0} = C'_{1M_0} = M_0; C'_{0M_1} = C'_{1M_1} = M_1$$

$$P\left(A'_{aM_{a^*}} = 1\right) = expit\left(-2 + 0.5C'_{aM_{a^*}} + 4a\right)$$

$$M'_{aM_{a^*}a'} = -0.4a' + C'_{aM_{a^*}} + \varepsilon_{M'_{aM_{a^*}a'}}$$

$$\log(T'_{aM_{a^*}a'}) = \log\left(6 - 0.3C'_{aM_{a^*}} - 0.4M'_{aM_{a^*}a'} + 0.5a' + 0.1M'_{aM_{a^*}a'} \times a'\right) + \log(2)\,\varepsilon_{T'_{aM_{a^*}a'}}$$

where all errors terms are independent $N(0,1)$.

The sample data generating equations at the first time-point are the same as the population data generating equations, but include the additional equations:

$$M = M_0(1 - A) + M_1 A$$

$$T = T_{0M_0}(1 - A) + T_{1M_1}A$$

$$\log(T_C) = \min(\log(4 - 0.3C + 0.5A + 0.1M) + \log(1.5)\varepsilon_{T_C}, \log(3))$$

where T_C is a provisional random censoring time and $\varepsilon_{T_C} \sim N(0,1)$. Observations for which $T_C < T$ have their follow-up time censored at T_C. Observations that either experienced the event or have been censored before the second time-point ($T < 3$ or $T_C < 3$) have missing data for their variables at the second time-point. Otherwise, the sample data generating equations at the second time point are also the same as the population data generating equations, but further include the following equations:

$$C' = C'_{0M_0}(1 - A) + C'_{1M_1}$$

$$A' = A'_{0M_0}(1 - A) + A'_{1M_1}$$

$$\mathrm{M}' = \mathrm{M}'_{0\mathrm{M}_0 0}(1 - \mathrm{A})(1 - \mathrm{A}') + \mathrm{M}'_{1\mathrm{M}_1 0}A(1 - \mathrm{A}') + \mathrm{M}'_{0\mathrm{M}_0 1}(1 - \mathrm{A})\mathrm{A}' + \mathrm{M}'_{1\mathrm{M}_1 1}\mathrm{AA}'$$

$$\mathrm{T}' = \mathrm{T}'_{0\mathrm{M}_0 0}(1 - \mathrm{A})(1 - \mathrm{A}') + \mathrm{T}'_{1\mathrm{M}_1 0}A(1 - \mathrm{A}') + \mathrm{T}'_{0\mathrm{M}_0 1}(1 - \mathrm{A})\mathrm{A}' + \mathrm{T}'_{1\mathrm{M}_1 1}\mathrm{AA}'$$

$$\log(T'_C) = \min(\log(3 - 0.3C' + 0.5A' + 0.1M') + \log(1.5)\varepsilon_{T'_C}, \log(3))$$

where T'_C is a random censoring time and $\varepsilon_{T'_C} \sim N(0,1)$. The total time to event for subjects participating at the second time-point is $T + T'$ and their total censoring time is $T_C + T'_C$. Observations for which $T_C + T'_C < T + T'$ have their follow-up time censored at $T_C + T'_C$.

To estimate the total effect and its component, we have computed weights for each observation as described in Appendix 3. More precisely, the numerator of the weight corresponded to $P(A' = A|C)$ and the denominator was $P(A' = A|C, M)$, both of which were estimated utilizing logistic regression models that were fitted separately for exposed subjects and unexposed subjects. Weights were then truncated at their 99th percentile. Afterward, an

augmented data set was constructed that contained one row for each observation at the first time-point and an additional row for each subject that participated at the second time-point and remained adherent to their initial exposure level. The weights for the rows at the first time-point were all set to 1, whereas the rows for the second time-point were those we previously described.

To compute $\hat{\mathrm{E}}\left[Y\middle|A=a,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$, we first fitted a Cox proportional hazard model that included A and C on the weighted augmented dataset. The predicted 5-year risk was then computed for the original dataset that included only one row per observation. Similarly, $\hat{\mathrm{E}}\left[Y\middle|A=a,M=m,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$ was estimated utilizing a Cox model that included A, M, C and an interaction term between A and M. We computed $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\boldsymbol{c}_{i,t}\right)$ using the output of a linear regression of M according to A and C fitted on the weighted augmented dataset; predicted values were however computed for the original dataset. We then took $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\ \boldsymbol{c}_{i,t}\right)$ as a normal distribution whose mean was the linear predictor from the model and whose variance was the residual variance.

