

1 Understanding the causes and consequences of low statin
2 adherence: evidence from UK Biobank primary care data

3

4 Deniz Türkmen* ¹, Xiaoran Liang* ¹, Jane A H Masoli^{1,2}, Dipender Gill³, Luke C Pilling¹, Jack
5 Bowden¹

6 *Joint first authors

7 Authors and the order not final and to be decided.

8

9 1. Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK

10 2. Royal Devon University Healthcare NHS Foundation Trust, Barrack Road, Exeter, EX2 5DW, UK

11 3. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,
12 90 Wood Lane, London, W12 0BZ, UK

13

14 **Keywords**

15 Cardiovascular, Statin, Primary Care, Adherence, Pharmacogenomics, Clinical Outcomes

16

17 **Corresponding Author**

18 Jack Bowden, j.bowden2@exeter.ac.uk

19

1 **Abstract**

2 **Background:** Statins are commonly prescribed to lower LDL cholesterol. Clinical guidelines
3 recommend 30-50% reduction within 3 months, yet many patients do not achieve this. We
4 investigated the impact of patient characteristics and genetics on LDL-c reduction, treatment
5 adherence, and adverse clinical outcomes.

6 **Methods:** We analysed 76,000 UK Biobank participants prescribed atorvastatin or simvastatin in
7 primary care: 41,000 had LDL-c measurements before statin initiation (median=16 days prior,
8 IQR=28) and within a year of starting treatment (median=89 days, IQR=125). Adherence was defined
9 as the “proportion of days covered” (PDC). We estimated associations between PDC within one year
10 of statin initiation, genetic factors, post-treatment LDL reduction, and clinical adverse outcomes. For
11 13,000 patients with ≥ 3 LDL-c measures, we used inverse probability weighting methods to estimate
12 the effect of sustained adherence intervention on LDL-c reduction longitudinally.

13 **Results:** Predictors of LDL-c reduction following statin initiation included the time until the 1st
14 measurement, PDC, and the pharmacogenetic variant *SLCO1B1**5. LDL-c reduction was greater in
15 those with high adherence versus lower adherence (38% reduction when PDC>95% [high] vs. 15%
16 when PDC<50% [low]). Longitudinal causal modelling showed that the most recent PDC measure
17 exerted the largest influence on overall LDL-c reduction, followed by the initial PDC.

18 Genetic predictors of reduced PDC included liability to schizophrenia (Coef_{top 20%}-1.94, 95%CI -2.69 to
19 -1.19), whilst genetic liability to cardiovascular diseases increased PDC (Coef_{top 20%}1.30, 95%CI 0.55 to
20 2.05). High PDC was associated with increased risk of incident iron deficiency anaemia (HR 1.30,
21 95%CI 1.09-1.54) and cataract (HR 1.20, 95%CI 1.07-1.34), and decreased risk of incident coronary
22 heart disease (HR 0.78, 95%CI 0.73-0.84).

23 **Conclusion:** We identify substantial variability in the time to first on-treatment LDL measurements
24 and also in adherence to statin medication, highlighting a gap between NHS guidelines, LDL
25 monitoring and statin adherence. We show its subsequent impact on long term health,
26 demonstrating the potential effect of targeted interventions to improve adherence. We identify

- 1 important predictors of reduced statin effectiveness, including pharmacogenetic variants, polygenic
- 2 scores, but most of all, adherence. Tailored statin therapy strategies with patient education on statin
- 3 indication and adherence could optimise treatment efficacy, safety, and long-term clinical outcomes.
- 4

1 Introduction

2

3 Cardiovascular diseases (CVDs) remain the leading cause of mortality and morbidity in adults globally
4 (1). Low-density lipoprotein cholesterol (LDL-c), a key modifiable risk factor, can be lowered with
5 statins, reducing CVD risk by ~22% per 1 mmol/L (18mg/dL) reduction (2). National guidelines,
6 including the UK National Institute for Health and Care Excellence (NICE) and the American Heart
7 Association (AHA), recommend a 30-50% target LDL-c reduction for those aged 40-75 with a high
8 CVD risk with a follow-up lipid check at 2-3 months post-statin initiation (3–5). In the UK, for
9 secondary prevention, the target LDL-C is <2 mmol/L (36mg/dL). Many individuals with detected
10 hypercholesterolaemia are not followed up in the desired time frame and do not achieve these
11 targets (6–9). Adherence to statin therapy in real-world studies can be <50% within the first year of
12 treatment (10–12), far lower than in randomized controlled trials (RCT) conducted with close clinical
13 monitoring and mostly concentrating on patients after an CVD event or hospitalization (13,14). Poor
14 long-term adherence to statin therapy is associated with higher hospitalization rates (15), yet the
15 patient-related causes and consequences of statin adherence in routine primary care remains
16 unclear.

17

18 Recent trials (16,17) show those at high genetic risk for CVD experience greater benefits from statins
19 in primary prevention, but whether this effect is pharmacogenetic or simply an artefact of increased
20 pre-treatment LDL levels is unclear. Understanding any effect of pharmacogenetic variants and
21 polygenic scores (aggregated over multiple genetic loci) on statin adherence and LDL-c reduction is of
22 vital importance, since they offer the potential for stratified medicine to better support the
23 identification and optimization of treatment outcomes in high-risk populations. Recent largescale
24 studies have investigated socio-demographic and genetic risk factors of statin adherence (18), or
25 genome-wide association studies of lipid response to statins (19), but have not considered
26 adherence, LDL-c reduction, and adverse clinical outcomes in parallel. To address this, we conducted

1 an extensive analysis of 76,000 UK Biobank participants prescribed statin medication in the linked
2 primary care records. We estimated the impact of time to first on-treatment follow-up,
3 pharmacogenetic variants, polygenic scores, and adherence on LDL-c reduction within 1 year of
4 initiating statin treatment. We also estimated the impact of low adherence on adverse clinical
5 outcomes. To further evaluate the effect of interventions to increase statin adherence on LDL levels
6 over time, we performed longitudinal causal modelling of individual patient LDL trajectories, paired
7 with inverse probability weighting (IPW) methods (20) to control for observable time-varying
8 confounders of adherence and LDL.

9

10

11

12

13

14

15

16

17

18

19

20

21

1

2 **Methods**

3

4 Cohort: UK Biobank

5

6 The UK Biobank (UKB) recruited 503,325 community-based volunteers aged 40-70 from Wales,
7 Scotland, or England over the period 2006-2010. Participants provided blood samples for genetic and
8 biochemical analyses. This study involves two analyses: 1) using the linked primary care data from
9 General Practice (GP) available in 230,096 participants (45.7% of UKB) (Figure 1) to examine LDL
10 reduction, and 2) Incorporating secondary care data (hospital) with the GP data for the incident
11 outcomes.

12 GP data includes drug name, date of prescription, number of tablets, and drug code (in Clinical Read
13 version 2, British National Formulary (BNF) or Dictionary of Medicines and Devices (DM+D) format,
14 depending on provider) and are up to September 2017 (EMIS/Vision system in Wales) and August
15 2016 (TPP system supplier in England). We included simvastatin and atorvastatin prescriptions,
16 classifying first doses by NICE guidelines (21) : 8% low-intensity, 76% medium-intensity, and 16%
17 high-intensity in the UKB cohort (Appendix Table 1 and 2).

18 We identified primary and secondary care-diagnosed CVDs (Abdominal Aortic Aneurysm, Coronary
19 Heart Disease, Ischemic Stroke, Ischemic Stroke or Transient Ischemic Attack, Peripheral Arterial
20 Disease, All Types of Stroke, and Transient Ischemic Attack) based on clinical guidelines for initiating
21 statin treatment for secondary prevention (22). Supplementary Table 1 for the ICD-10 codes.

22

1 Genetic Data

2

3 UKB obtained 805,426 directly genotyped variants via the Affymetrix Axiom UKB array (in 438,427
4 participants) and the Affymetrix UKBiLEVE array (in 49,950 participants). Imputation was performed
5 by the central UKB team for 487,442 participants and ~96 million genetic variants were obtained.
6 Neither participants' nor their healthcare providers received genotype data as part of the study, so
7 this information could not directly influence treatment. 451,367 (93%) participants genetically similar
8 to the 1000Genomes EUR reference population ("European-like") were included in our genetic
9 analyses to minimise the effect of population stratification (this was identified via Principal
10 Components Analysis using the 1000Genomes reference panels, see (23) for details).

11

12 We also included polygenic scores (PGS) derived by Genomics PLC (24) (UKB category 301): 8 PGS for
13 relevant biomarkers and patient characteristics (e.g., LDL, BMI) and 28 PGS for genetic liability to
14 disease (e.g., CVD, schizophrenia). See Supplementary Table 2 for the details.

15 Adherence as measured by Proportion of Days Covered (PDC)

16 For each patient with a history of statin treatment initiation as well as pre- and post-treatment LDL
17 measures (n=69,503), we sought all available prescribing data. We then calculated the number of
18 tablets prescribed to each individual in a given time period, to create the variable:

19

$$20 \quad PDC = \frac{\text{Number of tablets prescribed in study period}}{\text{Length of study period (days)}} \times 100$$

21

22 We first focused on adherence in the first year of treatment, which we call PDC_1 . If patients received
23 a new prescription that extended past the first 12 months, we included the time until the next
24 prescription in the 'length' calculation for PDC_1 . However, if there were no subsequent prescriptions,
25 we excluded that prescription from the calculation (n=7,446). We also excluded patients who died

1 within 3 months of starting statins (n=39) and those with a $PDC_1 > 200\%$ (n=714). See Figure 1 for a
2 flowchart of this process and the final number of included participants.

3

4 We used a three-category definition for PDC_1 when treating it as the exposure: <50% , 50-95% ,
5 >95% based on observed adherence patterns in the cohort and partly in line with previous research
6 [e.g., for <50%] (12). In addition, when treating it as the outcome variable and estimating its
7 predictors via multivariable regression modelling, we used the original (continuous) PDC_1 measure.
8 When performing longitudinal causal modelling of individuals' LDL trajectories over three post-
9 treatment follow up visits, continuous measures for PDC adherence over participant-specific study
10 windows were used as covariates and in the derivation of inverse probability weights.

11

12 Outcomes

13

14 LDL change in 1 year

15

16 We obtained GP-recorded LDL measurements for patients on statins (see Supplementary Table 3 for
17 the read codes used). We identified the baseline LDL measure (within 180 days and preceding first
18 prescription) and the first follow-up measure following statin initiation, restricting analyses to
19 participants with a follow-up measure within 12 months of statin initiation. LDL measurements for
20 patients prescribed their first statin within 90 days of hospital discharge for a cardiovascular event
21 were excluded, as these prescriptions may not have been recorded by the GP. Extreme
22 measurements (LDL <1 mmol/L: $n_{LDL0}=1,402$, $n_{LDL1}=2,081$ and LDL >8 mmol/L: $n_{LDL0}=29$, $n_{LDL1}=13$) were
23 excluded (see Figure 1). We defined an LDL change measure as the post-statin/pre-statin LDL ratio:

24

$$LDL\ change\ (Y) = \frac{LDL1\ (follow\ -\ up\ LDL\ measurement)}{LDL0\ (baseline\ LDL\ measurement)}$$

1

2 Long-term clinical outcomes

3

4 Eighty four long-term clinical outcomes were investigated to identify intended and unintended
5 adverse events. This encompassed all common chronic diseases lasting more than 3 months with a
6 prevalence of >0.5% in older adults (25).

