

Appendix A. Serum Cholesterol Dose-Response Function

This appendix describes EPA's literature review to identify studies to estimate relationships between cholesterol levels and serum per- and polyfluoroalkyl substances (PFAS) for inclusion in a meta-analysis of these relationships. Statistical analyses that combine the results of multiple studies, such as meta-analyses, are widely applied to investigate the dose-specific relationship between contaminant levels and associated health effects. Such analyses are suitable for economic assessments because they can improve precision and statistical power (Engels et al., 2000; Deeks, 2002; Rücker et al., 2009). This appendix also provides details on the meta-data development, results of the meta-analysis, and limitations and uncertainties associated with the estimated relationships. EPA used the estimated relationships to estimate cardiovascular disease (CVD) risk reduction associated with exposure to PFAS mediated by changes in serum cholesterol markers.

A.1 Data Sources

EPA relied on two literature review efforts to identify potential sources of exposure-response information for the effect of PFAS on serum cholesterol, lipids, and lipoproteins:

1. A literature review built on the one conducted by the Agency for Toxic Substances and Disease Registry (ATSDR) in the development of their Toxicological Review Public Comment Draft (ATSDR, 2018), which included literature through mid-2017.
2. The most recent systematic review of the newly published epidemiological literature for PFAS performed by EPA's Office of Science and Technology (EPA/OST), which included literature from 2016 to 2020 (U.S. EPA, 2021a; U.S. EPA, 2021b).

The relationships between exposure to PFAS and serum total cholesterol (TC) and high-density lipoprotein cholesterol (HDL) identified based on these literature reviews allowed EPA to generate inputs for the Pooled Cohort Atherosclerotic Cardiovascular Disease (ASCVD) risk model (Goff et al., 2014).¹

A.1.1 Literature Review and Studies Identification for the Meta-Analysis

Two reviewers independently screened references retrieved from the literature search by title and abstract, and then reviewed relevant studies in full text. EPA evaluated studies identified during the search according to the following criteria prior to inclusion in the meta-analysis to ensure validity, consistency, and applicability. Briefly, of interest were studies in general population adults, evaluating the outcomes of TC and HDL, and the exposures of perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS). Because EPA evaluates CVD risk among a general population of adults aged 40 to 89, studies performed on specific population subsets, such as occupational populations, were not considered for inclusion in the meta-analysis due to the potential for greater levels of exposure to PFOA and PFOS in these populations compared to the general population.

¹ The ASCVD model relies on the following inputs: demographic information, smoking and diabetes status, serum TC, and HDL.

Applicability: Each study was evaluated to determine whether it estimated the association between exposure to PFOA or PFOS (measured in serum or plasma) and a quantitative measure of TC or HDLC in general populations (age 20 and older). Of the 39 studies identified as part of the ATSDR-based literature review that provided information on the relationship between exposure to PFAS and TC and HDLC levels, 9 were general population studies. Of the 41 studies identified as part of the EPA/OST literature review that provided information on the relationship between exposure to PFAS and TC and HDLC levels, 14 were general population studies. These studies² were further evaluated for inclusion in the meta-analysis.

Research methods and study details: Each study was evaluated to determine whether it reported numbers of participants, quantitative effect estimates (beta coefficients), measures of effect estimate variance (95% confidence intervals [CIs], and standard errors [SEs] or standard deviations [SDs]). EPA retained studies with missing measures of effect estimate variance but with reported p-values for differences. For such studies, EPA used the approach in the Cochrane Handbook for Systematic Reviews (J. P. Higgins et al., 2019) to calculate SDs or SEs. Briefly, the approach estimates the SEs using the correspondence between the p-value and the t-statistic, with degrees of freedom equal to the difference between the sample size and the number of parameters in the model that provided the effect estimate. Then the SE is obtained by dividing the effect estimate by the t-statistic.

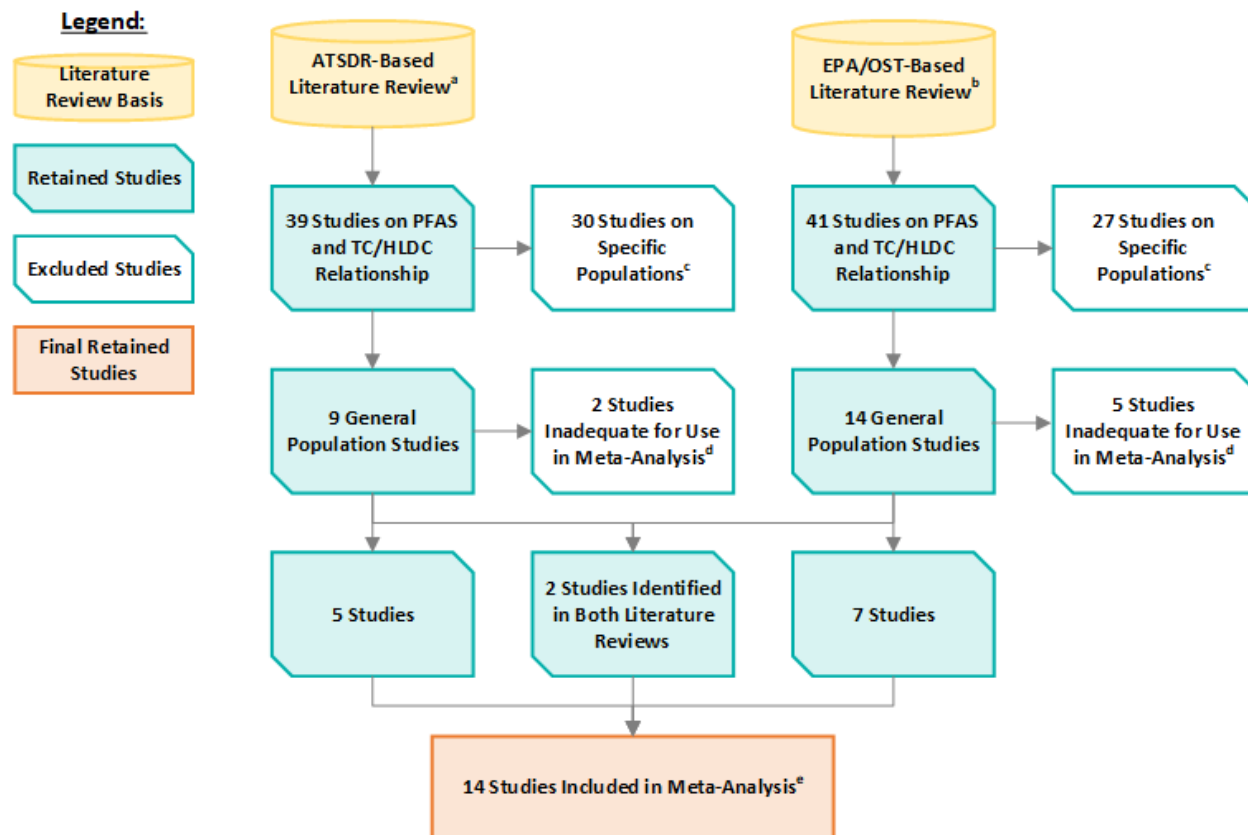
Additional exclusion criteria: EPA also excluded studies that reported data only for pregnant women, infants, or children. Although there is some evidence that PFAS exposure is associated with cardiometabolic impairment in children and younger adults (Rappazzo et al., 2017), EPA did not extract data from these studies because lipid levels are known to change during pregnancy from pre-pregnancy levels, and the relationships between lipid profiles at early life stages are not as well defined as they are at later life stages. Another frequent reason for study exclusion was the reporting of only relative risks or odds ratios for hypercholesterolemia or hyperlipidemia; results in this form could not be used to estimate continuous exposure-response relationships.