Simulation results

Table A1. Results of the simulation study investigating the performance of our semi-parametric g-formula estimator of the total effect and its components

Scenario	Effect	True value	Estimated value	Bias	Coverage probability
1	Total RD	-8.6%	-8.6%	0.0%	94.9%
	Direct RD	-6.6%	-6.8%	-0.2%	95.0%
	Indirect RD	-2.0%	-1.8%	0.2%	93.5%
	Total RRR	26%	25%	1%	95.0%
	Direct RRR	20%	19%	-1%	94.7%
	Indirect RRR	6%	6%	0%	93.4%
2	Total RD	-6.8%	-6.8%	0.0%	95.4%
	Direct RD	-6.8%	-6.8%	0.0%	95.3%
	Indirect RD	0.0%	0.0%	0.0%	94.6%
	Total RRR	21%	19%	1%	95.6%
	Direct RRR	21%	19%	1%	95.1%
	Indirect RRR	0%	0%	0%	94.5%
3	Total RD	-8.3%	-8.1%	0.2%	94.6%
	Direct RD	-6.4%	-6.3%	0.1%	94.7%
	Indirect RD	-1.9%	-1.8%	0.1%	93.7%
	Total RRR	27%	25%	-2%	94.0%
	Direct RRR	21%	19%	-2%	94.7%
	Indirect RRR	6%	6%	0%	94.2%
4	Total RD	-5.4%	-5.0%	0.4%	95.5%
	Direct RD	-4.3%	-4.2%	0.2%	95.6%
	Indirect RD	-1.1%	-0.9%	0.2%	91.6%
	Total RRR	25%	21%	-3%	95.4%
	Direct RRR	20%	17%	3%	95.5%
	Indirect RRR	5%	4%	1%	93.3%

These simulation results suggest that our proposed approach is able to unbiasedly estimate the total effect and its component when the models used are correctly specified, even with a small sample size. Moreover, no substantial bias was observed and coverage rates remained appropriate under mild misspecifications, suggesting that our approach is robust to some degree of misspecification of the models.

Appendix 3: Details concerning the adherence-adjusted analysis

In this appendix, we describe in more details how adherence-adjusted estimates were obtained. To do this, we have closely followed the approach proposed by Hernán et al. (2008)[27]. First, we created an augmented dataset for each trial that included one row for each participant at entry into the trial as well as one row for each subsequent follow-up visit to which the participant took part. The complete augmented dataset was obtained by concatenating all the trial-specific augmented datasets. We then computed the following visit-specific inverse probability weights for each participant in each trial:

$$w_{itk} = \prod_{k=1}^{K_t} \frac{P(Adh_{itk} = 1|\boldsymbol{Z}_{it0})}{P(Adh_{itk} = 1|\overline{\boldsymbol{Z}}_{itk})}, \quad (3)$$

where *k* is the follow-up visit, K_t is the total number of follow-up visits for Trial *t* and Adh_{itk} is an indicator of whether or not the participant *i* of Trial *t* was still adhering to their initial treatment at follow-up visit *k*, $\boldsymbol{Z}_{it0}$ and $\overline{\boldsymbol{Z}}_{itk}$ are sets of covariates according to which observations are weighted. Note that both the numerator and the denominator of Equation (3) equal 1 when $k = 1$, at entry into the trial. To estimate the numerator and the denominator for rows where $k \neq 1$, we have fitted pooled logistic regression models on the complete augmented dataset, excluding entry rows. The covariates to include in the weights calculations should be chosen such that discontinuing the initial treatment (i.e., taking statins or not taking statins) is independent of unmeasured risk factors of the incidence of the outcome (CHD, CVD or mortality) conditional on the weighting variables[27]. The explanatory variables for the

denominator models ($\overline{\boldsymbol{Z}}_{itk}$) included all pre-entry potential confounders, the most recent values of the same variables (when they were time-varying) and indicator variables for the follow-up examinations. Age and Agatston score were included as in the previous models. The explanatory variables for the numerator ($\boldsymbol{Z}_{it0}$) only included the pre-entry potential confounder variables. Because the factors that explain why statin users remain users might differ from those that explain why non-users remain non-users, separate logistic regression models were fitted. To reduce the potential influence of observations with large weights, the weights were truncated at their 99th percentile. The estimated inverse probability weights had mean 1.00 (standard-deviation = 0.13), the 99th percentile of the weights was 1.39 and the maximum was 7.79. After truncation, the mean was 0.99 (standard deviation = 0.09).

