Online supplementary file 2: CPRD technical report

1. LDL-C missingness approach

To assess how LDL-C levels varied across time, 2 mutually exclusive time periods of analysis were defined, selecting for each of these periods a corresponding lipid measurement.

a) Period 1 (P1 or baseline) was defined with the closest LDL-C, total cholesterol (TC), triglycerides (TG) and HDL-C measurements prior to FH diagnosis for the 'newly treated' and 'untreated' subgroups, or prior to first prescription date for the 'treatment retainers' subgroup.
b) Period 2 (P2) was defined as the period between 3 months and 2 years from FH diagnosis for the 'newly treated' and 'untreated' subgroups (or first prescription for the 'treatment retainers' subgroup), with the closest to 2 years LDL-C, TC, TG and HDL-C measurements being selected.

The presence of lipid measurements with LDL-C missing was tackled through the use of the Friedewald equation (1). This was possible given the availability of TC, TG and HDL-C measurements, which enabled the use of this calculated method to obtain missing LDL-C measurements. The Friedewald equation was initially applied to measurements collected on the same date (that is, on day *A* and for FH patient *i*, LDL-C_{Ai} was missing and was derived through TC_{Ai}, TG_{Ai} and HDL-C_{Ai} measured on that same day, *A*). Subsequently it was applied according to time periods (that is, and for example, if no LDL-C measurement was available for period *2* for FH patient *i*, this was derived through TC_{2i}, TG_{2i} and HDL-C_{2i} selected for period *2* for patient *i*, if existing). A high level of missingness for LDL-C measurements across the 2 periods was still observed after applying the Friedewald equation.

To tackle this issue, and particularly for missing baseline LDL-C measurements, an 'alternative' Friedewald equation was developed that considers: a) the pre-treatment TC; b) the lowest post-treatment HDL-C if pre-treatment HDL-C was not available (under the assumption that FH patients have normal levels of HDL-C and that the effect of LLTs on HDL-C is small (2)); and c) the highest post-treatment TG if pre-treatment TG was not available (assuming that FH patients generally have normal levels of TG (3, 4) and that the effect of LLTs on TG is modest (5, 6)). Thus, the bespoke "alternative" Friedewald equation considers: [pre-treatment LDL-C] = [pre-treatment TC] - [post-treatment-lowest-HDL] - ([post-treatment-highest-TG] / 2.19), for [post-treatment-highest-TG]>4.52 as per Friedewald equation requirements.

A multivariate multiple imputation approach was considered in order to generate LDL-C values (7). This approach allowed accounting for the uncertainty around the true LDL-C values, and obtain approximately unbiased estimates. The key commonly applied assumption for the application of this imputation approach is that the missingness is not completely random, but that the propensity of missingness depends on the observed data, not the missing data itself – the so called Missing at Random (MAR) assumption.

Following recent published recommendations on multiple imputation, the imputation process for the LDL-C response analysis was carried out separately by patient' subgroups, that is, separately for the 'untreated, the 'newly treated' and the 'treatment retainers' subgroups (8). For the risk modelling analysis, imputation of LDL-C was only applied to the treated cohort (i.e. 'newly treated' and the 'treatment retainers' subgroups combined). The number of imputed datasets was estimated using a two-step approach, as recommended in recent guidance (9), and set to 34 imputed datasets.

The covariate set considered for the multiple imputation modelling was selected according to their use in key publications relating to: i) prediction of statin use (10); ii) predictors of statin adherence (11, 12); and iii) predictors of LDL-C response to LLTs. The imputation model considered: age at baseline (years), gender, history of CVD, post-baseline CVD, pre-treatment LDL-C, ethnicity, deprivation index, body mass index, smoking, diabetes, hypertension, systolic blood pressure, other comorbidities (includes: inflammatory disease, HIV and chronic kidney disease), atrial fibrillation, family history of CHD, anti-hypertensive medication, polypharmacy (patients getting more than one drug, considering: antipsychotics, corticosteroid, antihypertensives and immunosuppressants), post-baseline statin potency (categorised into low, medium and high potency (13)), QRisk2 score and lipids measurements (in mmol/l). We used Rubin's rules (Rubin, 2004) to obtain combined estimates from imputed datasets.