A.1.2 Assessment of Study Applicability to the Meta-Analysis

Figure A-1 presents a flow diagram of the studies reviewed as part of the ATSDR-based and EPA/OST-based literature reviews and the selection of studies retained for inclusion in the meta-analysis. Using the study inclusion criteria described in section A.1.1, EPA retained 14 studies for use in the meta-analysis. Of these, five were identified as part of the ATSDR literature review (Château-Degat et al., 2010; Fisher et al., 2013; Fu et al., 2014; Nelson et al., 2010; Steenland et al., 2009), seven were identified from the EPA/OST systematic review (Dong et al., 2019; Fan et al., 2020; Jain et al., 2019; Y. Li et al., 2020; C. Y. Lin et al., 2020; P.-I. D. Lin et al., 2019; Yang et al., 2018), and two were identified in both literature reviews (He et al., 2018; Liu et al., 2018).

² Of the general population studies identified as part of the EPA/OST literature review, five overlapped with studies identified as part of the ATSDR-based literature review.

Figure A-1. Diagram of Literature Retained for Use in the Meta-Analysis and Data Sources.



Notes:

ATSDR = Agency for Toxic Substances and Disease Registry, EPA = Environmental Protection Agency, OST = Office of Science and Technology, PFAS = per- and polyfluoroalkyl substances, TC = Total Cholesterol, HDLC = high-density lipoprotein in cholesterol

^aIncluded literature through mid-2017.

^bIncluded literature published from 2016 to 2020.

^cFor example, studies based on occupational data or data only for pregnant women, infants, or children.

^dSome studies did not include the estimates required for meta-analysis calculations. For example, certain studies did not report effect estimates or interquartile ranges.

^eOf these studies, 8 are based on data from the United States and 6 are based on data outside of the United States.

Table A-1 summarizes the 14 studies that were identified in the ATSDR-based and EPA/OST literature review that EPA used to derive slope estimates for PFOA and PFOS associations with serum TC and HDLC levels.³ Six of the studies that EPA retained for use in the meta-analysis were based on PFAS and serum lipid measurements from the U.S. general population (National Health and Nutrition Examination Survey [NHANES]) (Dong et al., 2019; Fan et al., 2020; He et al., 2018; Jain et al., 2019; Liu et al., 2018; Nelson et al., 2010); there were also general population studies from Canada (Fisher et al., 2013), Sweden (Y. Li et al., 2020), Taiwan (Yang et al., 2018; C. Y. Lin et al., 2020), and Henan Province, China (Fu et al., 2014). Château-Degat et al. (2010) reported on the relationship between PFOS and serum lipids in a Canadian Inuit

³ For this effort, EPA focused on PFOA and PFOS, since these are by far the most well-studied perfluorinated compounds.

population. EPA also retained the results from a study of a highly exposed population in the United States (the C8 cohort) (Steenland et al., 2009) and from a study using participants in a U.S. diabetes prevention program (P.-I. D. Lin et al., 2019).

EPA excluded two general population studies identified in the ATSDR-based literature review (Eriksen et al., 2013; Seo et al., 2018) and two general population studies identified based on the EPA/OST systematic review (Convertino et al., 2018; Huang et al., 2018) that were inadequate for use in the meta-analysis because they did not include the estimates required for meta-analysis calculations. For example, EPA excluded the studies identified in the ATSDR literature review from the meta-analysis because the authors did not report either the effect estimates (Seo et al., 2018) or interquartile ranges (Eriksen et al., 2013) needed for calculations.⁴ Similarly, EPA excluded the studies identified as part of the EPA/OST systematic review because they involved a Phase 1 controlled trial with modeled exposures in cancer patients dosed with ammonium perfluorooctanoate (Convertino et al., 2018) or reported effect estimates (Spearman correlation coefficients) that were not suitable for use in the meta-analysis (Huang et al., 2018).

Table A-1: Studies Selected for Inclusion in the Meta-Analyses

Author and Year	Title	Cholesterol and PFAS Relationship Evaluated in Study			
		TC		HDL	
		PFOA	PFOS	PFOA	PFOS
Steenland et al., 2009 ^{a,d}	Association of Perfluorooctanoic Acid and Perfluorooctane Sulfonate With Serum Lipids Among Adults Living Near a Chemical Plant	X	X	X	X
Château-Degat et al., 2010 ^{a,d}	Effects of Perfluorooctanesulfonate Exposure on Plasma Lipid Levels in the Inuit Population of Nunavik (Northern Quebec)		X		X
Nelson et al., 2010 ^{a,d}	Exposure to Polyfluoroalkyl Chemicals and Cholesterol, Body Weight, and Insulin Resistance in the General U.S. Population	X	X	X	X
Fisher et al., 2013 ^{a,d}	Do Perfluoroalkyl Substances Affect Metabolic Function and Plasma Lipids?—Analysis of the 2007–2009, Canadian Health Measures Survey (CHMS) Cycle 1	X	X	X	X
Fu et al., 2014 ^{a,d}	Associations Between Serum Concentrations of Perfluoroalkyl Acids and Serum Lipid Levels in a Chinese Population	X	X	X	X
He et al., 2018 ^c	PFOA is Associated with Diabetes and Metabolic Alteration in US Men: National Health and Nutrition Examination Survey 2003-2012	X	X	X	X

⁴ Efforts to contact the study authors for the missing data were unsuccessful at the time of this report.