The estimation procedure then carried forward similarly to what was described in the "Intention to treat estimation" section. The only differences are that the parameters of the linear and the Cox regression models were estimated on the complete augmented dataset we have just described, censoring subjects once they discontinued their initial treatment and weighting other rows according to the inverse probability weights (3). Predicted values of $\hat{\mathrm{E}}\left[Y\middle|A=a,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$, $\hat{\mathrm{E}}\left[Y\middle|A=a,M=m,\boldsymbol{C}=\boldsymbol{c}_{i,t}\right]$ and $\hat{f}\left(m\middle|A=a^{*},\boldsymbol{C}=\boldsymbol{c}_{i,t}\right)$ were then computed for the original dataset that included only one row per participant in each trial.

Appendix 4: Results of the sensitivity analyses

Atorvastatin vs non-statin users

The results of the comparison between Atorvastatin users and non-statin users are presented in Table A2. Since these analyses were conducted on a reduced sample, statistical power was reduced, leading to generally wider confidence intervals. The total association with mortality is similar, but the association with CVD and CHD is larger than in the main analysis. Moreover, a larger portion of all three associations investigated seem to be attributable to the LDL pathway. These analyses also suggest that a small, but statistically significant part of the association between Atorvastatin and both mortality and CVD might be attributable to their effect on HDL, in opposition to the main findings. This may suggest that the importance of the cholesterol pathways would vary according to statin type. Given that this analysis is underpowered, these results would need to be confirmed by future studies.

Table A2. Decomposition of the adherence-adjusted effect estimate of Atorvastatin on Coronary Heart Disease, Cardiovascular Disease and Mortality (95% confidence intervals in parenthesis)

	Measure	Total effect of Atorvastatin	Direct effect of Atorvastatin	Indirect effect of Atorvastatin (via the mediator)
CHD				
LDL	RD	-0.7% (-1.7%, 0.3%)	0.5% (-1.4%, 3.8%)	-1.3% (-4.0%, 0.2%)
	RRR	27% (-10%, 57%)	-19% (-135%, 49%)	46% (-6%, 146%)
HDL	RD	-0.7% (-1.7%, 0.3%)	-0.7% (-1.6%, 0.3%)	-0.1% (-0.2%, 0.1%)
	RRR	27% (-10%, 57%)	25% (-13%, 56%)	2% (-2%, 8%)
Trig.	RD	-0.7% (-1.7%, 0.3%)	-0.7% (-1.7%, 0.3%)	0.0% (-0.2%, 0.1%)
	RRR	27% (-10%, 57%)	27% (-11%, 56%)	0% (-5%, 6%)
CVD				
LDL	RD	-1.3% (-2.4%, 0.0%)	-0.3% (-2.3%, 2.9%)	-1.0% (-3.8%, 0.7%)
	RRR	30% (0%, 55%)	6% (-71%, 55%)	24% (-15%, 91%)
HDL	RD	-1.3% (-2.4%, 0.0%)	-1.1% (-2.3%, 0.1%)	-0.2% (-0.5%, 0.0%)
	RRR	30% (0%, 55%)	27% (-3%, 53%)	4% (0%, 11%)
Trig.	RD	-1.3% (-2.4%, 0.0%)	-1.2% (-2.3%, 0.0%)	0.0% (-0.2%, 0.1%)
	RRR	30% (0%, 55%)	30% (0%, 54%)	1% (-3%, 5%)
Mortality				
LDL	RD	-0.8% (-2.1%, 0.5%)	-0.3% (-2.2%, 2.3%)	-0.5% (-2.7%, 0.8%)
	RRR	17% (-11%, 43%)	7% (-51%, 45%)	11% (-18%, 57%)
HDL	RD	-0.8% (-2.1%, 0.5%)	-0.5% (-1.9%, 0.8%)	-0.3% (-0.7%, 0.0%)
	RRR	17% (-11%, 43%)	11% (-18%, 39%)	6% (1%, 15%)
Trig.	RD	-0.8% (-2.1%, 0.5%)	-0.8% (-2.0%, 0.5%)	0.0% (-0.3%, 0.1%)
	RRR	17% (-11%, 43%)	16% (-11%, 42%)	1% (-3%, 6%)

The total estimated effect is decomposed in an indirect effect, due to the effect of Atorvastatin on a given mediator, and a direct component, not due to the effect of Atorvastatin on that given mediator. LDL = low-density lipoprotein, HDL = high-density lipoprotein, Trig. = triglycerides, RD = risk difference, RRR = relative risk reduction, CHD = coronary heart diseases, CVD = cardiovascular diseases.