2. Validity of fitted survival models

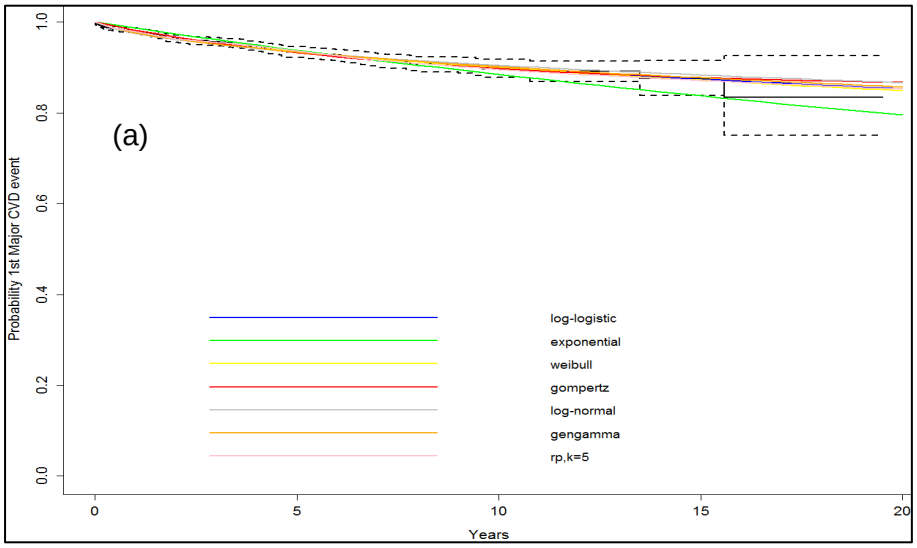
i. Internal validity

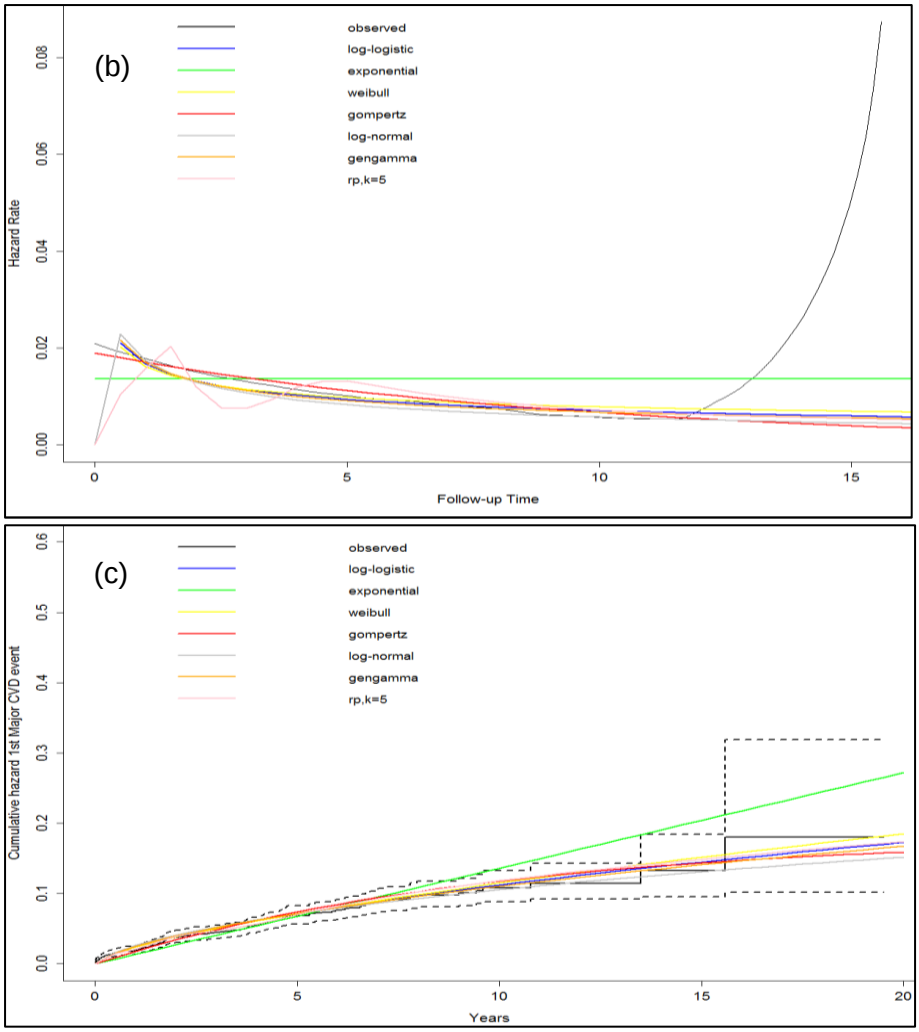
Figure 3.1 shows the survival, hazard and cumulative hazard curves for each fitted survival curve. Table 3.1 shows the AIC statistic for all fitted parametric survival models.