Table A-1: Studies Selected for Inclusion in the Meta-Analyses

Author and Year	Title	Cholesterol and PFAS Relationship Evaluated in Study			
		TC		HDL	
		PFOA	PFOS	PFOA	PFOS
Liu et al., 2018 ^c	Association Among Total Serum Isomers of Perfluorinated Chemicals, Glucose Homeostasis, Lipid Profiles, Serum Protein and Metabolic Syndrome in Adults: NHANES, 2013–2014	X	X	X	X
Yang et al., 2018 ^b	Association of Serum Levels of Perfluoroalkyl Substances (PFASs) With the Metabolic Syndrome (MetS) in Chinese Male Adults: A Cross-Sectional Study			X	X
Dong et al., 2019 ^b	Using 2003–2014 U.S. NHANES Data to Determine the Associations Between Per- and Polyfluoroalkyl Substances and Cholesterol: Trend and Implications	X	X	X	
Jain et al., 2019 ^b	Roles of Gender and Obesity in Defining Correlations Between Perfluoroalkyl Substances and Lipid/Lipoproteins	X	X	X	X
P.-I. D. Lin et al., 2019 ^b	Per- and Polyfluoroalkyl Substances and Blood Lipid Levels in Pre-Diabetic Adults—Longitudinal Analysis of the Diabetes Prevention Program Outcomes Study	X	X	X	X
Fan et al., 2020 ^b	Serum Albumin Mediates the Effect of Multiple Per- and Polyfluoroalkyl Substances on Serum Lipid Levels	X	X	X	X
Y. Li et al., 2020 ^b	Associations Between Perfluoroalkyl Substances and Serum Lipids in a Swedish Adult Population With Contaminated Drinking Water	X	X	X	X
C. Y. Lin et al., 2020 ^b	The Association Between Total Serum Isomers of Per- and Polyfluoroalkyl Substances, Lipid Profiles, and the DNA Oxidative/Nitrative Stress Biomarkers in Middle-Aged Taiwanese Adults			X	X

Notes:

^a Studies identified based on ATSDR literature review.

^b Studies identified based on EPA/OST literature review.

^c Studies available in both assessments.

^d Studies available in PFOA and/or PFOS health effects support documents (U.S. EPA, 2016a, 2016b).

A.2 Meta-Analysis

Based on the study inclusion criteria discussed in section A.1.1, EPA included 14 studies in the meta-analysis. Of these 14 studies, 11 were used to develop exposure-response relationships for serum PFOA and TC, 13 were used to develop exposure-response relationships for serum PFOA and HDL, 12 studies were used to develop exposure-response relationships for serum PFOS and TC, and 13 studies were used to develop exposure-response relationships for serum

PFOA and HDLC (Table A-1). EPA conducted four separate meta-analyses: one analysis for each combination of chemical (PFOA or PFOS) and health outcome (TC or HDLC).

A.2.1 Extraction of Slope Values for TC and HDLC

If studies reported linear slope relationships (change in serum TC or HDLC in mg/dL per ng/mL change in serum PFOA/PFOS), EPA extracted these values, along with their confidence limits, directly as reported by the study authors. If results from multiple models with different adjustments for confounders were reported within a single study, either the most adjusted results or the main model results as presented by the study authors were selected. When studies provided results for both untransformed and log-transformed PFOA/PFOS, EPA used untransformed PFOA/PFOS to reduce bias due to back-transformations of effect estimates. For studies that provided results only for log-transformed PFOA/PFOS (five studies) or log-transformed outcomes (two studies), or log-transformed both PFOA/PFOS and outcomes (two studies), EPA approximated the results for an untransformed analysis using the approach outlined by Rodríguez-Barranco et al. (2017) and Dzierlenga et al. (2020). When not reported, EPA assumed that the natural logarithm was the basis of the transformation. An independent EPA reviewer evaluated the extracted slope values for quality assurance.

A.2.2 Methods and Key Assumptions

The summary measure of association was a beta coefficient relating changes in TC or HDLC in mg/dL to increases in serum or plasma⁵ PFOA or PFOS in ng/mL. EPA conducted random-effects meta-analyses using the DerSimonian et al. (1986) approach, which uses weights based on the inverse of the variance of the coefficient of each study plus the addition of an extra component of variance between studies. When studies reported beta coefficients by quartiles (e.g., He et al., 2018), EPA estimated a linear coefficient using a weighted linear regression of the midpoints of the quartiles and the reported beta coefficients, using the inverse of standard errors as the regression weights.

EPA assessed between-study heterogeneity using Cochran's Q test (Cochran, 1954) and the I^2 statistic (J. P. Higgins et al., 2003). EPA developed forest plots to display the results. EPA developed funnel plots and performed an Egger regression on the estimates of effect size to assess potential publication bias (Begg et al., 1994; M. Egger et al., 2008; Matthias Egger et al., 1997). Because back-transformations of effect estimates with log-transformed outcomes or exposures could introduce bias and could be a source of heterogeneity, EPA also conducted sub-analyses by type of model that provided the study-specific effect estimate (e.g., only including studies that reported linear associations [six studies] or linear-log associations [five studies]).

If publication bias was observed, EPA conducted sensitivity analyses using trim-and-fill methods (Duval et al., 2000a, 2000b) to estimate the number of missing studies and predict the impact of the hypothetical "missing" studies on the pooled effect estimate. To investigate sources of heterogeneity, EPA conducted several sensitivity analyses:

⁵ PFOA or PFOS concentrations in serum or plasma were treated interchangeably.

- EPA first evaluated the impact of using other estimation methods for the between-study variance (τ^2) besides the DerSimonian et al. (1986) approach, such as restricted maximum likelihood (Raudenbush, 2009) or Sidik et al. (2005).
- To assess potential impact of a single study on the overall effect estimate, EPA conducted leave-one-out meta-analyses.
- To assess the impact of using multiple beta coefficients from the same study (which are correlated), EPA excluded a study that contributed four effect estimates (gender- and obesity-specific) for each analysis, which also accounted for most of the weight in the overall pooled beta coefficient (Jain et al., 2019).
- EPA also assessed the impact of non U.S or Canadian general population studies in sensitivity analyses excluding studies conducted in China (Fu et al., 2014), Taiwan (Yang et al., 2018; C. Y. Lin et al., 2020), or Sweden (Y. Li et al., 2020), the Canadian Inuit population study (Château-Degat et al., 2010), and the U.S. high-exposure community study (Steenland et al., 2009).