Estimates are adjusted for age, education, gender, health insurance, race/ethnicity, Field Center, previous LDL, previous HDL, previous triglycerides, Agatston score, BMI, cigarette smoking, diabetes, family history of CVD, hypertension and total number of medication.

Adjusting for further potential confounders

Table A3 reports the results of the adherence-adjusted sensitivity further adjusting for alcohol consumption, whereas Table A4 reports the results further adjusting for walking pace. In both cases, the strength of the associations is generally marginally larger, but otherwise the results are similar to those of the main analysis.

Table A3. Results of the adherence-adjusted sensitivity analysis when further adjusting for alcohol consumption

	Measure	Total effect of statins	Direct effect of statins	Indirect effect of statins (via the mediator)
CHD				
LDL	RD	-0.5% (-1.3%, 0.2%)	0.3% (-1.0%, 1.8%)	-0.8% (-1.9%, 0.1%)
	RRR	17% (-8%, 39%)	-9% (-61%, 30%)	26% (-2%, 64%)
HDL	RD	-0.5% (-1.3%, 0.2%)	-0.5% (-1.2%, 0.3%)	0.0% (-0.1%, 0.0%)
	RRR	17% (-8%, 39%)	17% (-9%, 39%)	1% (-1%, 3%)
Trig.	RD	-0.5% (-1.3%, 0.2%)	-0.6% (-1.3%, 0.2%)	0.0% (-0.1%, 0.1%)
	RRR	17% (-8%, 39%)	18% (-7%, 40%)	-1% (-5%, 2%)
CVD				
LDL	RD	-1.1% (-2.0%, -0.2%)	-0.7% (-2.0%, 0.7%)	-0.4% (-1.4%, 0.4%)
	RRR	25% (5%, 42%)	16% (-16%, 42%)	8% (-10%, 32%)
HDL	RD	-1.1% (-2.0%, -0.2%)	-1.1% (-1.9%, -0.1%)	-0.1% (-0.1%, 0.0%)
	RRR	25% (5%, 42%)	24% (3%, 42%)	1% (0%, 3%)
Trig.	RD	-1.1% (-2.0%, -0.2%)	-1.1% (-2.0%, -0.2%)	0.0% (-0.1%, 0.1%)
	RRR	25% (5%, 42%)	25% (4%, 42%)	0% (-3%, 2%)
Mortality				
LDL	RD	-0.9% (-1.8%, 0.1%)	-0.9% (-2.1%, 0.3%)	0.0% (-0.8%, 0.8%)
	RRR	18% (-1%, 34%)	19% (-7%, 40%)	-1% (-15%, 16%)
HDL	RD	-0.9% (-1.8%, 0.1%)	-0.8% (-1.7%, 0.1%)	-0.1% (-0.2%, 0.0%)
	RRR	18% (-1%, 34%)	16% (-3%, 33%)	1% (0%, 4%)
Trig.	RD	-0.9% (-1.8%, 0.1%)	-0.9% (-1.8%, 0.0%)	0.0% (-0.1%, 0.2%)
	RRR	18% (-1%, 34%)	19% (0%, 35%)	-1% (-3%, 1%)

The total estimated effect is decomposed in an indirect effect, due to the effect of Atorvastatin on a given mediator, and a direct component, not due to the effect of Atorvastatin on that given mediator. LDL = low-density lipoprotein, HDL = high-density lipoprotein, Trig. = triglycerides, RD = risk difference, RRR = relative risk reduction, CHD = coronary heart diseases, CVD = cardiovascular diseases.

Estimates are adjusted for age, education, gender, health insurance, race/ethnicity, Field Center, previous LDL, previous HDL, previous triglycerides, Agatston score, BMI, cigarette smoking, diabetes, family history of CVD, hypertension, total number of medication and alcohol consumption.