7

8 Analyses

9

10 Figure 2 provides a detailed visualisation of our analysis strategy. All models in Analysis 1 are
11 adjusted for sex, age at first statin prescription, the baseline assessment centre, first dose and
12 education unless stated otherwise.

13 Analysis 1: Cross-sectional analysis of LDL change

14 *LDL change explained by week of follow-up*

15

16 We tested associations between LDL change and the time interval between their pre- and post-
17 treatment measurements, using a linear regression model, adjusted additionally for dose of first
18 atorvastatin and simvastatin prescribed in year 1 and the prevalent cardiovascular disease diagnoses.
19 Doses were organised into a categorical variable based on NICE guidelines (21) (Appendix - Table 1).
20 We repeated this analysis in the $\leq 95\%$ and $>95\%$ PDC_1 groups separately.

21

22 *Modelling LDL change by genotype, PDC and time of follow-up measure.*

23

24 We next tested associations between LDL change and genetic variant rs4149056 using a linear model
25 that allowed for a quadratic effect of time-to-follow-up on treatment in European-like participants
26 and an interaction between time on treatment and rs4149056. We also conducted a linear

1 regression analysis to test the associations between polygenic scores and LDL change. We
2 additionally adjusted for 10 genetic principal components (PCs) in these two analyses.

3

4 Next, we tested the association between LDL change and PDC_1 using the same model, but with PDC
5 in place of rs4149056. This analysis did not adjust for genetic PCs but accounted for the prevalence
6 of CVD before statin use, including abdominal aortic aneurysm, atrial fibrillation, coronary heart
7 disease, peripheral arterial disease, all types of stroke, and thromboembolic disease. For further
8 technical details on the statistical modelling see the Appendix.

9

10 Analysis 2: Longitudinal analysis on LDL trajectories

11

12 *Causal modelling of hypothetical LDL change via sustained PDC intervention*

13

14 We applied and refined causal modelling approaches to evaluate how intervening to increase statin
15 adherence would affect a person's LDL reduction over a longer sustained period, adjusting for fixed
16 confounders of PDC and LDL, as well as time-varying confounding of previous PDC and LDL measures
17 on their future values. For patients prescribed statins, we included individuals with at least three
18 post-statin LDL measures over a ten-year window for the analysis (n=13,315). We defined LDL
19 reduction (LR) as the percentage change between each of their post-statin LDL measures and the
20 pre-statin baseline level (denoted by LDL_0) as

$$21 \quad LR_k = \frac{LDL_k - LDL_0}{LDL_0} \text{ for time point } k = 1, 2, 3.$$

22 Figure 3 shows the distribution of patients' three follow-up times over the full study window. The
23 average follow-up time to the third LDL measure was three years and the maximum was ten. For this
24 analysis, we redefined PDC_k as the average PDC value between time point $k-1$ and k . For example, if
25 $k = 1$, then $k-1$ indexes the baseline LDL and k indexes the first follow-up LDL. We assumed that PDC_1
26 up to PDC_k exerted additive linear effects on LR_k . Our goal was to estimate these effects, averaged

1 over the study population, by adjusting for fixed and time-varying confounding. The causal diagram
2 in Figure 4 Panel (a) shows the assumed data structure for our analysis.

3 We made the critical assumption that all the confounders of PDC and LR were observable and could
4 be controlled for in the analysis. We include age at the first statin prescription, sex, pre-statin
5 baseline LDL (i.e., LDL_0), education and assessment centre as fixed confounding factors. Previous
6 PDC and LR were assumed to act as time varying confounders of future PDC and LR measures. We
7 also included patient visit time for each post-statin LDL measure as the time-varying confounder.
8 We utilized the inverse probability weighting (IPW) method, which is a two-step procedure for causal
9 estimation under the observable confounding assumption. In stage 1, we create a pseudo-population by assigning
10 everyone in the sample a weight so that, in this weighted pseudo-population, the exposure PDC is
11 balanced across all the confounding covariates (see Figure 4 Panel (b) for illustration). We applied the
12 Covariate Balancing Propensity Score (CBPS) method (26) to estimate the weight for each individual,
13 using both the parametric and non-parametric (npCBPS) options. After obtaining the weights, we
14 evaluated the covariate balance in the pseudo-population by calculating the weighted Pearson and
15 Spearman correlation between each exposure and confounding covariate. Following the practice in
16 (27), if the size of the correlation fell below the cut-off value 0.1, then the corresponding covariate
17 was deemed sufficiently balanced. In stage 2, for each time point k , we estimated the effect of each
18 PDC_1 to PDC_k on LR_k by the weighted OLS regression (regressing LR_k on PDC_1 to PDC_k
19 simultaneously) using the weights obtained from the first step. The effect of the contemporaneous
20 PDC_k on LR_k is the direct effect, while the effect of intervening on a previous PDC on LR_k
21 incorporates both the direct effect and the indirect effect via the earlier LR measures. If a covariate
22 failed the balancing test, we included it as a covariate in the regression (28). See the Appendix for
23 further technical details on the causal modelling and estimation methods.

24 We conducted LDL analyses (Analysis 1 and Analysis 2) separately for patient groups with and
25 without prevalent CVD.
26

1

2 Genome Wide Association Studies (GWAS)

3

4 We performed GWAS analysis of PDC traits using REGENIE v3.3(29) adjusted for age, sex, assessment
5 centres, and genotyping microarray (2 categories) to identify genetic variants predicting PDC.
6 REGENIE adjusts for sample relatedness and population structure during analysis. . We studied
7 16million genetic variants from the imputed data released by UKB (category 100319, (30)), where the
8 imputation quality was >30% and Minor Allele Frequencies were >0.1%. P-values < 5×10^{-8} were taken
9 as genome-wide significant.

10

11 Time-to-event analysis of long-term clinical outcomes

12

13 We conducted a time-to-event analysis beginning 3 months post-statin initiation. Patients exited the
14 model at the event date or end of follow-up. The model adjusted for sex, age at first statin
15 prescription, education level, assessment centre, and baseline binary event diagnosis. Adherence
16 levels were categorized as $\leq 50\%$, 50–95%, and $>95\%$ and used as exposure.

17

18 Unless otherwise stated all analyses were performed using R v4.3.1 and the package `survival` (v3.5-
19 8) was used for time-to-event analysis. Package `CBPS` (version: 0.23) was used for longitudinal
20 causal analyses.

21

22

1 Results

2 LDL

3

4 Of the 76,909 UKB participants ever prescribed statins in their available primary care data (33% of all
5 participants with primary care data), 41,316 had LDL measurements prior to statin initiation (<180
6 days pre-statin; Mean: 32, SD: 38, Median: 16) and within 1 year of first statin prescription (post-
7 statin LDL measurement; Mean: 123 days, SD: 96, Median: 92). Of these, 37,297 were of European-
8 like genetic ancestry (Figure 1). Mean age at statin initiation was 61 years (SD 7.3, range 40-79), and
9 44.4% were female (Table 1).

10

11 *Analysis 1: Cross-sectional analysis of LDL change*

12

13 The mean pre- and post-statin LDL was 3.8 mmol/L (SD 1.05) and 2.5 mmol/L (SD 0.95) respectively.
14 Individuals with pre-statin CVD (secondary prevention group) had, on average, baseline LDL levels
15 0.59 units lower (95% CI 0.62 to 0.56, $p=7.6 \times 10^{-355}$) and pre-statin CVD was associated with a 0.12
16 unit increase in dose intensity (95% CI 0.1 to 0.13, $p=3.7 \times 10^{-71}$) than those without pre-statin CVD.

17 Across all individuals, we found a U-shaped relationship between LDL-c reduction and the time to
18 first follow-up, with a maximum reduction for those returning for follow-up between weeks 5-8 of
19 26% (95%CI 23-29) (Appendix - Figure 1). The mean PDC_1 value was 94.7% (IQ 85.5-105) and the
20 prevalence of PDCs is as follows: PDC = 50% (8%), PDC = 50-95% (33%), and PDC > 95% (59%).

21 Stratifying the analysis by Individuals who had $\leq 95\%$ and $>95\%$ PDC_1 revealed the U-shape was
22 driven by the former patient group rather than the latter (Appendix - Figure 2).

23 In the primary prevention group, the high PDC group, which was also more adherent to post-statin
24 LDL follow-up (Figure 5-A), achieved the greatest LDL reduction, with a 40% reduction (LDL ratio
25 ~ 0.6) by 15 weeks. In the secondary prevention group, post-statin LDL levels were reduced to below
26 2 mmol/L, as recommended by NICE guidelines, in 31% of individuals. Those in the high PDC group

1 (>95%) showed the greatest LDL reduction by 15 weeks, with 57% achieving LDL ≤ 2 , compared to
2 23% in the low PDC group ($p=2 \times 10^{-31}$). The group sizes and densities indicate greater adherence to
3 LDL check-ups in the high PDC group, which was associated with the most significant LDL reduction
4 over time (Figure 5-B).

5 Finally, *SLCO1B1**5 genotype was seen to alter the relationship between follow-up time and LDL
6 reduction: the homozygotes (i.e., rs4149046 CC genotype) had the smallest LDL-c reduction in the
7 quadratic model in year 1 (Appendix - Figure 3). The polygenic score for Alzheimer's disease was
8 associated with a reduced LDL reduction (Coef: 4.5×10^{-3} , 95% CI 1.7×10^{-3} to 7.3×10^{-3} , $p=0.002$).
9 Supplementary Table 4 for the other polygenic scores.

10

11 *Analysis 2: Longitudinal analysis on LDL trajectories*

12 Using the weights obtained from the stage 1 weighting procedure, we examined the weighted pair-
13 wise correlations between PDC_k and each of its constituent variables with both the Pearson and
14 Spearman correlation. All covariates except for the patient visit times had average correlations (both
15 Pearson and Spearman) with PDC_k that fell below the 0.1 threshold (Supplementary Tables 5-6). We
16 therefore explicitly included patient visit times as covariates in the stage two weighted regressions.

17 Model-derived estimates of the effects of PDCs on LR with the 95% confidence intervals using the
18 CBPS weights are shown in Figure 4 Panel (c). All three PDC measures have significant ($p < 0.05$)
19 negative effects on LDL reduction (Supplementary Table 7). Firstly, for all three time points, the most
20 recent PDC measure, PDC_k , had the largest estimated effect on LR_k . Secondly, the initial adherence
21 level PDC_1 had a larger estimated impact on LR_3 than PDC_2 . Thirdly, while PDCs at different time
22 points had different estimated effects, the total effect estimate (expressed as a summation across
23 the time points) remained remarkably constant. For illustration, we calculated the average difference
24 in LDL reduction if all individuals' PDCs were set to 100 versus if all PDCs were set to 0 in Figure 4
25 Panel (d). This revealed that LDL percentage reductions of approximately 35% can be achieved for
26 sustained, full adherence to statins. Another relevant measure for each patient is the LDL reduction if

1 they could maintain their own personal maximum PDC level across each time point compared to
2 their observed PDC level at each time point. Our model suggests this difference would be
3 approximately 6%. We conducted the analysis using the non-parametric option of the CBPS method
4 (npCBPS) as well, with consistent results (Supplementary Table 8).