Six studies that EPA retained for use in the meta-analysis were based on PFAS and serum lipid measurements using data from overlapping NHANES cycles: Dong et al. (2019) used data from 2003–2014, while He et al. (2018) used 2003–2012 data; Jain et al. (2019) used 2005–2014 data; Fan et al. (2020) used 2011–2014 data; Liu et al. (2018) used 2013–2014; and Nelson et al. (2010) used data from 2003–2004. Although the datasets and models were not exactly the same in all NHANES-based studies, to avoid estimate dependency issues due to overlapping populations in the meta-analysis, EPA also performed a sensitivity analysis including only Dong et al. (2019) for TC, He et al. (2018) for HDLC, and Liu et al. (2018) for PFOS and HDLC.

EPA performed statistical analyses using the software STATA, version 16.1 (StataCorp, 2019), with the *combine*, *meta esize*, *meta set*, *meta summarize*, *metainf*, *meta funnel*, *meta bias*, and *meta trimfill* packages (Palmer et al., 2016). Results of the meta-analyses are presented in Table A-2 and Table A-3. Overall, there is a high degree of heterogeneity when all studies are combined. Excluding Jain et al. (2019) did not significantly reduce the heterogeneity; however restricting analyses to studies reporting linear or linear-log associations did reduce heterogeneity in most cases.

A.2.3 Slope Estimation for PFOA

When including the six studies reporting linear associations, there was a statistically significant positive increase in TC of 1.57 (95% CI: 0.02, 3.13) mg/dL per ng/mL serum PFOA (p -value=0.048, I^2 =87%). The association for HDLC and PFOA was positive (0.15; 95% CI: -0.18, 0.48) but not statistically significant (Table A-2, Figure A-2). Adjusting for possible publication bias through funnel plots and trim-and-fill analysis suggested the imputation of two additional studies for HDLC and PFOA with a smaller effect (0.008, 95% CI: -0.38, 0.40). For TC and PFOA, the pooled associations did not change when adjusting for possible publication bias (Figure A-3). However, methods to assess heterogeneity and publication bias have limitations in small sample-size meta-analyses, thus these results should be interpreted cautiously (von

Hippel, 2015). Similar results for both TC- and HDLC-PFOA relationships were observed when the analysis excluded three overlapping NHANES studies (Table A-2).

The overall pooled associations (including both linear and nonlinear studies) for TC and PFOA and for HDLC and PFOA were positive but did not reach statistical significance at the 0.05 level when all study types were included (Table A-2, Figure A-4). There was evidence of significant heterogeneity (p-values < 0.001, I^2 > 70%) and publication bias. Adjusting for possible publication bias through funnel plots and trim-and-fill analysis suggested the imputation of six additional studies for TC and four for HDLC; however, the pooled estimates did not change significantly (Figure A-5). EPA observed similar results in leave-one-out analyses, sensitivity analyses restricted to U.S or Canadian general population studies, and analyses excluding Jain et al. (2019) estimates.

When EPA restricted analyses to only studies reporting linear-log associations, the results showed reduced heterogeneity and positive but non-significant associations for both outcomes and exposures (Table A-2). Given the large degree of heterogeneity in the pooled associations when all data were included and the likelihood of bias that back-transformation of effect estimates with log-transformed outcomes or exposures could introduce, EPA relied on the analyses restricted to studies reporting similar models, favoring the pooled slope (from the six studies reporting linear associations) of 1.57 mg/dL per ng/mL serum PFOA for interpretability and use in the CVD risk reduction analysis.⁶ Because the association for PFOA and HDLC was not statistically significant and the available evidence does not support an association, EPA does not use an effect estimate for the PFOA-HDLC relationship in the CVD risk reduction analysis.

⁶ EPA will characterize uncertainty surrounding this estimate as described in Appendix E.

Table A-2: Results for PFOA Meta-Analyses

Group	Outcome	Number of Studies/ Number of Estimates	Beta (mg/dL per ng/mL)	95% CIs		p-value	Q ^a	p-value for Q	I ²	Tau ²
All Studies	TC	11/14	0.003	-0.001	0.006	0.177	123.68	< 0.001	89.49	0
	HDLC	13/18	0.001	-0.001	0.004	0.307	61.02	< 0.001	72.14	0
Linear Models Only	TC	4	1.574	0.018	3.130	0.048	23.43	< 0.001	87.19	1.910
	HDLC	6	0.150	-0.183	0.482	0.378	20.08	0.001	75.10	0.091
Linear-Log Models Only ^b	TC	3/6	0.006	-0.014	0.025	0.582	31.57	< 0.001	84.16	0.0004
	HDLC	5/9	0.005	-0.009	0.019	0.488	13.52	0.095	40.84	0.0001
Exclude Jain et al. (2019)	TC	10	0.004	-0.002	0.01	0.179	82.04	< 0.001	89.03	0
	HDLC	12/14	0.001	-0.004	0.006	0.516	56.46	< 0.001	76.97	0
All Studies ^c	TC	6	0.017	-0.033	0.067	0.505	21.56	0.001	76.9	0.001
	HDLC	9/10	0.001	-0.002	0.005	0.418	32.61	< 0.001	72.4	0
Linear Models Only – No Overlap ^c	TC	1 ^d	1.480	0.180	2.780	0.026	0.00	NA	NA	NA
	HDLC	3	0.308	-0.305	0.921	0.325	5.32	0.070	62.40	0.165

Notes:

^a Q statistics for heterogeneity. Tau² is the between-studies variance.

^b Units for the pooled beta coefficient are mg/dL per ln(ng/mL).

^c No overlapping NHANES data.

^d Data from Dong et al. (2019). Statistics for heterogeneity do not apply when only one study is used.

Figure A-2. Forest Plots Showing the Beta Coefficients Relating PFOA Concentrations to TC and HDLC in Each Study Reporting Linear Associations, and Pooled Estimates After Random-Effects Meta-Analysis.