Table A4. Results of the adherence-adjusted sensitivity analysis when further adjusting for walking pace

	Measure	Total effect of statins	Direct effect of statins	Indirect effect of statins (via the mediator)
CHD				
LDL	RD	-0.5% (-1.3%, 0.4%)	0.2% (-1.1%, 1.7%)	-0.7% (-1.8%, 0.1%)
	RRR	17% (-12%, 41%)	-7% (-59%, 36%)	24% (-4%, 61%)
HDL	RD	-0.5% (-1.3%, 0.4%)	-0.5% (-1.3%, 0.4%)	0.0% (-0.1%, 0.0%)
	RRR	17% (-12%, 41%)	16% (-13%, 41%)	1% (-1%, 3%)
Trig.	RD	-0.5% (-1.3%, 0.4%)	-0.5% (-1.3%, 0.4%)	0.0% (-0.1%, 0.1%)
	RRR	17% (-12%, 41%)	17% (-13%, 41%)	0% (-3%, 3%)
CVD				
LDL	RD	-1.1% (-2.0%, -0.2%)	-0.7% (-1.9%, 0.8%)	-0.4% (-1.5%, 0.4%)
	RRR	26% (4%, 45%)	16% (-19%, 44%)	10% (-9%, 35%)
HDL	RD	-1.1% (-2.0%, -0.2%)	-1.0% (-1.9%, -0.1%)	0.0% (-0.1%, 0.0%)
	RRR	26% (4%, 45%)	25% (3%, 44%)	1% (0%, 3%)
Trig.	RD	-1.1% (-2.0%, -0.2%)	-1.1% (-1.9%, -0.1%)	0.0% (-0.1%, 0.1%)
	RRR	26% (4%, 45%)	25% (3%, 44%)	0% (-2%, 3%)
Mortality				
LDL	RD	-1.1% (-2.0%, -0.2%)	-1.2% (-2.3%, -0.1%)	0.2% (-0.6%, 0.8%)
	RRR	23% (5%, 39%)	26% (2%, 46%)	-4% (-17%, 12%)
HDL	RD	-1.1% (-2.0%, -0.2%)	-1.0% (-1.9%, -0.1%)	0.0% (-0.1%, 0.0%)
	RRR	23% (5%, 39%)	22% (3%, 39%)	1% (0%, 3%)
Trig.	RD	-1.1% (-2.0%, -0.2%)	-1.1% (-2.0%, -0.2%)	0.0% (-0.1%, 0.1%)
	RRR	23% (5%, 39%)	23% (5%, 40%)	0% (-3%, 2%)

The total estimated effect is decomposed in an indirect effect, due to the effect of Atorvastatin on a given mediator, and a direct component, not due to the effect of Atorvastatin on that given mediator. LDL = low-density lipoprotein, HDL = high-density lipoprotein, Trig. = triglycerides, RD = risk difference, RRR = relative risk reduction, CHD = coronary heart diseases, CVD = cardiovascular diseases.

Estimates are adjusted for age, education, gender, health insurance, race/ethnicity, Field Center, previous LDL, previous HDL, previous triglycerides, Agatston score, BMI, cigarette smoking, diabetes, family history of CVD, hypertension, total number of medication and walking pace.

Appendix 5 (Table A5): Pre-entry Characteristics of Participants According to Statin Usage at the Entry Examination and Trial