5 The longitudinal analysis of LDL changes with sustained PDC intervention in the primary prevention
6 subgroup (n = 10,971) yielded results consistent with those in the main analysis of the full sample
7 (Supplementary Tables 9-10). There was insufficient statistical power for analysis of the secondary
8 prevention subgroup due to small sample size (n = 2,344).

9

10 Predictors of PDC₁

11

12 Male participants (Coef -0.64, 95% CI -1.25 to -0.04, p=0.03) and current smokers (Coef -3.66, 95% CI
13 -4.61 to -2.40, p=7x10⁻¹⁴) were less adherent (lower PDC). Participants with higher pre-statin LDL-c,
14 higher educational attainment, more prevalent diseases, and who were older at first statin
15 prescription were more adherent (higher PDC). BMI was not associated with PDC (Coef 0.01, 95% CI -
16 0.05 to 0.07). Genetic predictors of reduced PDC included liability to schizophrenia (Coef_{top 20%} -1.94,
17 95%CI -2.69 to -1.19), whilst genetic liability to cardiovascular diseases increased PDC (Coef_{top 20%}
18 1.30, 95%CI 0.55 to 2.05) (Supplementary table 11). rs4149046 CC (*SLCO1B1**5), was not significantly
19 associated with PDC in linear regression models (Coef -0.002, 95% CI -1.56 to 1.55, p=0.1). See table
20 2 for the details.

21

22 GWAS

23

24 We performed GWAS analysis of two PDC traits, using REGENIE to analyse 16million genetic variants
25 with MAF>0.001 and INFO>0.3. One genetic variant was significantly (p<5*10⁻⁸) associated with "PDC

1 until LDL1” in linear regression models: rs548267220 (C>G, MAF=0.006, INFO=0.8, BETA=12.7,
2 $p=4.8*10^{-8}$), intronic for gene *RUNX3*. Two genetic variants were significantly ($p<5*10^{-8}$) associated
3 with “PDC >95% vs. <=95%” in logistic regression models: rs9439705 (G>A, MAF=0.92, INFO=0.99,
4 BETA=-0.13, $p=1.4*10^{-8}$), located 30kbp from *KLHDC7A*, and rs75103961 (C>T, MAF=0.002,
5 INFO=0.92, BETA=-0.77, $p=4.0*10^{-8}$), intergenic on chromosome 12. For details see Supplementary
6 Tables 12.

7
8 rs548267220 has not appeared in the GWAS catalog previously (5th Nov 2024), though there are 142
9 entries for *RUNX3* including haematological traits, kidney function, allergic diseases, blood pressure,
10 and obesity. Neither rs9439705 nor rs75103961 have appeared in the GWAS catalog previously (5th
11 Nov 2024). *KLHDC7A*, encoding protein kelch domain containing 7A, has previously been linked to
12 LDL cholesterol, kidney function, and other traits in the GWAS catalog.

13

14

15 Time-to-event analysis

16

17 We analysed up to 55,784 patients across three PDC groups (low: <50%, moderate: 50-95%, high:
18 >95%). Hypertension was the most prevalent disease (43%). Key differences between PDC groups
19 included CHD (19% in low PDC, 17% in high PDC), cataract (8% in low PDC, 10% in high PDC), and
20 chronic kidney disease (8.5% in low PDC, 10% in high PDC) (Supplementary Table 13).

21 Patients with high PDC had a decreased risk of ischemic stroke (HR 0.67, 95% CI 0.54-0.83, $p=10^{-2}$),
22 coronary heart disease (HR 0.78, 95% CI 0.73-0.84, $p=2*10^{-8}$), transient ischemic attack (HR 0.79, 95%
23 CI 0.68-0.92, $p=4*10^{-2}$); but an increased risk of cataract (HR 1.20, 95% CI 1.07-1.34, $p=4*10^{-2}$), iron
24 deficiency anaemia (HR 1.3, 95% CI 1.09-1.54, $p=5*10^{-2}$), chronic kidney disease (HR 1.19, 95% CI
25 1.06-1.32, $p=4*10^{-2}$) and enthesopathy of the upper limbs (HR 1.16, 95% CI 1.06-1.26, $p=4*10^{-2}$)

1 compared to those with low PDC (Table 3). P-values were adjusted using the Benjamini-Hochberg
2 correction method.

3

4 **Discussion**

5

6 Using real-world data, we identify substantial variability in the time to first on-treatment LDL and
7 adherence to statin medication, highlighting a gap between NHS guidelines and the actual level of
8 achieved treatment response monitoring. Our work highlights important predictors of reduced statin
9 effectiveness, such as pharmacogenetic variants, polygenic scores, but most of all, sustained
10 adherence to statin medications. Higher adherence to treatment was associated with greater
11 adherence to therapy overall, leading to more significant LDL reductions over time. This suggests that
12 improved service provision and patients' factors—such as increased statin use, better understanding
13 of treatment by patients, more frequent GP visits for LDL checks, and tailored treatments addressing
14 low-adherence risk factors—could enhance efficacy, safety, and long-term patient outcomes.

15

16 When quantifying adherence using the Proportion of Days Covered (PDC) measure, we assumed that
17 patients adhered to their prescribed medication regimens. However, the UKB lacks detailed data on
18 actual medication intake and GP dosage instructions (e.g., whether a patient is taking 2x40 mg versus
19 1x80 mg). Therefore, we restricted our PDC analysis to patients with a PDC of less than 200%, under
20 the assumption that they might be taking two tablets per day as prescribed. As a result, our
21 adherence rate is higher than in previous Finnish cohorts, likely due to their smaller PDC cut-off and
22 the fact that 96% of statin users in Finland were prescribed one tablet daily (31,32). Despite these
23 differences, the extent to which LDL change is strongly predicted by our PDC measure indicates that
24 it is capturing the underlying relationship between medication adherence and effectiveness of
25 treatment.

26

1 To add further weight to this, our finding of an increased risk of iron deficiency anemia (IDA) in
2 patients with higher statin adherence (HR 1.30: 95% CI 1.12–1.70) is consistent with previous
3 research: a hypothesis-free analysis of a Korean cohort (7,847 statin users vs. 39,235 non-users)
4 similarly identified a higher risk of IDA (95% CI 2.11–12.03) linked to statin use among over 100
5 tested diseases, supported by Mendelian Randomisation analysis (33). The possible mechanism
6 between iron and statins can be linked to statin’s capacity to favourably modulate iron homeostasis
7 (34), which might be beneficial for CVD risk as abnormal iron accumulation in cells increases
8 oxidative stress, vascular inflammation, and the development of atherosclerosis (34,35). A review of
9 over 313,200 patients (36) found a moderate increase in cataract risk with statin use in cohort
10 studies, which had long follow-up periods (~ >5 years) and various baseline characteristics, while
11 case-control studies and randomized trials showed no significant risk. Although our findings support
12 previous cohort studies with long-term follow-up, the review concludes that there is likely no
13 significant link between statin treatment and cataracts, emphasizing that the benefits of statin
14 therapy outweigh potential risks.

15

16 In exploratory analysis to evaluate the effect of a sustained PDC intervention on LDL reduction using
17 longitudinal causal models, we made a critical assumption that all confounders between PDC and
18 LDL across three time points were observable and could be controlled for in the analysis using
19 inverse probability weighting. The resulting estimates suggest that approximately 35% reductions in
20 LDL are eminently achievable in the UKB population under full adherence. Ensuring high adherence
21 in the most recent period was estimated to have, unsurprisingly, the largest influence on patients’
22 LDL reduction, but ensuring high adherence in their initial treatment phase was interestingly the next
23 most effective way to achieve this. This again highlights the importance of early follow-up monitoring
24 after statin initiation. A possible direction for future research is to utilize pharmacogenetic predictors
25 of treatment adherence within a Mendelian randomization framework (37,38), to facilitate this
26 analysis under a relaxation of the observable confounder assumption. Unfortunately, we did not

1 identify any suitably strong genetic predictors of PDC in the UKB primary data reaching genome wide
2 significance. This is consistent with findings from the FinnGen/EstoniaBiobank study (18), which also
3 failed to identify any genome-wide significant genetic variants associated with PDC in statin users
4 (18). Future work will expand this search to include other cohort studies and to consider other
5 measures of adherence beyond PDC.

6 A PDC below 50% is particularly important, as it indicates that even if patients are taking 1/2 tablets
7 daily, their adherence is still insufficient to fully meet the prescribed regimen. Consequently, patients
8 with low adherence had a 27% higher risk of coronary heart disease and a 50% increased risk of
9 ischemic stroke compared to those with high adherence. Understanding and improving adherence is
10 crucial to reducing cardiovascular events. Predictors of low adherence to statins included genetic
11 liability to schizophrenia. Likewise, a study from the Estonian Bank (39) highlighted that genetic
12 liability to schizophrenia was associated with schizophrenia diagnoses in this cohort. For these
13 patients, higher adherence to antipsychotic medications reduced metabolic syndromes, which are
14 known to elevate the risk of heart disease and stroke. Our research suggests that this could be due
15 indirectly to higher levels of statin adherence rather than a direct effect of antipsychotic medication.

16
17 When investigating genetic predictors of on-statin LDL response, we initially focused on the
18 pharmacogenetic polygenic risk score reported by Mayerhofer et al (40). We did not find the score to
19 be predictive of an LDL reduction as a whole. rs2900478, a single SNP included in the polygenic
20 score, was highly correlated ($r^2=0.96$) with the well-known statin pharmacogenetic variant
21 *SLCO1B1**5 (rs4149056). We suspected that the score effect could be attributed to the statin
22 transporter gene *SLCO1B1*, and we continued our analysis focusing on this gene alone. Consistent
23 with our previous work, where we found that *SLCO1B1**5 was associated with discontinuation of
24 statin therapy (41), we here report that *5 is associated with attenuated LDL reduction following
25 statin initiation. We found no association between *SLCO1B1**5 and PDC (adherence) perhaps
26 because, by definition, those with multiple prescriptions meeting inclusion criteria have not yet

1 discontinued. When taken together, our analyses support the known pharmacogenetic effect of *5
2 on statin effectiveness, and potential impact on adverse events.

3

4 In conclusion, this is the first study, to our knowledge, that uses a large cohort with linked genetics
5 and primary care health records to examine the effect of adherence on LDL change and long-term
6 outcomes, and identify risk factors for adherence using genetic and causal inference models. Our
7 work emphasizes insufficient LDL control with a need for LDL monitoring in line with clinical
8 guidelines, and a clear clinical benefit of adhering to prescribing guidance, which reinforces the
9 importance of patient education on statin indication and adherence.