Table 3.1: AIC statistic for all fitted parametric survival models.

Parametric survival model	AIC statistic
Exponential	1413.3
Weibull	1387.0
Gompertz	1405.0
Log-normal	1387.8
Log-logistic	1384.9
Generalised Gamma	1387.4
Royston Parmar spline with 5 knots	1380.9

Figure 3.1: (a) Survival curve; (b) Hazard rates; and (c) Cumulative hazard for each fitted survival model.



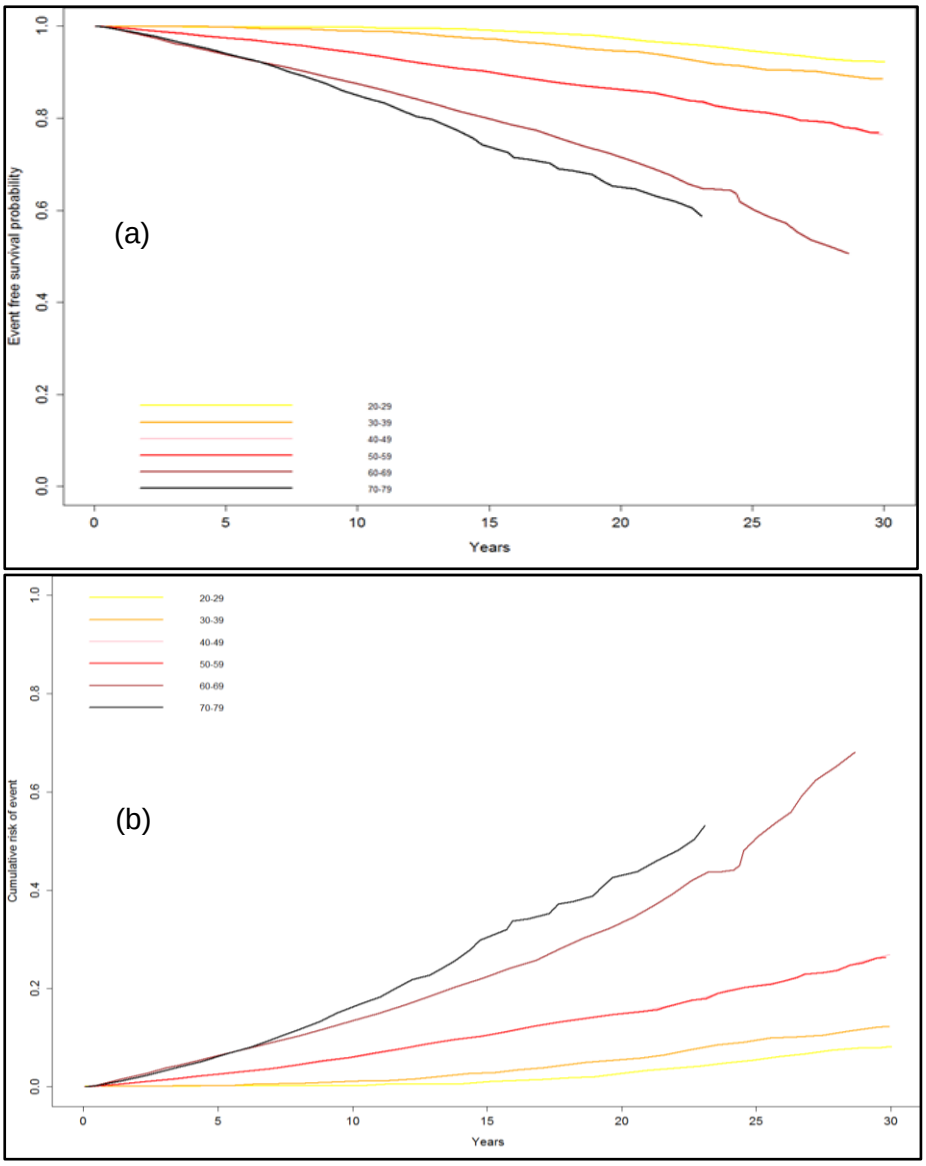


ii. External validity

The external validity of the fitted parametric models was assessed through the use of external data from a recent publication by Perak et al (14). The authors looked at the long-term risk of atherosclerotic cardiovascular disease in US adults with