	Trial 1		Trial 2		Trial 3		Trial 4	
	Non-Users (n = 4,405)	Users (n = 417)	Non-Users (n = 3,739)	Users (n = 361)	Non-Users (n = 3,291)	Users (n = 341)	Non-Users (n = 2,267)	Users (n = 532)
Age, mean (SD)	61.1 (10.3)	64.3 (9.5)	62.5 (10.1)	63.9 (9.5)	63.5 (10.0)	66.1 (9.1)	63.6 (9.6)	65.1 (8.9)
Education (highest level reached)								
- Less than high school	16.7	18.5	15.6	20.8	14.7	17.6	12.6	16.4
- High school	17.0	19.9	17.2	16.9	17.3	20.2	16.0	19.2
- More than high school	66.2	61.6	67.2	62.3	68.0	62.2	71.5	64.5
Gender female	53.0	50.1	53.5	60.4	54.1	55.4	55.1	53.9
Health insurance	9.1	6.2	7.3	5.0	6.8	3.8	7.4	4.5
Race/ethnicity								
- White, Caucasian	37.5	43.6	38.4	34.1	39.3	37.0	36.0	37.4
- Chinese American	12.3	9.1	12.7	10.8	12.0	12.6	12.7	8.8
- Black, African-American	27.8	26.6	27.4	30.7	26.9	26.4	26.5	28.6
- Hispanic	22.4	20.6	21.6	24.4	21.8	24.0	21.8	25.2
Site								
- WFU	15.1	18.0	15.4	13.0	15.8	10.9	14.7	21.6
- COL	17.1	15.1	17.0	18.8	16.8	15.2	18.1	16.9
- JHU	14.4	17.3	14.1	172	12.9	17.3	11.9	11.5
- UMN	15.8	13.7	15.8	13.9	16.0	19.6	16.1	17.3
- NWU	18.6	15.3	19.6	15.8	19.0	15.2	20.4	15.6
- UCLA	18.9	20.6	18.1	21.3	19.3	21.7	18.7	17.1
LDL in mmol/L, mean (SD)	3.0 (0.8)	3.7 (0.8)	3.0 (0.8)	3.6 (1.0)	3.0 (0.8)	3.4 (0.8)	3.0 (0.8)	3.4 (0.9)
HDL in mmol/L, mean (SD)	1.3 (0.4)	1.3 (0.3)	1.4 (0.4)	1.3 (0.3)	1.4 (0.4)	1.3 (0.3)	1.4 (0.4)	1.3 (0.4)
Triglycerides in mmoL/L, mean (SD)	1.4 (0.7)	1.6 (0.8)	1.4 (0.7)	1.7 (0.8)	1.3 (0.7)	1.6 (0.8)	1.3 (0.7)	1.5 (0.8)
Agatston score > 0	44.5	61.2	42.4	55.4	40.0	56.3	34.8	44.4
Ln of Agatston score when > 0, mean (SD)	4.2 (1.8)	4.7 (1.8)	4.1 (1.8)	4.3 (1.8)	4.1 (1.8)	4.4 (1.9)	3.8 (1.8)	4.2 (1.7)
BMI								
- Normal	30.9	22.8	31.8	22.4	32.8	27.0	33.0	25.8
- Overweight	38.5	40.0	38.8	35.2	37.6	38.7	36.5	38.5
- Obesity	30.6	37.2	29.4	42.4	29.6	34.3	30.4	35.7
Cigarette smoking								
- Never	50.9	48.7	47.5	47.6	47.2	46.6	46.5	48.5
- Former	36.0	41.7	40.9	40.7	42.0	47.2	43.3	42.5
- Current	13.2	9.6	11.6	11.6	10.8	6.2	10.2	9.0
Diabetes	8.8	22.8	9.9	23.5	8.8	25.2	8.0	19.6
Family history of CVD	26.2	29.3	26.0	30.2	25.9	25.5	25.5	29.9
Hypertension	39.5	56.8	39.3	57.1	41.0	59.2	39.8	56.0
Number of medications, mean (SD)	2.8 (2.6)	3.4 (2.6)	4.6 (3.6)	5.3 (3.4)	4.7 (3.9)	5.6 (3.7)	4.6 (4.0)	5.4 (3.9)

Results are expressed as percentage unless otherwise indicated

Appendix 6: Results of the ITT analyses

Table A6. Decomposition of the intention to treat effect estimate of statins on Coronary Heart Disease, Cardiovascular Disease and Mortality (95% confidence intervals in parenthesis)