10

11

12

13

14

15

16

17

18

19

20

21

22

1

2

3

4

5 **Declarations**

6 ***Ethics approval and consent to participate***

7 The collection and use of UK Biobank data was approved by the North West Multi-Centre Research

8 Ethics Committee (Research Ethics Committee reference 11/NW/0382).

9

10 ***Consent for publication***

11 UK Biobank participants gave informed consent for their data to be used in health research.

12

13 ***Availability of data and materials***

14 The data that support the findings of this study are available from UK Biobank but restrictions apply

15 to the availability of these data, which were used under license for the current study, and so are not

16 publicly available. Data are however available from the authors upon reasonable request and with

17 permission of UK Biobank (www.ukbiobank.ac.uk/register-apply). GWAS summary statistics will be

18 posted to Zenodo and a DOI provided following acceptance.

19

20 ***Competing Interests***

21 JB is employed part-time by Novo Nordisk, his company work is unrelated to that presented here. DG

22 is the Chief Executive Officer of Sequoia Genetics, a private company that works with investors,

23 pharma, biotech, and academia by performing research that leverages genetic data to help inform

1 drug discovery and development. DG serves on the advisory board of, and has financial interests in,
2 several biotechnology companies, his company work is unrelated to that presented here.

3

4

5

6 ***Funding***

7 DT, XL, DG, LCP, and JB are funded by UK Medical Research Council research grant MR/X011372/1.
8 JM is funded by a National Institute for Health Research Fellowship (NIHR301445). This publication
9 presents independent research funded by the National Institute for Health Research (NIHR). This
10 study is supported by the NIHR Exeter Biomedical Research Centre (BRC). The views expressed are
11 those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health
12 and Social Care. The funders had no involvement in the study design; in the collection, analysis, and
13 interpretation of data; in the writing of the report; or in the decision to submit the paper for
14 publication.

15

16 ***Authors' contributions***

17 DT and XL generated data, performed analyses, interpreted results, created the figures, searched the
18 literature, and co-wrote the manuscript. JM and DG provided expert clinical interpretation of the
19 data and contributed to the manuscript. LP interpreted results, searched the literature, and co-wrote
20 the manuscript. JB oversaw interpretation, contributed to data analysis, and co-write the manuscript.

21

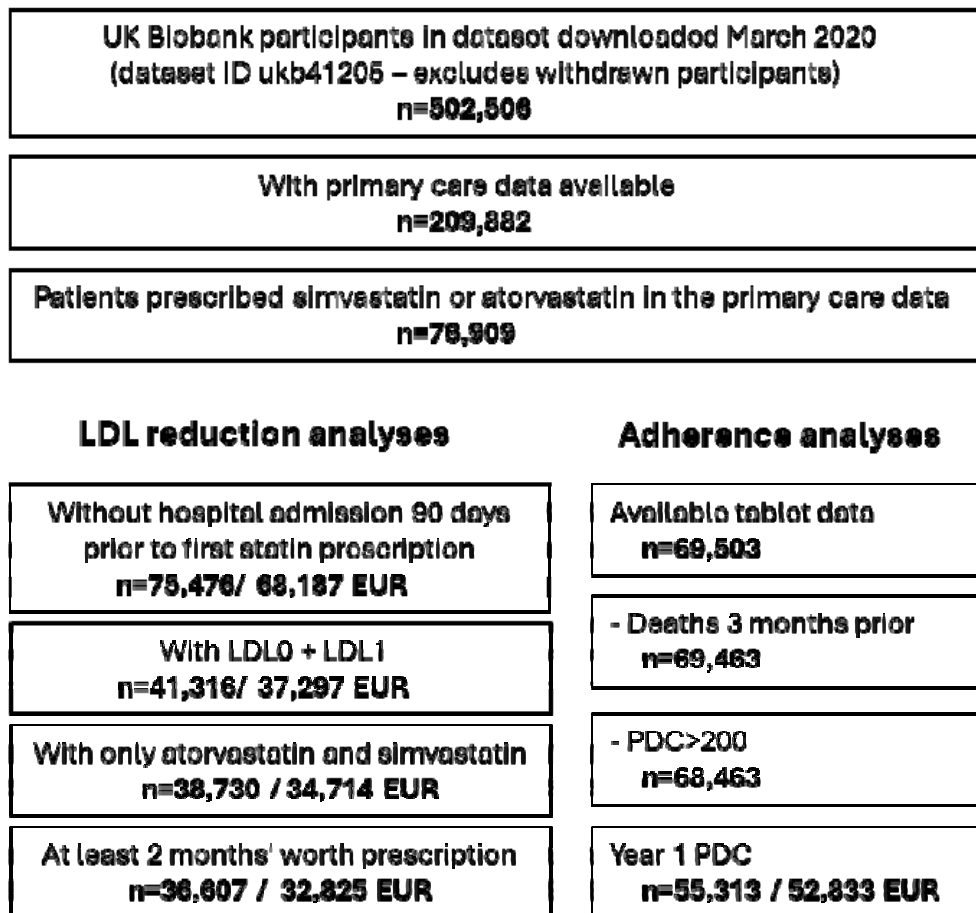
22 ***Acknowledgements***

23 Access to UK Biobank was granted under Application Number 14631. The authors would like to
24 acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in
25 carrying out this work. For the purpose of open access, the author has applied a 'Creative Commons
26 Attribution (CC BY)' licence to any Author Accepted Manuscript version arising from this submission.

1 **Figures:**

2 **Figure 1 – Flowchart of the UK Biobank cohort**

3



4

5

6 LDL0: pre-statin LDL-c measurement, LDL1: post-statin LDL-c measurement, EUR: Genetically

7 European ancestry

8

9

10

11

12

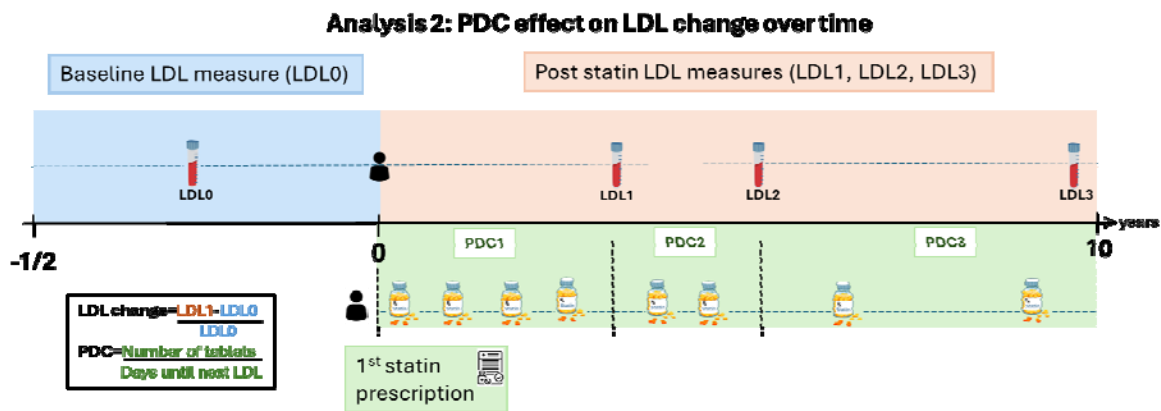
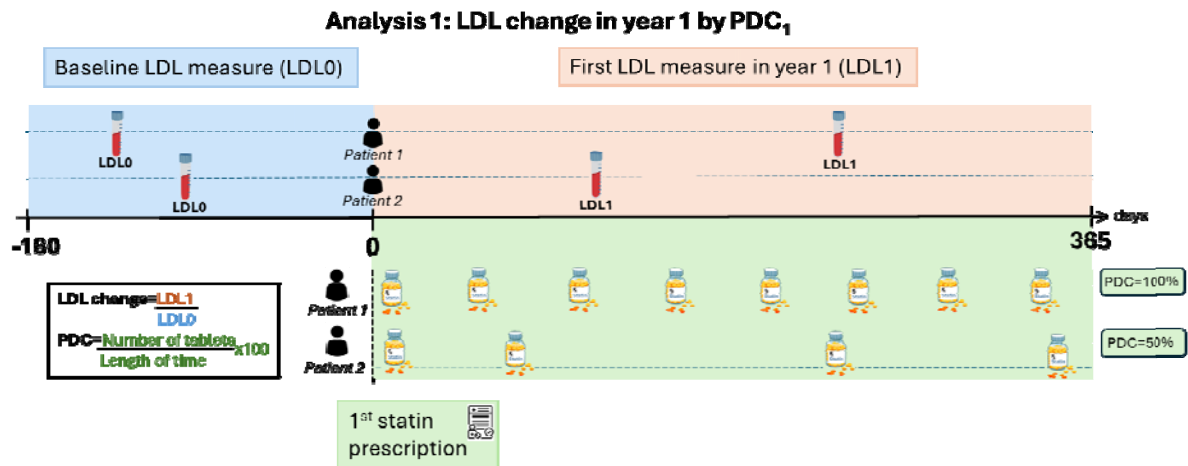
13

14

15

1 **Figure 2 – Design of the analyses**

2



3

4

5

6 For each patient with a history of statin treatment initiation as well as pre and post treatment LDL
 7 measures (n=69,503), we sought all available data on the date and quantity their prescription. We
 8 then calculated the number of tablets prescribed to each individual in a given time period, to create
 9 the PDC variable.

10 Analysis 1: We estimated LDL change for varying the adherence in the first year of treatment, which
 11 we call PDC_1 . If patients received a new prescription just before the end of year 1, we included the
 12 time until the next prescription in the 'length' calculation for PDC_1 . However, if there were no
 13 subsequent prescriptions, we excluded that latest prescription from the calculation (n=7,446). We
 14 also excluded patients who died within 3 months of starting statins due to pre-existing illness (n=39)
 15 and those with a $PDC_1 > 200\%$ (n=714).

16 Analysis 2: We attempted to evaluate how intervening to increase statin adherence would affect a
 17 person's LDL change over a longer sustained period, adjusting for fixed confounders of PDC and LDL,
 18 as well as time-varying confounding of previous PDC and LDL measures on their future values. For
 19 patients prescribed statins, we included individuals with at least three post-statin LDL measures over
 20 a ten-year window for the analysis (n=13 315).