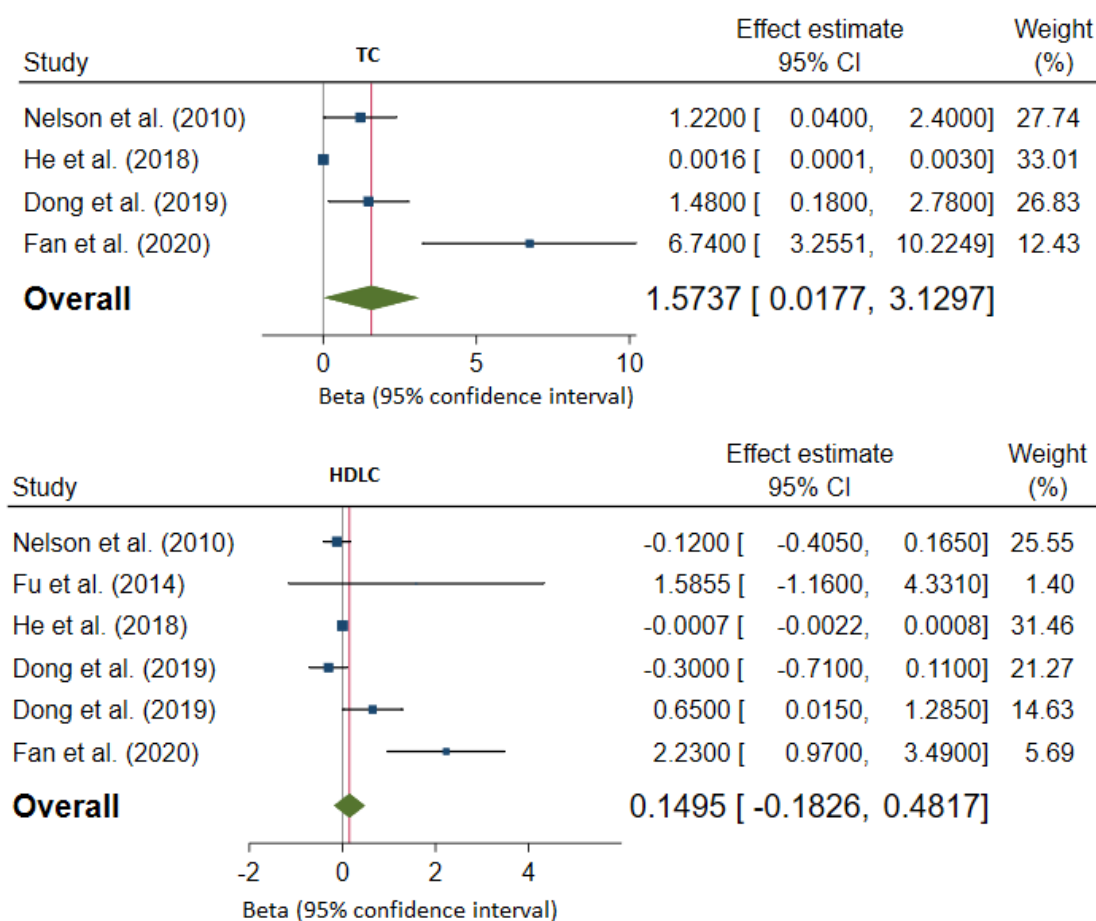
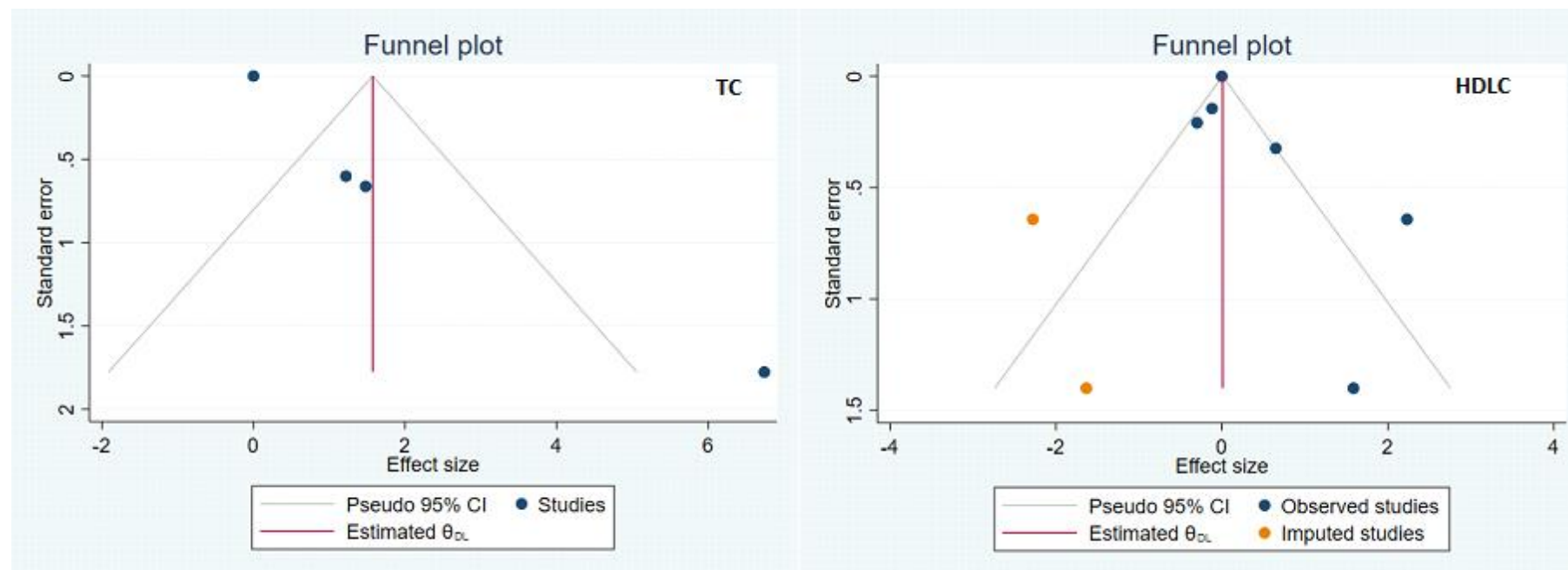


Figure A-3. Filled-in Funnel Plots to Evaluate Publication Bias of the PFOA and TC (Left) or HDLC (Right) Association in Studies Reporting Linear Associations.



Note: The funnel plot shows individual studies included in the analysis according to random-effect beta estimates (x-axis) and the standard error of each study-specific beta (y-axis). The red vertical line indicates the pooled estimate for all studies combined and the gray lines indicate pseudo 95% confidence limits around the pooled estimate. Number of observed studies: 4 (TC) and 6 (HDLC).

Figure A-4. Forest Plots Showing the Beta Coefficients Relating TC and HDLC to PFOA Concentrations in Each Study, and Pooled Estimates After Random-Effects Meta-Analysis.