	Measure	Total effect of statins	Direct effect of statins	Indirect effect of statins (via the mediator)
CHD				
LDL	RD	-0.7% (-1.4%, 0.0%)	-0.3% (-1.2%, 0.8%)	-0.4% (-1.1%, 0.2%)
	RRR	20% (0%, 38%)	8% (-23%, 34%)	12% (-4%, 33%)
HDL	RD	-0.7% (-1.4%, 0.0%)	-0.7% (-1.4%, 0.0%)	0.0% (-0.1%, 0.0%)
	RRR	20% (0%, 38%)	19% (-1%, 37%)	1% (0%, 3%)
Trig.	RD	-0.7% (-1.4%, 0.0%)	-0.7% (-1.4%, 0.0%)	0.0% (0.0%, 0.1%)
	RRR	20% (0%, 38%)	21% (0%, 39%)	-1% (-3%, 1%)
CVD				
LDL	RD	-1.2% (-2.0%, -0.4%)	-0.8% (-1.9%, 0.4%)	-0.4% (-1.2%, 0.3%)
	RRR	23% (7%, 37%)	16% (-7%, 36%)	-7% (-23%, 6%)
HDL	RD	-1.2% (-2.0%, -0.4%)	-1.1% (-2.0%, -0.3%)	-0.1% (-0.1%, 0.0%)
	RRR	23% (7%, 37%)	22% (6%, 37%)	-1% (-2%, 0%)
Trig.	RD	-1.2% (-2.0%, -0.4%)	-1.2% (-2.0%, -0.4%)	0.0% (-0.1%, 0.1%)
	RRR	23% (7%, 37%)	23% (7%, 37%)	0% (-1%, 2%)
Mortality				
LDL	RD	-0.7% (-1.5%, 0.0%)	-0.9% (-1.8%, 0.1%)	0.2% (-0.5%, 0.7%)
	RRR	15% (0%, 30%)	19% (-3%, 36%)	3% (-10%, 15%)
HDL	RD	-0.7% (-1.5%, 0.0%)	-0.7% (-1.4%, 0.1%)	-0.1% (-0.1%, 0.0%)
	RRR	15% (0%, 30%)	14% (-2%, 29%)	-1% (-3%, 0%)
Trig.	RD	-0.7% (-1.5%, 0.0%)	-0.8% (-1.5%, 0.0%)	0.0% (-0.1%, 0.1%)
	RRR	15% (0%, 30%)	16% (1%, 30%)	0% (-1%, 2%)

The total estimated effect is decomposed in an indirect effect, due to the effect of statins on a given mediator, and a direct component, not due to the effect of statins on that given mediator. LDL = low-density lipoprotein, HDL = high-density lipoprotein, Trig. = triglycerides, RD = risk difference, RRR = relative risk reduction, CHD = coronary heart diseases, CVD = cardiovascular diseases.

Estimates are adjusted for age, education, gender, health insurance, race/ethnicity, Field Center, previous LDL, previous HDL, previous triglycerides, Agatston score, BMI, cigarette smoking, diabetes, family history of CVD, hypertension and total number of medication.

Appendix 7: Expanded substantive discussion

In our study, we have obtained a beneficial indirect effect estimate of statins on CHD through LDL and a detrimental direct effect estimate (through other pathways than LDL). This type of reversed direct effect is plausible, as CHD is the original target of statins (so we would expect this association to be present) but all drugs have side effects. For instance, statins might increase the incidence of type 2 diabetes[5].

Our results regarding CHD are in concordance with those of previous studies that found that greater reduction in LDL cholesterol was associated with greater reduction in CHD risk[6,18,20,21,22,23]. They thus support the current guidelines that advocate for reducing LDL cholesterol as a means to reducing the risk of CHD events, where lower LDL targets are recommended for individuals at higher risk of CHD[44].

To our knowledge, only one previous study attempted quantifying the importance of the cholesterol pathways in the effect of pravastatin on CHD in humans using causal mediation methods[25]. This study suggested that most of the effect was LDL-independent, in opposition to the current study. This discrepancy might be due to the fact that pravastatin is less effective at reducing cholesterol than most other types of statins[45].

More studies investigating the importance of the cholesterol pathways in the effect of statins on CHD, CVD and mortality utilizing appropriate mediation methods are certainly needed. Future studies should also investigate other pathways by which statins might affect those outcomes. In that respect, inflammation appears to be particularly promising as an additional pathway by which statins may act on clinical endpoints[10]. Randomized trials

specifically designed to estimate the importance of these pathways would attenuate the risk that results are biased by confounding by indication.

There might be very important implications if it were confirmed that an important portion of the mechanisms relating statin use to reduced mortality and improved cardiovascular health are cholesterol-independent. Notably, this might suggest that statin medication could be pursued in patients not achieving any LDL reduction and still yield a reduction in CVD and mortality risks. New populations that might benefit from statins might also be identified.

Additional references

44. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the national cholesterol education program expert panel on detection, evaluation, and treatment of high blood cholesterol in adults. *JAMA*. 2001;285(19):2486-2497.

45. Schaefer EJ, McNamara JR, Tayler T, Daly JA, Gleason JL, Seman LJ, Ferrari A, Rubenstein JJ. Comparisons of effects of statins (atorvastatin, fluvastatin, lovastatin, pravastatin, and simvastatin) on fasting and postprandial lipoproteins in patiens with coronary heart disease versus control subjects. *Am J Cardiol*. 2004;93(1):31-39.