elevated LDL-C (≥190mg/dL). The paper was selected as it provided a longer period of follow-up than the data available from CPRD and represents a large cohort. This publication provided adjusted[1] 30-year survival free from CHD death or non-fatal MI for different age groups of the cohort. Please see Figure 3.2 for the digitised Perak study data. Across age groups event hazards appear to be increasing over time.

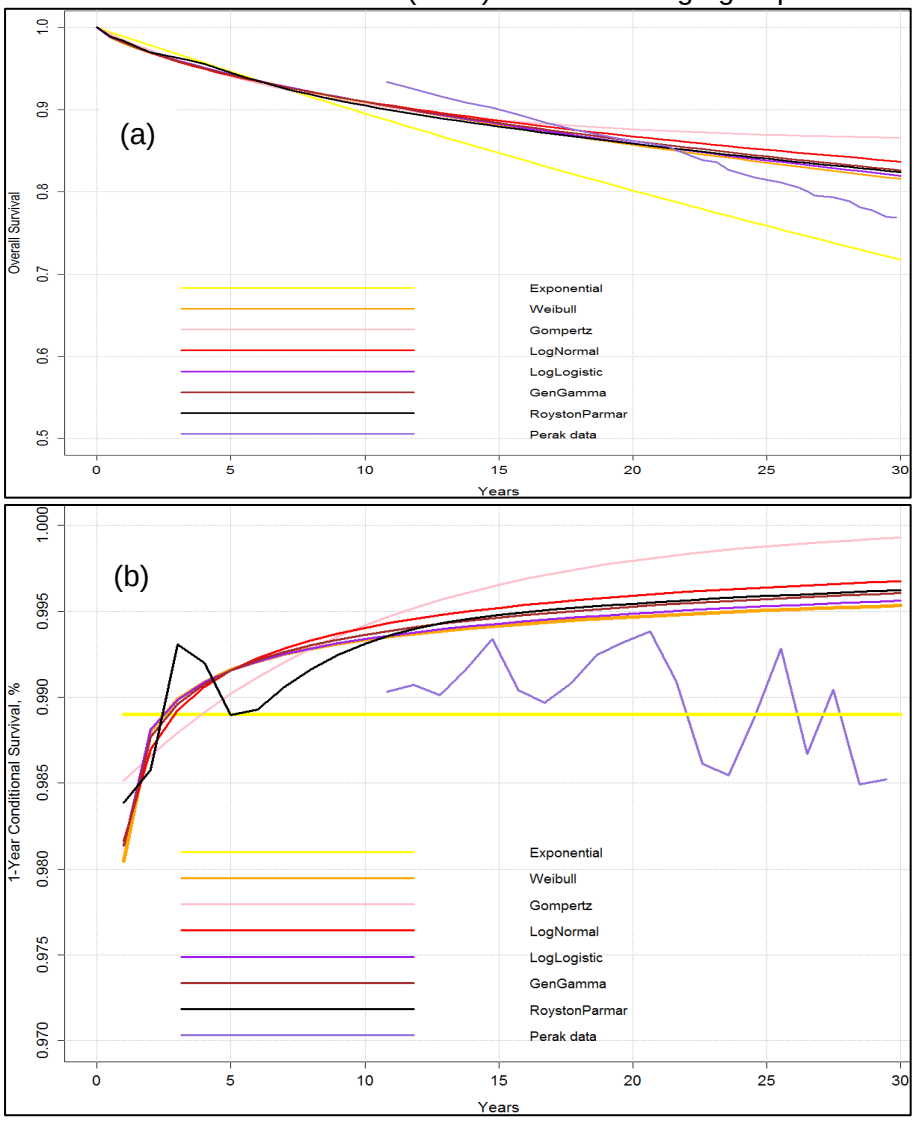
Figure 3.2: (a) Perak study 30-year survival free from CHD death or non-fatal MI for different age groups; and (b) Perak study 30-year cumulative hazard for survival free from CHD death or non-fatal MI for different age groups.