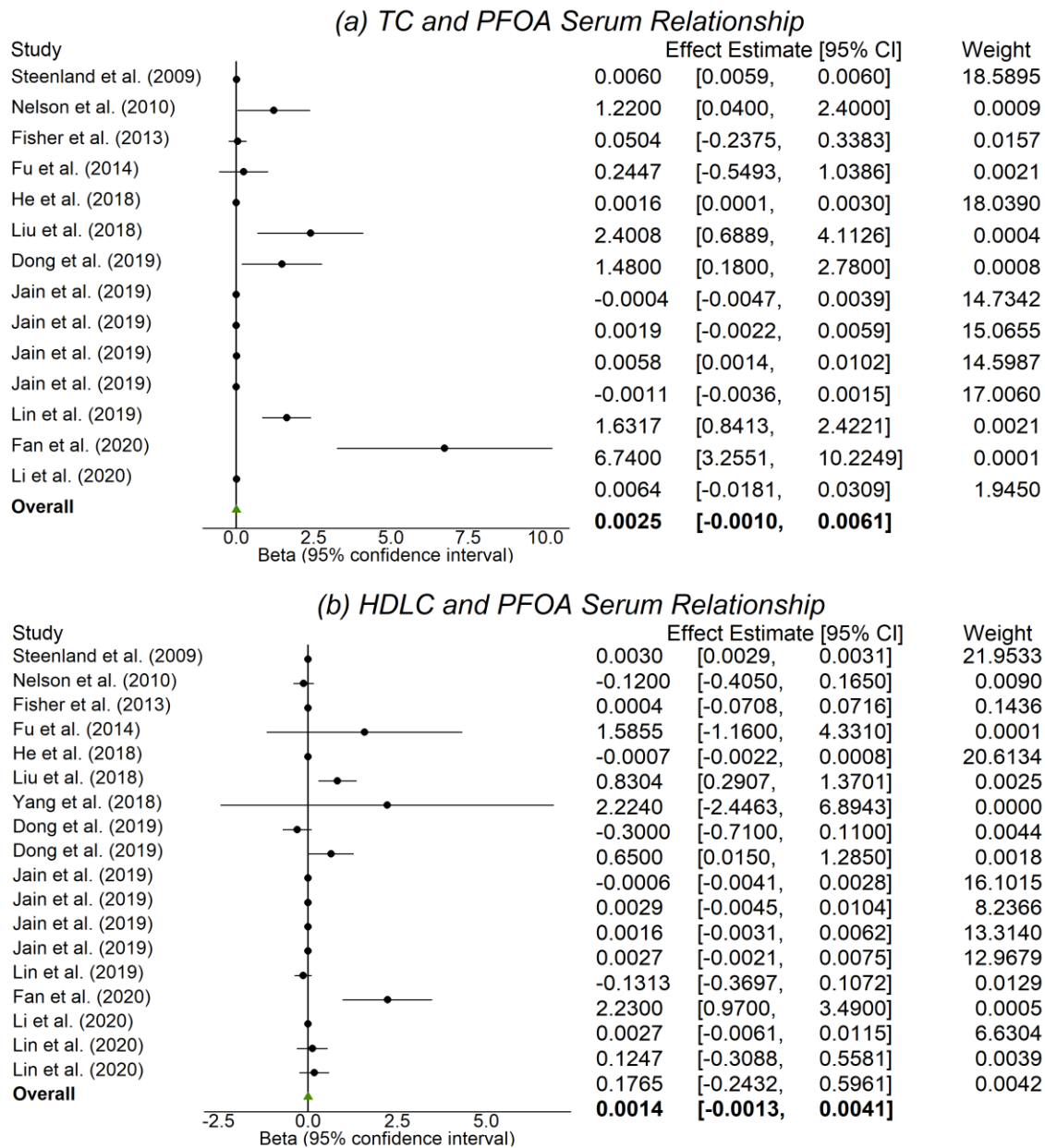
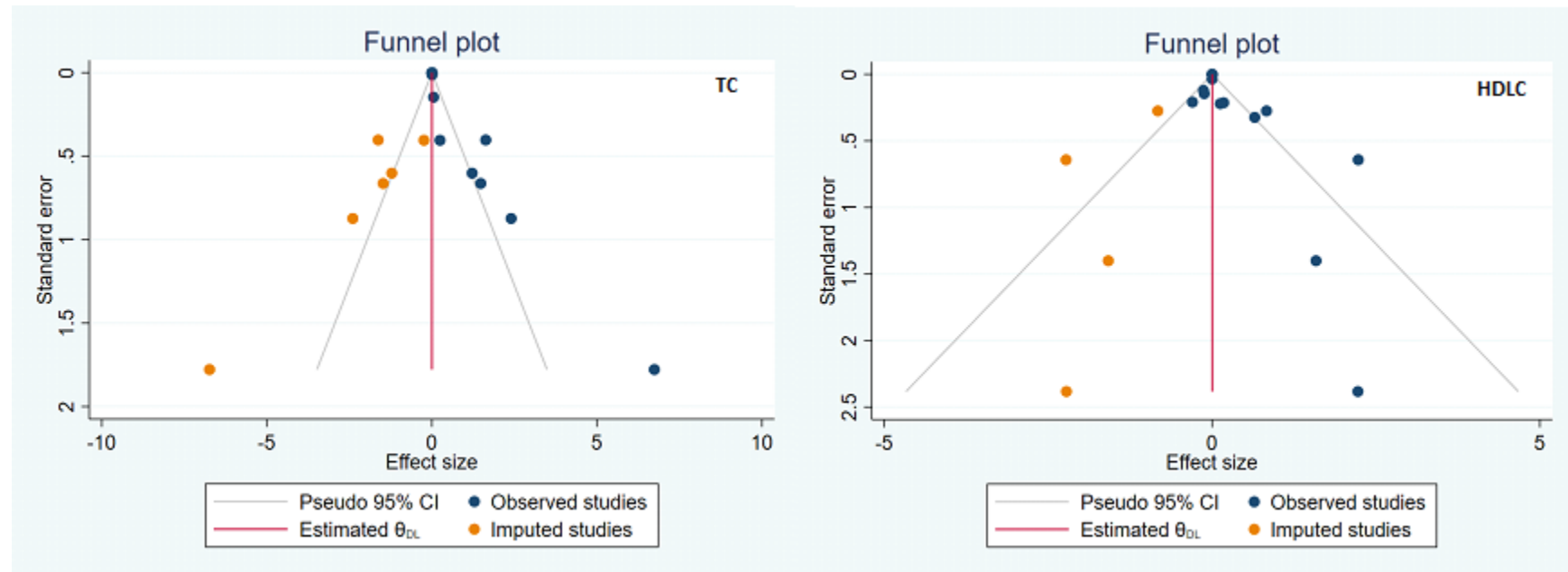


Figure A-5. Filled-in Funnel Plots to Evaluate Publication Bias of the PFOA and TC (Left) or HDLC (Right) Association.