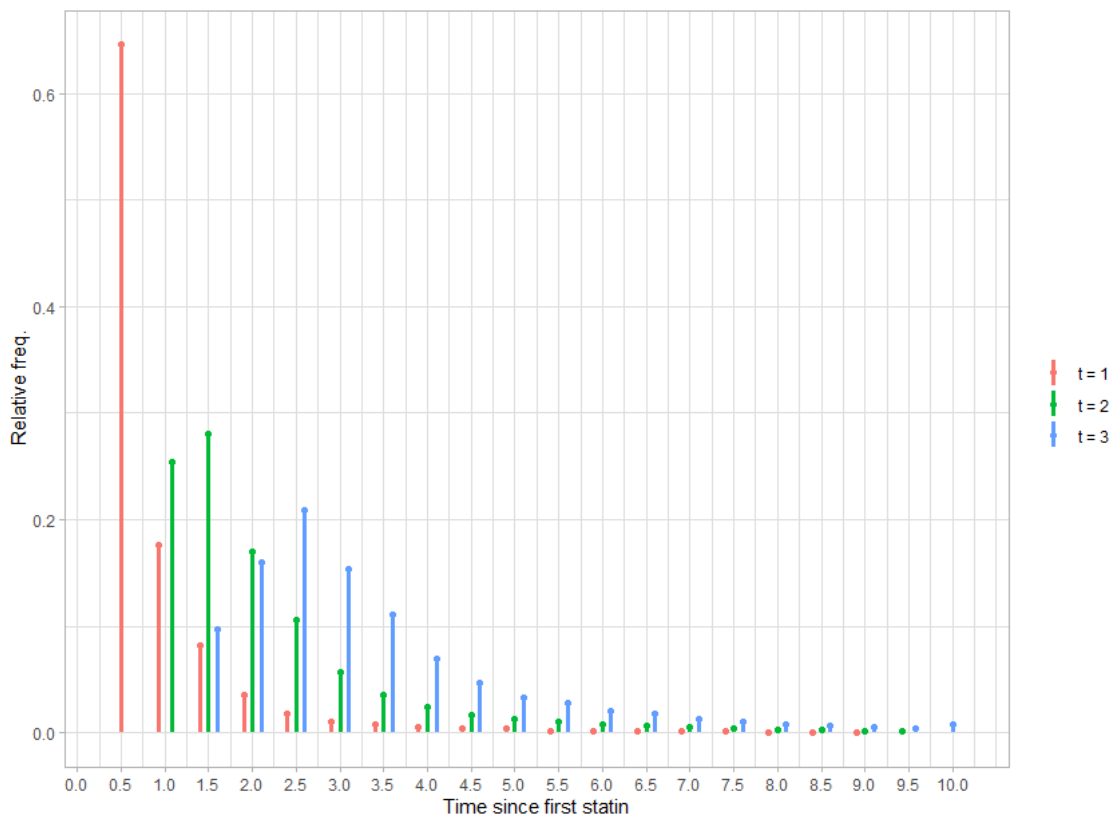
21

1 **Figure 3 – Frequency distribution of first three post-statin LDL measures times**

2

3

4



5

6 The x-axis shows the time length by year, and y-axis shows the relative frequency

7

8

9

10

11

12

13

14

15

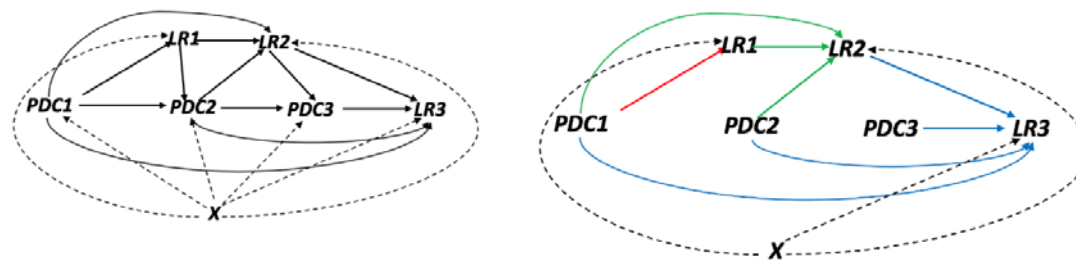
16

17

18

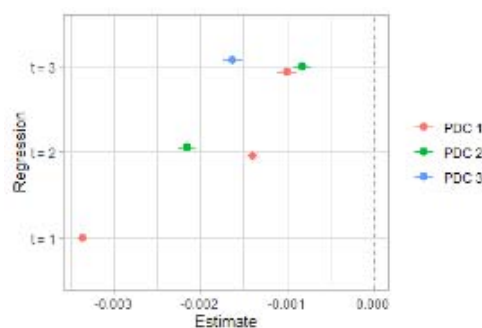
1 **Figure 4 – Causal diagrams illustrating the model set up of and results of the inverse**
 2 **probability weighted analysis.**

3

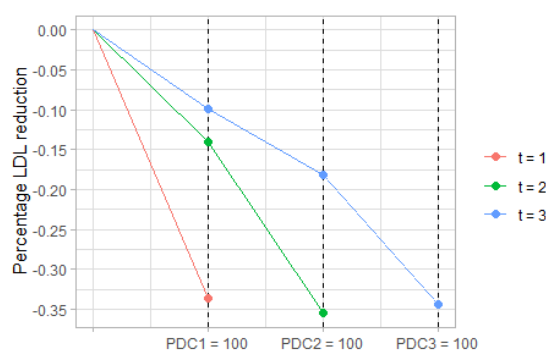


(a): Causal diagram illustrating the conditional dependencies and independencies between three adherence measurements PDC1-PDC3 and three outcome measurement LR1-LR3 in the presence of a set of confounders X.

(b): Causal diagram illustrating the conditional dependencies and independencies between PDC and LR after inverse probability weighting.



(c) Estimates and 95% confidence intervals of the effects of PDCs on LR for each time point from the weighted OLS using the CBPS weights.



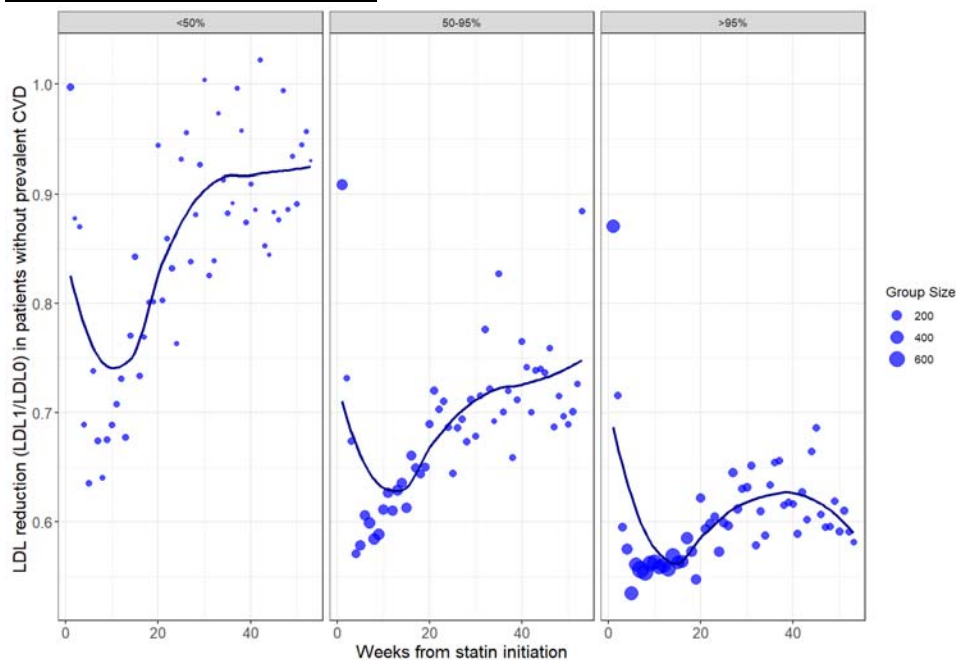
(d) The difference in percentage LDL reduction between the cases with all PDCs equal to 0 and all PDCs equal to 100. The red line represents the effect of PDC1 on LR1, the two segments of the green line (from left to right) represent the effect of PDC1 and PDC2 on LR2 respectively, and the three segments of the blue line (from left to right) represent the effect of PDC1, PDC2 and PDC3 on LR3 respectively.

4
 5
 6
 7
 8
 9
 10
 11

1 **Figure 5 – LDL Change in Primary Prevention & Mean LDL in Secondary Prevention Groups**

2

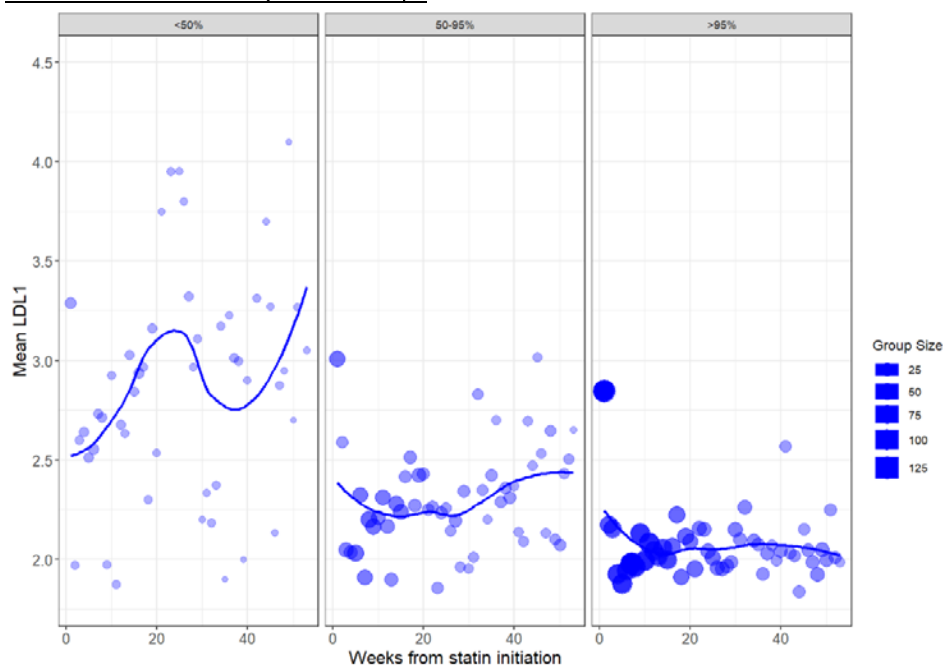
3 **A- LDL Change in the Primary Prevention Group and Distribution of Follow-Up LDL**
4 **Measurements by PDC Groups**



5

6 Change in LDL levels (LDL1/LDL0) after statin initiation, stratified by treatment group. Points represent the
7 average LDL change for each time bin (weeks from statin initiation) in patients without prevalent cardiovascular
8 disease. The size of the points reflects the number of patients in each group. The plot is divided by three
9 treatment groups, categorised by the percentage of patients in each group (<50%, 50-95%, and >95% of the
10 population).