[1] Adjusted for age, sex, race, body mass index, diabetes mellitus, smoking, systolic blood pressure, antihypertensive therapy, HDL-C, cholesterol treatment, and cohort.



Survival predictions for the covariate values in Perak were generated for each model. For comparability reasons, these were then adjusted to Perak et al (14). The adjustment involved removing the impact of TIA and stroke first events from the predicted CVD risks obtained from applying different survival models to the CPRD data. Also, most patients in the Perak et al (14) cohort were untreated at baseline, and treatment rates were assumed low during the (up to) 30 year's follow-up. Thus, the treatment effect from the predicted CVD risks in CPRD was removed by considering an estimated LDL-C reduction from response to LLT of 32.6% (see Table 3.3) and a rate ratio of 0.79 reduction in CVD per 1 mmol/L reduction in LDL-C (15). These adjustments produced comparable parametric predictions for the different survival models and for each age-related patient profile in Perak et al (14). Figure 3.3 shows the survival and 1-year conditional survival curves for the 50-59 FH patient profile, which matches the average FH patient age in the CPRD cohort. Overall the Perak data suggested a higher event rate from year 10 onwards than all models fitted to CPRD with the exception of the exponential model. Therefore, although the exponential model did not provide a good fit to the observed data, it was selected for use within a sensitivity analysis.

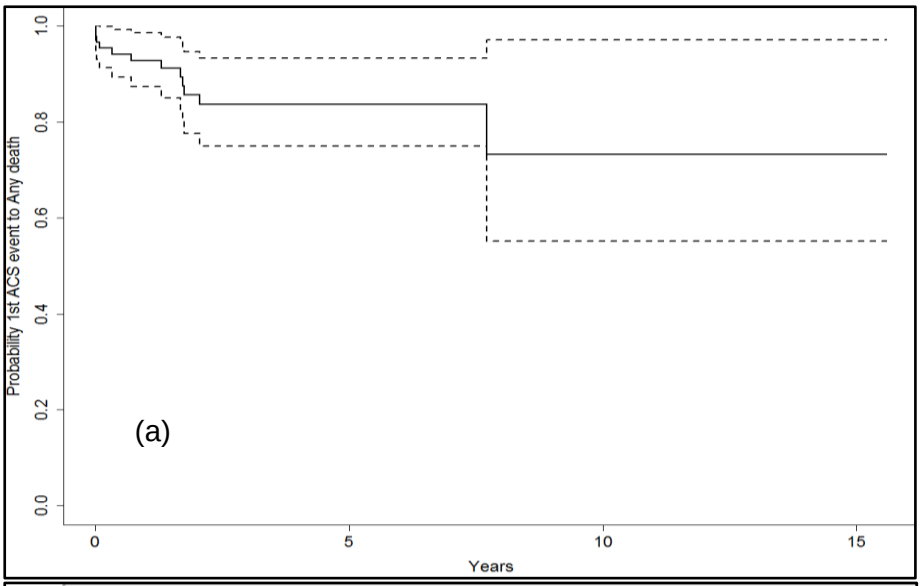
Figure 3.3: (a) Predicted CVD risk in the 50-59 age group for each fitted survival model, overlaid by Perak et al (2016) Kaplan-Meier data for this age group; and (b) 1-year conditional survival in the 50-59 age group for each fitted survival model, overlaid by 1-year conditional survival Perak et al (2016) data for this age group.