Note: The funnel plot shows individual studies included in the analysis according to random-effect beta estimates (x-axis) and the standard error of each study-specific beta (y-axis). The red vertical line indicates the pooled estimate for all studies combined and the gray lines indicate pseudo 95% confidence limits around the pooled estimate. Number of observed studies: 11 (TC) and 13 (HDLC).

A.2.4 Slope Estimation for PFOS

When including the six studies reporting linear associations, there was a positive increase in TC of 0.08 (95% CI: -0.01, 0.16) mg/dL per ng/mL serum PFOS (p-value=0.064, I^2 =84%) that was significant at the 0.10 level. The association for PFOS and HDLC was positive but not statistically significant (Table A-3, Figure A-6). Adjusting for possible publication bias through funnel plots and trim-and-fill analysis suggested the imputation of additional studies; however, the magnitude or significance of the pooled associations did not change significantly (Figure A-7). Similar results were observed when the analysis excluded three overlapping NHANES studies.

When all studies were combined (11 studies, 14 results), EPA observed a borderline statistically significant positive increase in TC of 0.066 (95% CI: -0.001, 0.132) mg/dL per ng/mL serum PFOS (p-value=0.055, I^2 =100%) (Table A-3, Figure A-8). Adjusting for possible publication bias through funnel plots and trim-and-fill analysis suggested the imputation of three additional studies for TC and six for HDLC; however, the pooled effect estimates did not change significantly (Figure A-9). EPA observed similar results in leave-one-out analyses, sensitivity analyses restricted to U.S or Canadian general population studies, and analyses excluding Jain et al. (2019), estimates.

The pooled estimate based on the studies reporting linear associations was 0.079 (95% CI: -0.01, 0.16) and significant at the 0.10 level (p-value=0.064) and there is evidence supporting a positive and significant relationship between PFOS and TC: OST's review of 41 recent epidemiological studies showed positive associations between PFOS and TC in the general population and the meta-analysis performed with all studies combined showed a positive increase in TC per ng/mL serum PFOS that was significant at the 0.10 level. Given this weight of evidence, the large degree of heterogeneity in the pooled associations when all data were included, and the likelihood of bias that back-transformation of effect estimates with log-transformed outcomes or exposures could introduce (and difficulty with estimating the directionality of this bias towards or away from the null), EPA relied on the results from analyses restricted to studies reporting similar models, favoring the pooled slope (from the six studies reporting linear associations) of 0.08 mg/dL per ng/mL serum PFOS from the linear models for interpretability and use in the CVD risk reduction analysis.⁷ Because the association for PFOS and HDLC was not statistically significant and the available evidence does not support an association, EPA did not use an effect estimate for the PFOS-HDLC relationship in the CVD risk reduction analysis.

⁷ EPA will characterize uncertainty surrounding this estimate as described in Appendix E.

Table A-3: Results for PFOS Meta-Analyses

Group	Outcome	N Studies/ Number of Estimates	Beta (mg/dL per ng/mL)	95% CIs		p-value	Q ^a	p-value for Q	I ²	Tau ²
All Studies	TC	12/15	0.066	-0.001	0.132	0.055	630000	< 0.001	100	0.012
	HDLC	13/18	0.0002	-0.001	0.001	0.604	163.21	< 0.001	89.58	0
Linear Models Only	TC	5	0.079	-0.005	0.162	0.064	25.84	< 0.001	84.52	0.004
	HDLC	6	0.058	-0.005	0.120	0.070	31.61	< 0.001	84.18	0.003
Linear-Log Models Only ^b	TC	3/6	0.003	-0.003	0.009	0.342	7.82	0.166	36.07	0
	HDLC	5/9	0.007	-0.007	0.021	0.318	19.15	0.014	58.22	0.0002
Exclude Jain et al. (2019)	TC	11	0.136	0.02	0.253	0.028	510000	< 0.001	100	0.019
	HDLC	12/14	0	-0.002	0.003	0.833	131.29	< 0.001	90.10	0
All Studies ^c	TC	7	0.109	-0.016	0.234	0.088	1.2+e05	< 0.001	100	0.022
	HDLC	10/12	-0.0002	-0.0017	0.0014	0.806	123.94	< 0.001	91.1	0
Linear Models Only – No Overlap ^c	TC	2 ^d	0.192	-0.162	0.546	0.289	6.880	0.009	85.500	0.057
	HDLC	3/4	0.061	-0.009	0.132	0.089	24.330	< 0.001	87.70	0.032

Notes:

^a Q statistics for heterogeneity. Tau² is the between-studies variance.

^b Units for the pooled beta coefficient are mg/dL per ln(ng/mL).

^c No overlapping NHANES data.

^d Data from Dong et al. (2019) and Château-Degat et al. (2010).

Figure A-6. Forest Plots Showing the Beta Coefficients Relating TC and HDLC to PFOS Concentrations in Each Study Reporting Linear Associations, and Pooled Estimates After Random-Effects Meta-Analysis.