11 **B- Mean Post-Statin LDL in the Secondary Prevention Group and Distribution of Follow-Up**
12 **LDL Measurements by PDC Groups**



13

14 The mean first post-statin LDL measurements in secondary prevention group.

1 **Table 1 – Descriptive table of patients who have GP prescribed statin**

2

Female (N, %)	32,545 (42.3%)
Treatment duration (years)	0.01-25 (mean 6.3)
Number of participants with PDC ₁	55,313
PDC ₁ <50 (%)	8%
PDC ₁ = 50-95 (%)	33%
PDC ₁ >95 (%)	59%
Age at first statin (mean)	61 (SD 7.3)
PDC ₁ <50	59 (SD 7.6)
PDC ₁ = 50-95	60 (SD 7.6)
PDC ₁ >95	62 (SD 7.1)
BMI kg/m ² (mean)	29 (SD 5 – IQR 26-31)
PDC ₁ <50	29 (SD 5 – IQR 26-31)
PDC ₁ = 50-95	29 (SD 5 – IQR 26-31)
PDC ₁ >95	29 (SD 5 – IQR 26-31)
LDL pre statin (GP) mmol/L (mean)	3.8 (SD 1.05)
PDC ₁ <50	3.9 (SD 1.01)
PDC ₁ = 50-95	3.8 (SD 1.06)
PDC ₁ >95	3.8 (SD 1.07)
No Prevalent CVD (primary prevention)	4 (SD 1)
Prevalent CVD (secondary prevention)	3.4 (SD 1)
LDL post statin (GP) mmol/L (mean)	2.5 (SD 0.95)
PDC ₁ <50	3.2 (SD 1.1)
PDC ₁ = 50-95	2.6 (SD 0.9)
PDC ₁ >95	2.3 (SD 0.8)
No Prevalent CVD (primary prevention)	2.6 (SD 0.9)
Prevalent CVD (secondary prevention)	2.3 (SD 0.8)
CVD*	
Pre statin (N)	17,849
PDC ₁ <50 (%)	34%
PDC ₁ = 50-95 (%)	37%
PDC ₁ >95 (%)	40%
Post statin (N)	16,250
PDC ₁ <50 (%)	25%
PDC ₁ = 50-95 (%)	22%
PDC ₁ >95 (%)	21%

3

4 PDC₁: Proportion of days covered in year 1 of statin initiation; CVD: Cardiovascular diseases included:
 5 Abdominal Aortic Aneurysm, Coronary Heart Disease, Peripheral Arterial Disease, All Types of
 6 Strokes, and Transient Ischemic Attack. Post statin: Within one year of last statin prescription.

7

1 **Table 2 - Associations between risk factors and PDC (proportion of days covered) as**
 2 **adherence measure using linear regression in GP-prescribed statin patients in UK Biobank**

Exposure	n	Coef	95%	CI	p
Observational[^]					
Sex (male)		-0.64	-1.25	-0.04	3.7E-02
Age at first statin		0.34	0.30	0.39	5.8E-54
BMI		0.01	-0.05	0.07	7.5E-01
LDL pre statin (GP)		0.55	0.30	0.80	1.3E-05
Smoking					
Never smoked	24932	<i>ref</i>	-	-	-
Ex-smoker	22982	0.63	-0.003	1.26	5.1E-02
Current smoker	7014	-3.66	-4.61	-2.70	7.2E-14
Education					
None	15583	<i>ref</i>	-	-	-
CSEs	1584	0.17	-1.63	1.98	8.5E-01
GCSEs/O-levels	6729	1.38	0.38	2.38	6.9E-03
A-levels/NVQ/HND/HNC	9403	1.24	0.32	2.15	7.9E-03
Prof.qual	7952	1.92	0.98	2.87	6.5E-05
Degree	13110	2.52	1.69	3.34	2.4E-09
Multimorbidity*					
No add. Disease	3127	<i>ref</i>	-	-	-
1 Disease	7644	2.85	1.33	4.37	2.4E-04
2 Diseases	9128	3.97	2.49	5.45	1.5E-07
3 Diseases	8584	3.87	2.38	5.36	3.7E-07
4 Diseases	7087	4.30	2.78	5.82	3.2E-08
4+ Diseases	19743	4.44	3.05	5.84	4.4E-10
Polygenic Scores[`]					
SCZ	49873	-0.58	-0.81	-0.34	5.9E-05
CAD	49873	0.49	0.25	0.72	8.4E-04
CVD	49873	0.44	0.20	0.67	3.4E-03
Genotype					
rs4149056 TT	36079	<i>ref</i>	-	-	-
rs4149056 TC	12759	0.09	-0.45	0.62	7.56E-01
rs4149056 CC	1162	-0.002	-1.56	1.55	9.98E-01

3 Positive estimates (Coef: Coefficient) have increased PDC.

4 [^] Age at first statin: age at the first prescription by GP

5 *Multimorbidity includes 84 diseases including cardiovascular diseases. No add. Disease represents
 6 patients with no disease prior to their first statin prescription.

7 [`]Three of 36 polygenic scores derived by Genomics PLC tested are significant after The Benjamini-
 8 Hochberg multiple testing adjustments. The units are scaled; distribution has a zero mean and unit
 9 variance. SCZ: Schizophrenia, CAD: Coronary Arteria Disease, CVD: Cardiovascular Disease. Polygenic
 10 scores are continuous, and genotype are tested in European-like participants only to minimise
 11 population stratification effect.

12

13

1 **Table 3- Associations between future diseases and PDC (low/high) PDC>95 versus PDC<50**

2

3

Outcome	% high PDC (low PDC)	n case high PDC	n case total	HR	95%	CI	p_bh
CHD	16.62 (18.56)	5455	9421	0.78	0.73	0.84	2.E-08
Isch Stroke	1.56 (2.30)	513	928	0.67	0.54	0.83	1.E-02
Cataract	9.68 (7.81)	3177	5237	1.20	1.07	1.34	4.E-02
CKD	10.04 (8.57)	3297	5370	1.19	1.06	1.32	4.E-02
Enthesopathy U.	12.62 (12.66)	4144	7162	1.16	1.06	1.26	4.E-02
TIA	4.00 (4.73)	745	2221	0.79	0.68	0.92	4.E-02
IDA	3.93 (3.38)	1289	2145	1.30	1.09	1.54	5.E-02

4

5

6 We performed a time-to-event analysis in statin patients, starting 3 months post-initiation.
7 Patients exited the model at either the event date or the end of follow-up. The model
8 adjusted for sex, age at first statin use, education, assessment center, and prevalent
9 diagnosis of the event (binary). We compared adherence levels ($\leq 50\%$, $50-95\%$, and $>95\%$),
10 showing here on results for the high adherence group (PDC $>95\%$).

11

12 Outcomes were derived from General Practice (GP) and hospital records. Abbreviations: CHD
13 (Coronary Heart Disease), Isch (Ischemic), CKD (Chronic Kidney Disease), Enthesopathy U
14 (Upper), TIA (Transient Ischemic Attack), IDA (Iron Deficiency Anemia), % (prevalence), n
15 (total number of cases), HR (Hazard Ratio), p_bh (Benjamini-Hochberg corrected p-value).

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

1 REFERENCES

2

3

4

5 1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden
6 of Cardiovascular Diseases and Risk Factors, 1990–2019. *J Am Coll Cardiol*. 2020
7 Dec;76(25):2982–3021.

8 2. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from
9 90,056 participants in 14 randomised trials of statins - PubMed [Internet]. [cited 2024 Oct 22].
10 Available from: <https://pubmed.ncbi.nlm.nih.gov/16214597/>

11 3. Recommendations | Cardiovascular disease: risk assessment and reduction, including lipid
12 modification | Guidance | NICE [Internet]. NICE; 2023 [cited 2024 Jun 24]. Available from:
13 [https://www.nice.org.uk/guidance/ng238/chapter/recommendations#assessing-response-to-](https://www.nice.org.uk/guidance/ng238/chapter/recommendations#assessing-response-to-treatment-5)
14 [treatment-5](https://www.nice.org.uk/guidance/ng238/chapter/recommendations#assessing-response-to-treatment-5)

15 4. Virani SS, Newby LK, Arnold SV, Bittner V, Brewer LC, Demeter SH, et al. 2023
16 AHA/ACC/ACCP/ASPC/NLA/PCNA Guideline for the Management of Patients With Chronic
17 Coronary Disease: A Report of the American Heart Association/American College of Cardiology
18 Joint Committee on Clinical Practice Guidelines. *Circulation* [Internet]. 2023 Aug 29 [cited 2024
19 Dec 10];148(9):e9–119. Available from:
20 <https://www.ahajournals.org/doi/10.1161/CIR.0000000000001168>

21 5. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA
22 Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American
23 College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.
24 *Circulation* [Internet]. 2019 Sep 10 [cited 2024 Dec 10];140(11):e596–646. Available from:
25 <https://www.ahajournals.org/doi/10.1161/CIR.0000000000000678>

26 6. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS
27 guidelines for the management of dyslipidaemias: Lipid modification to reduce cardiovascular
28 risk. *Atherosclerosis*. 2019 Nov;290:140–205.

29 7. Lloyd-Jones DM, Morris PB, Ballantyne CM, Birtcher KK, Covington AM, DePalma SM, et al. 2022
30 ACC Expert Consensus Decision Pathway on the Role of Nonstatin Therapies for LDL-Cholesterol
31 Lowering in the Management of Atherosclerotic Cardiovascular Disease Risk. *J Am Coll Cardiol*.
32 2022 Oct;80(14):1366–418.

33 8. Cholesterol Treatment Trialists' (Ctt) Collaboration. Efficacy and safety of more intensive
34 lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised
35 trials. *The Lancet*. 2010 Nov;376(9753):1670–81.

36 9. Wilkinson MJ, Lepor NE, Michos ED. Evolving Management of Low-Density Lipoprotein
37 Cholesterol: A Personalized Approach to Preventing Atherosclerotic Cardiovascular Disease
38 Across the Risk Continuum. *J Am Heart Assoc*. 2023 Jun 6;12(11):e028892.

39 10. Maningat P, Gordon BR, Breslow JL. How Do We Improve Patient Compliance and Adherence to
40 Long-Term Statin Therapy? *Curr Atheroscler Rep* [Internet]. 2013 Jan [cited 2024 Oct
41 21];15(1):291. Available from: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC3534845/>