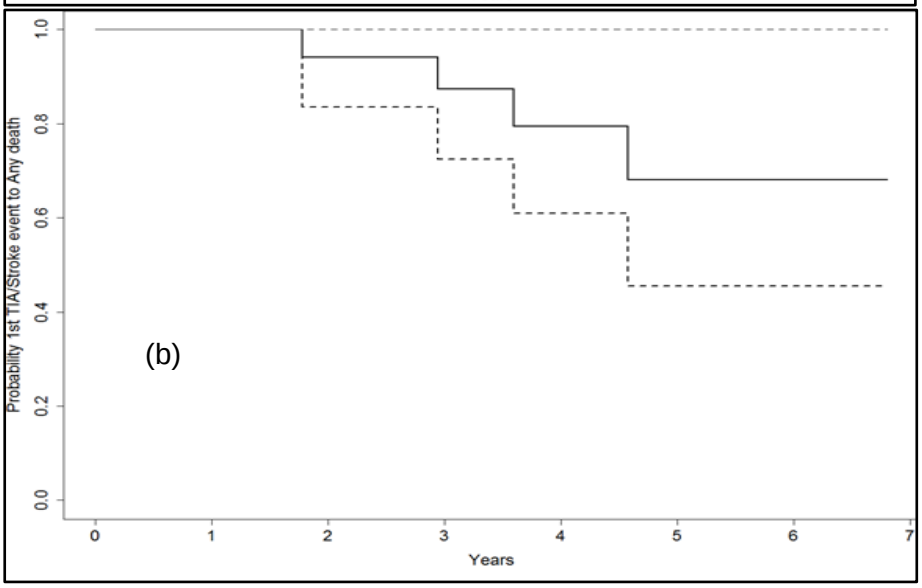


3. Mortality following non-fatal CVD events

Figure 3.4 (a) and (b) shows the Kaplan-Meier survival estimates for death for any reason from relevant non-fatal CVD events, respectively.

Figure 3.4: (a) Kaplan-Meier of time to any death following ACS; and (b) Kaplan-Meier of time to any death following TIA/Stroke.


Probability 1st ACS event to Any death
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
Years
(a)


Probability 1st TIA/Stroke event to Any death
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
Years
(b)

References

1. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clinical chemistry: Oxford Academic; 1972. p. 499-502.
2. McTaggart F, Jones P. Effects of statins on high-density lipoproteins: a potential contribution to cardiovascular benefit. Cardiovasc Drugs Ther. 2008;22(4):321-38.
3. Futema M, Kumari M, Boustred C, Kivimaki M, Humphries SE. Would raising the total cholesterol diagnostic cut-off from 7.5 mmol/L to 9.3 mmol/L improve detection rate of patients with monogenic familial hypercholesterolaemia? Atherosclerosis. 2015;239(2):295-8.
4. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk: The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS). European Heart Journal. 2019;41(1):111-88.
5. NICE. Clinical guideline 181. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. 2014.
6. NICE. Clinical guideline 71: identification and management of familial hypercholesterolaemia. London, Manchester: NICE; 2019.
7. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011. 2011;45(3):67.
8. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res. 2018;27(9):2610-26.
9. von Hippel PT. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule. Sociological Methods & Research. 2020;49(3):699-718.
10. Wu J, Zhu S, Yao GL, Mohammed MA, Marshall T. Patient factors influencing the prescribing of lipid lowering drugs for primary prevention of cardiovascular disease in UK general practice: a national retrospective cohort study. PLoS One. 2013;8(7):e67611.
11. Chowdhury R, Khan H, Heydon E, Shroufi A, Fahimi S, Moore C, et al. Adherence to cardiovascular therapy: a meta-analysis of prevalence and clinical consequences. Eur Heart J. 2013;34(38):2940-8.
12. Mann DM, Woodward M, Muntner P, Falzon L, Kronish I. Predictors of nonadherence to statins: a systematic review and meta-analysis. Ann Pharmacother. 2010;44(9):1410-21.
13. NICE. Cardiovascular disease: risk assessment and reduction, including lipid modification. 2014.
14. Perak AM, Ning H, de Ferranti SD, Gooding HC, Wilkins JT, Lloyd-Jones DM. Long-Term Risk of Atherosclerotic Cardiovascular Disease in US Adults With the Familial Hypercholesterolemia Phenotype. Circulation. 2016;134(1):9-19.
15. Fulcher J, O'Connell R, Voysey M, Emberson J, Blackwell L, Mihaylova B, et al. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials. Lancet. 2015;385(9976):1397-405.