- 1 11. Bosworth HB, Granger BB, Mendys P, Brindis R, Burkholder R, Czajkowski SM, et al. Medication
2 adherence: A call for action. *Am Heart J* [Internet]. 2011 Sep [cited 2024 Oct 11];162(3):412–24.
3 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002870311004753>
- 4 12. Rodriguez F, Maron DJ, Knowles JW, Virani SS, Lin S, Heidenreich PA. Association of Statin
5 Adherence With Mortality in Patients With Atherosclerotic Cardiovascular Disease. *JAMA*
6 *Cardiol*. 2019 Mar 1;4(3):206.
- 7 13. Hirsh BJ, Smilowitz NR, Rosenson RS, Fuster V, Sperling LS. Utilization of and Adherence to
8 Guideline-Recommended Lipid-Lowering Therapy After Acute Coronary Syndrome:
9 Opportunities for Improvement. *J Am Coll Cardiol* [Internet]. 2015 Jul 14 [cited 2024 Oct
10 21];66(2):184–92. Available from:
11 <https://www.sciencedirect.com/science/article/pii/S0735109715024754>
- 12 14. de Lemos JA, Blazing MA, Wiviott SD, Lewis EF, Fox KAA, White HD, et al. Early Intensive vs a
13 Delayed Conservative Simvastatin Strategy in Patients With Acute Coronary SyndromesPhase Z of
14 the A to Z Trial. *JAMA* [Internet]. 2004 Sep 15 [cited 2024 Oct 21];292(11):1307–16. Available
15 from: <https://doi.org/10.1001/jama.292.11.1307>
- 16 15. Aubert RE, Yao J, Xia F, Garavaglia SB. Is there a relationship between early statin compliance and
17 a reduction in healthcare utilization? *Am J Manag Care*. 2010 Jun;16(6):459–66.
- 18 16. Natarajan P, Young R, Stitzel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic Risk Score
19 Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From
20 Statin Therapy in the Primary Prevention Setting. *Circulation*. 2017 May 30;135(22):2091–101.
- 21 17. Mega JL, Stitzel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, et al. Genetic risk, coronary
22 heart disease events, and the clinical benefit of statin therapy: an analysis of primary and
23 secondary prevention trials. *The Lancet*. 2015 Jun;385(9984):2264–71.
- 24 18. Cordioli M, Corbetta A, Kariis HM, Jukarainen S, Vartiainen P, Kiiskinen T, et al. Socio-
25 demographic and genetic risk factors for drug adherence and persistence across 5 common
26 medication classes. *Nat Commun* [Internet]. 2024 Oct 23 [cited 2024 Dec 17];15(1):9156.
27 Available from: <https://www.nature.com/articles/s41467-024-53556-z>
- 28 19. Sadler MC, Apostolov A, Cevallos C, Ribeiro DM, Altman RB, Kutalik Z. Leveraging large-scale
29 biobank EHRs to enhance pharmacogenetics of cardiometabolic disease medications [Internet].
30 2024 [cited 2024 Oct 17]. Available from:
31 <http://medrxiv.org/lookup/doi/10.1101/2024.04.06.24305415>
- 32 20. Olarte Parra C, Daniel RM, Bartlett JW. Hypothetical Estimands in Clinical Trials: A Unification of
33 Causal Inference and Missing Data Methods. *Stat Biopharm Res* [Internet]. 2023 Apr 3 [cited
34 2024 Oct 21];15(2):421–32. Available from:
35 <https://www.tandfonline.com/doi/full/10.1080/19466315.2022.2081599>
- 36 21. Cardiovascular disease: risk assessment and reduction, including lipid modification. *Cardiovasc*
37 *Dis*.
- 38 22. Sampson M, Ling C, Sun Q, Harb R, Ashmaig M, Warnick R, et al. A New Equation for Calculation
39 of Low-Density Lipoprotein Cholesterol in Patients With Normolipidemia and/or
40 Hypertriglyceridemia. *JAMA Cardiol* [Internet]. 2020 May 1 [cited 2024 Dec 6];5(5):540. Available
41 from: <https://jamanetwork.com/journals/jamacardiology/fullarticle/2761953>

- 1 23. Casanova F, Tian Q, Atkins JL, Wood AR, Williamson D, Qian Y, et al. Iron and risk of dementia:
2 Mendelian randomisation analysis in UK Biobank. *J Med Genet* [Internet]. 2024 Jan 8 [cited 2024
3 Oct 23];jmg-2023-109295. Available from: [https://jmg.bmj.com/lookup/doi/10.1136/jmg-2023-](https://jmg.bmj.com/lookup/doi/10.1136/jmg-2023-109295)
4 109295
- 5 24. Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, et al. UK Biobank release and
6 systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits
7 [Internet]. 2022 [cited 2024 Jun 24]. Available from:
8 <http://medrxiv.org/lookup/doi/10.1101/2022.06.16.22276246>
- 9 25. Murrin O, Mounier N, Voller B, Tata L, Gallego-Moll C, Roso-Llorach A, et al. A systematic analysis
10 of the contribution of genetics to multimorbidity and comparisons with primary care data
11 [Internet]. 2024 [cited 2024 Jun 24]. Available from:
12 <http://medrxiv.org/lookup/doi/10.1101/2024.05.13.24307009>
- 13 26. Fong C, Hazlett C, Imai K. Covariate balancing propensity score for a continuous treatment:
14 Application to the efficacy of political advertisements. *Ann Appl Stat* [Internet]. 2018 Mar 1
15 [cited 2024 Oct 21];12(1). Available from: [https://projecteuclid.org/journals/annals-of-applied-](https://projecteuclid.org/journals/annals-of-applied-statistics/volume-12/issue-1/Covariate-balancing-propensity-score-for-a-continuous-treatment--Application/10.1214/17-AOAS1101.full)
16 [statistics/volume-12/issue-1/Covariate-balancing-propensity-score-for-a-continuous-treatment--](https://projecteuclid.org/journals/annals-of-applied-statistics/volume-12/issue-1/Covariate-balancing-propensity-score-for-a-continuous-treatment--Application/10.1214/17-AOAS1101.full)
17 [Application/10.1214/17-AOAS1101.full](https://projecteuclid.org/journals/annals-of-applied-statistics/volume-12/issue-1/Covariate-balancing-propensity-score-for-a-continuous-treatment--Application/10.1214/17-AOAS1101.full)
- 18 27. Austin PC. Assessing covariate balance when using the generalized propensity score with
19 quantitative or continuous exposures. *Stat Methods Med Res* [Internet]. 2019 May [cited 2024
20 Oct 21];28(5):1365–77. Available from:
21 <https://journals.sagepub.com/doi/10.1177/0962280218756159>
- 22 28. Bishop CD, Leite WL, Snyder PA. Using Propensity Score Weighting to Reduce Selection Bias in
23 Large-Scale Data Sets. *J Early Interv* [Internet]. 2018 Dec [cited 2024 Oct 21];40(4):347–62.
24 Available from: <https://journals.sagepub.com/doi/10.1177/1053815118793430>
- 25 29. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally
26 efficient whole-genome regression for quantitative and binary traits. *Nat Genet* [Internet]. 2021
27 Jul [cited 2024 Aug 31];53(7):1097–103. Available from:
28 <https://www.nature.com/articles/s41588-021-00870-7>
- 29 30. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with
30 deep phenotyping and genomic data. *Nature* [Internet]. 2018 Oct [cited 2022 Dec
31 12];562(7726):203–9. Available from: <http://www.nature.com/articles/s41586-018-0579-z>
- 32 31. Cordioli M, Corbetta A, Kariis HM, Jukarainen S, Vartiainen P, Kiiskinen T, et al. Socio-
33 demographic and genetic risk factors for drug adherence and persistence: a retrospective
34 nationwide and biobank study across 5 medication classes and 1 845 665 individuals [Internet].
35 2023 [cited 2024 Aug 13]. Available from:
36 <http://medrxiv.org/lookup/doi/10.1101/2023.10.09.23296740>
- 37 32. Aarnio EJ, Martikainen JA, Helin-Salmivaara A, Huupponen RK, Hartikainen JEK, Peura PK, et al.
38 Register-based predictors of adherence among new statin users in Finland. *J Clin Lipidol*
39 [Internet]. 2014 Jan [cited 2024 Nov 19];8(1):117–25. Available from:
40 <https://linkinghub.elsevier.com/retrieve/pii/S1933287413002973>
- 41 33. Ahn J, Lee S, Won S. Possible link between statin and iron deficiency anemia: A South Korean
42 nationwide population-based cohort study. *Sci Adv* [Internet]. 2023 Oct 27 [cited 2024 Oct
43 21];9(43):eadg6194. Available from: <https://www.science.org/doi/10.1126/sciadv.adg6194>

- 1 34. Mascitelli L, Goldstein MR. Might the beneficial effects of statin drugs be related to their action
2 on iron metabolism? QJM [Internet]. 2012 Dec 1 [cited 2024 Oct 21];105(12):1225–9. Available
3 from: <https://academic.oup.com/qjmed/article-lookup/doi/10.1093/qjmed/hcs204>
- 4 35. Zacharski LR, DePalma RG, Shamayeva G, Chow BK. The Statin–Iron Nexus: Anti-Inflammatory
5 Intervention for Arterial Disease Prevention. Am J Public Health [Internet]. 2013 Apr [cited 2024
6 Oct 21];103(4):e105. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3673278/>
- 7 36. Yu S, Chu Y, Li G, Ren L, Zhang Q, Wu L. Statin Use and the Risk of Cataracts: A Systematic Review
8 and Meta-Analysis. J Am Heart Assoc Cardiovasc Cerebrovasc Dis [Internet]. 2017 Mar 20 [cited
9 2024 Oct 21];6(3):e004180. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5523994/>
- 10 37. Bowden J, Madsen J, Goldman B, Iversen AT, Liang X, Vansteelandt S. Instrumental Variable
11 methods to target Hypothetical Estimands with longitudinal repeated measures data:
12 Application to the STEP 1 trial [Internet]. arXiv; 2024 [cited 2024 Dec 3]. Available from:
13 <http://arxiv.org/abs/2407.02902>
- 14 38. Bowden J, Pilling LC, Türkmen D, Kuo CL, Melzer D. The Triangulation Within a Study (TWIST)
15 framework for causal inference within pharmacogenetic research. Kutalik Z, editor. PLOS Genet
16 [Internet]. 2021 Sep 8 [cited 2023 Jan 25];17(9):e1009783. Available from:
17 <https://dx.plos.org/10.1371/journal.pgen.1009783>
- 18 39. Alver M, Kasela S, Haring L, Luitva LB, Fischer K, Möls M, et al. Genetic predisposition and
19 antipsychotic treatment effect on metabolic syndrome in schizophrenia: a ten-year follow-up
20 study using the Estonian Biobank. Lancet Reg Health - Eur [Internet]. 2024 Jun [cited 2024 Nov
21 20];41:100914. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2666776224000814>
- 22 40. Mayerhofer E, Malik R, Parodi L, Burgess S, Harloff A, Dichgans M, et al. Genetically predicted on-
23 statin LDL response is associated with higher intracerebral haemorrhage risk. Brain J Neurol.
24 2022 Aug 27;145(8):2677–86.
- 25 41. Türkmen D, Masoli JAH, Kuo C, Bowden J, Melzer D, Pilling LC. Statin treatment effectiveness and
26 the *SLCO1B1* *5 reduced function genotype: Long-term outcomes in women and men. Br J Clin
27 Pharmacol [Internet]. 2022 Jul [cited 2022 Dec 1];88(7):3230–40. Available from:
28 <https://onlinelibrary.wiley.com/doi/10.1111/bcp.15245>

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

1 **Appendix**

2

3 **Methods**

4 **Cross sectional modelling of LDL reduction explained by week of follow up**

5

6 We tested associations between LDL change and the time interval between their pre- and post-
7 treatment measurements, using a linear regression model, adjusted additionally for dose group of
8 first atorvastatin and simvastatin prescribed in year 1. Specifically:

9

10 $\log(Y) = \beta_0 + \beta_1 \cdot factor(T)$

11

12 Y = LDL change (LDL1/LDL0)

13 T = Time interval between their pre- and post-treatment measurements

14

15 We organised doses into a categorical variable based on NICE guidelines (21) for dose equivalence, as
16 shown below:

17

18 Table 1- Dose intensity classification based on the NICE Guideline

19

Drug	10 mg	20 mg	40 mg	80 mg
Simvastatin	low	middle	Middle	High
Atorvastatin	middle	high	High	High

20

21

1 Table 2- Prevalence of the dose intensity groups in individuals prescribed statins in the UK Biobank

2 GP data

Dose intensity	N total	%	N (in both LDL0 and LDL1 available)	%
low	5,808	7.7	2,355	6.4
middle	57,293	76.3	29,560	80.4
high	12,013	16.0	4,833	13.2
total:	75114		36,748	

3

4 We repeated this analysis in PDC_1 groups ($\leq 95\%$ and $>95\%$).

5

6

7 Modelling LDL reduction by genotype, PDC and time of follow up measure.

8

9 We tested associations between LDL change and rs4149056 using a linear model that allowed for a
10 quadratic effect of time to follow up on treatment in European-like participants and an interaction
11 between time on treatment and rs4149056:

12

13
$$E[Y] = \beta_0 + \beta_1 \cdot T_{off} + \beta_2 \cdot T_{on} + \beta_3 \cdot (T_{on})^2 + \beta_4 \cdot G + \beta_5 \cdot T_{on}G + \beta_6 \cdot (T_{on} \cdot G)^2 \quad (1)$$

14

15 Y = LDL change (LDL1/LDL0)

16 T_{off} = Time between pre-statin LDL measurement and the statin initiation

17 T_{on} = Time to first LDL measures after the statin initiation x (dose group).

18 G = rs4149056 (number of copies)

19

20 Model (1) additionally adjusted for 10 genetic principal components (PCs).

1 Next, we tested the association between LDL change and PDC_1 using model (1), but with PDC in
 2 place of the G variable, rs4149056. This analysis did not adjust for genetic PCs but accounted for the
 3 prevalence of CVD before statin use, including abdominal aortic aneurysm, atrial fibrillation,
 4 coronary heart disease, peripheral arterial disease, all types of stroke, and thromboembolic disease.

5 Causal modelling of hypothetical LDL reduction via sustained PDC intervention 6

7 Let the vector \overline{PDC}_{ik} denote the adherence history of patient i up to time point k so that:
 8 $\overline{PDC}_{ik} = (PDC_{i1}, \dots, PDC_{ik})$. We define LR_{ik}^a as the potential LDL reduction at time point k given
 9 that \overline{PDC}_{ik} takes the value \mathbf{a} . We assume that PDC_{i1} to PDC_{ik} exert linear additive effects on LR_{ik} .

10 More concretely, for each time point k , we define the estimand of interest as the contrast between
 11 the potential outcomes under PDC history profile $\mathbf{a} + 1$ and \mathbf{a} , which we assume obeys the
 12 following parametric relationship: $E(LR_{ik}^{\mathbf{a}+1}) - E(LR_{ik}^{\mathbf{a}}) = \sum_{j=1}^k \beta_k(j)$

13 where $k = 1, 2, 3$ and $\beta_k(j)$ represents the causal effect of PDC_{ij} on LR_{ik} with $j = 1, \dots, k$. We aim
 14 to estimate $\beta_k(j)$ with the inverse probability weighting (IPW) method under the assumption that all
 15 the confounders between PDC and LR are observable and can be controlled for in the analysis. Since
 16 we have a continuous exposure variable PDC, according to (20) and (26), the weight for individual i at
 17 time point k is given by

$$W_{i,k} = \prod_{j=1}^k w_{i,j} = \prod_{j=1}^k \frac{f(PDC_{i,j})}{f(PDC_{i,j} | L_{i,j})}$$

18 with $k = 1, 2, 3$. $f(PDC_{i,j})$ and $f(PDC_{i,j} | L_{i,j})$ are the marginal and conditional density functions
 19 respectively. $PDC_{i,j}$ is the observed value of the PDC for individual i at time point j and $L_{i,j}$
 20 summarizes all fixed covariates mentioned in the main text, the previous exposure $PDC_{i,j-1}$ and
 21 previous outcome $LR_{i,j-1}$ (for $j = 2, 3$), and the patient visit time, denoted by $t_{i,j}$.

22 We estimate each $w_{i,j}$ ($j = 1, 2, 3$) with the Covariate Balancing Propensity Score (CBPS) method
 23 proposed by (26). The method was implemented with the R functions 'CBPS' and 'npCBPS' (for the

1 non-parametric option) from the R package ‘CBPS’ ([https://cran.r-](https://cran.r-project.org/web/packages/CBPS/index.html)
2 [project.org/web/packages/CBPS/index.html](https://cran.r-project.org/web/packages/CBPS/index.html)). We ran the functions with default settings except that
3 we set ‘method = exact’ for CBPS. For each $j = 1, 2, 3$, we estimated $w_{i,j}$ by fitting the model of
4 $PDC_{i,j} \sim L_{i,j}$. For a given time point k ($k = 1, 2, 3$), $W_{i,k}$ is then obtained with $W_{i,k} = \prod_{j=1}^k w_{i,j}$. We
5 multiplied each $W_{i,k}$ by 10^{13} for the second stage weighted OLS to avoid zero truncation due to
6 machine tolerance issues. In the second-stage regression, we omit the intercept.

7

8

9

10

11

12

13

14

15

16

17

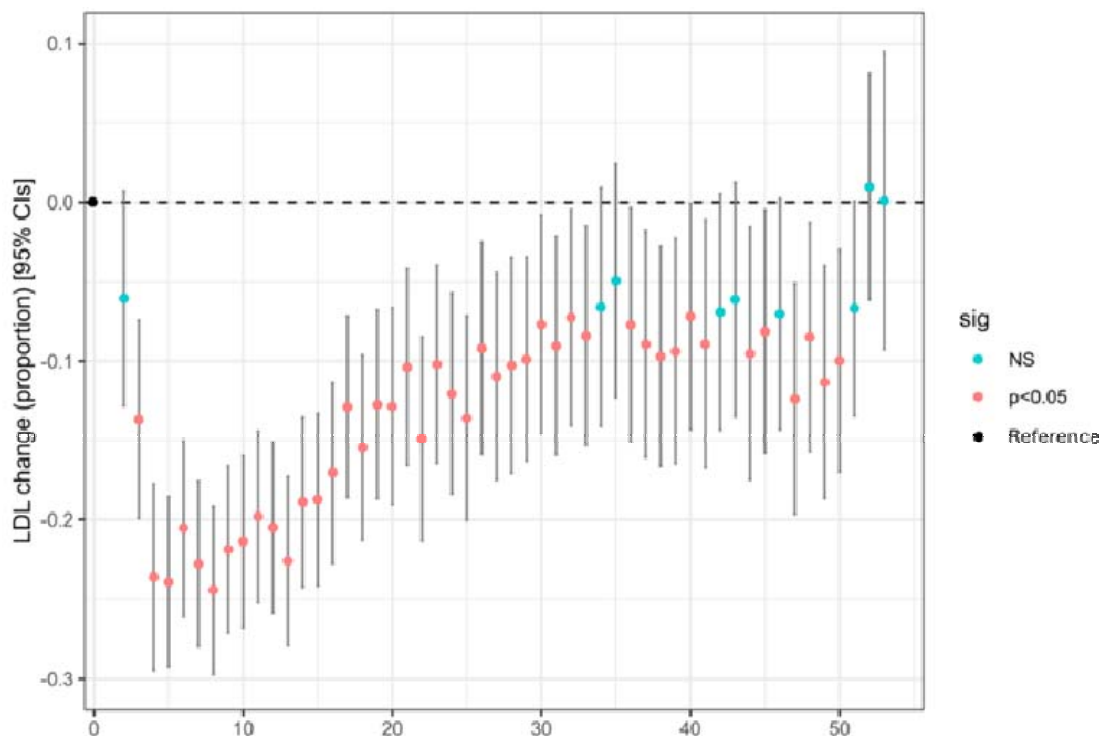
18

19

20 Results

21

1 **Figure 1 – Associations between time to first GP LDL measure following statin initiation**
2 **and LDL change**

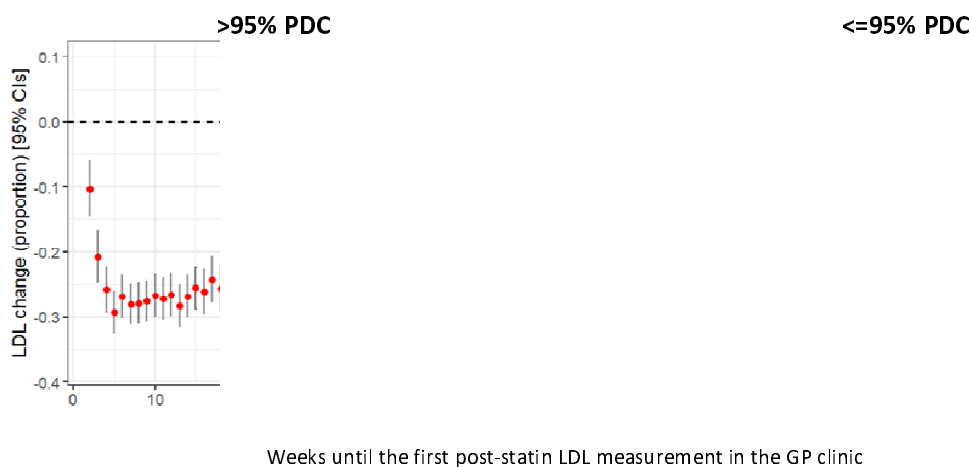


5 This plot illustrates associations between LDL reduction and visit days, based on available LDL-C data
6 for each week, not individual-level data. The maximum reduction occurs between weeks 5-8 of 26%
7 (95%CI 23-29).
8

9
10
11
12
13
14
15
16
17
18

16 **Figure 2 - Associations between time to first GP LDL measure following statin initiation and**
17 **LDL change by PDC groups**

1



2

3

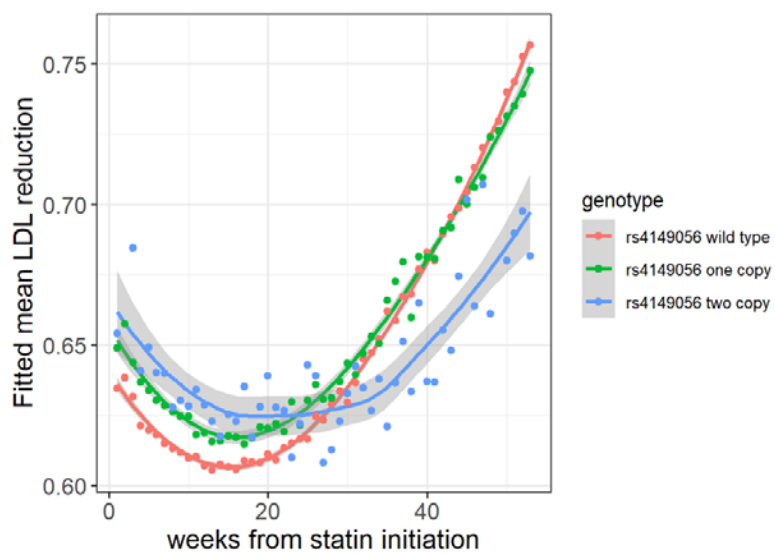
4 This plot illustrates the associations between LDL reduction and GP visit days, based on available LDL-
5 C measurement post stain initiation for each week in two PDC groups.

6

7

8 **Figure 3 - Quadratic model of LDL reduction estimated by SLCO1B1 rs4149046 genotype**

9



10

11

12 The quadratic model revealed that those with the rs4149046 CC allele had the smallest LDL-
13 c reduction in the in year 1.