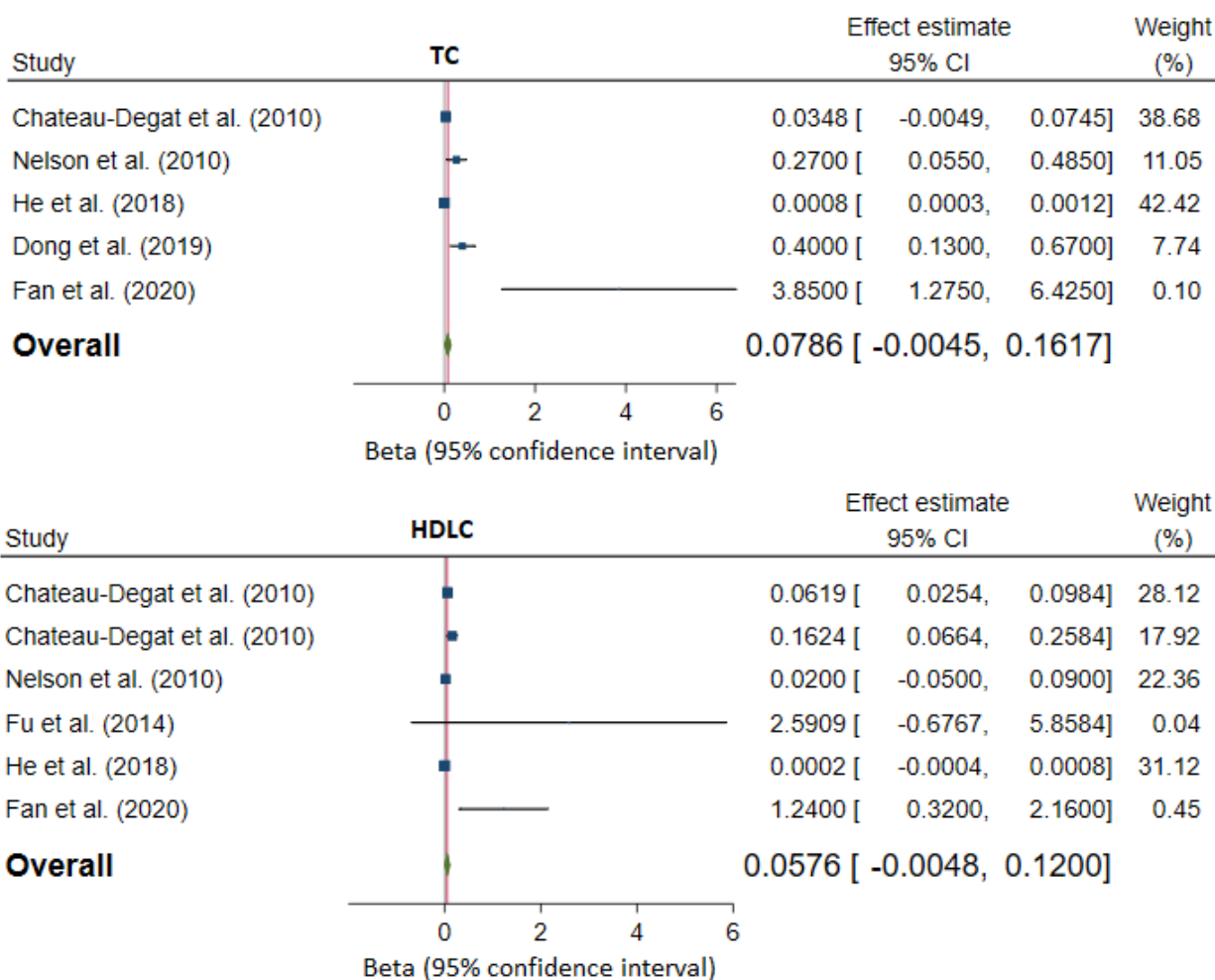
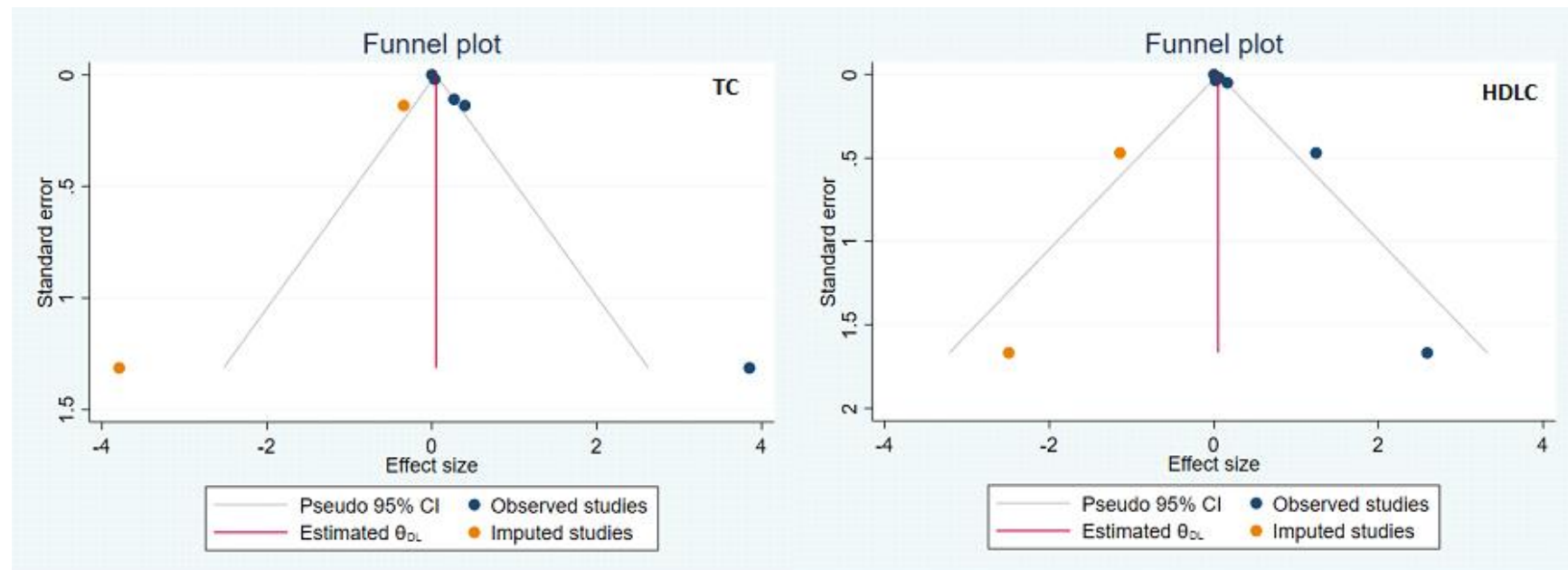


Figure A-7. Filled-in Funnel Plots to Evaluate Publication Bias of the PFOS and TC (Left) or HDLC (Right) Association in Studies Reporting Linear Associations.



Note: The funnel plot shows individual studies included in the analysis according to random-effect beta estimates (x-axis) and the standard error of each study-specific beta (y-axis). The red vertical line indicates the pooled estimate for all studies combined and the gray lines indicate pseudo 95% confidence limits around the pooled estimate. Number of observed studies: 5 (TC) and 6 (HDLC).

Figure A-8. Forest Plots Showing the Beta Coefficients Relating PFOS Concentrations to TC and HDLC in Each Study, and Pooled Estimates After Random-Effects Meta-Analysis.

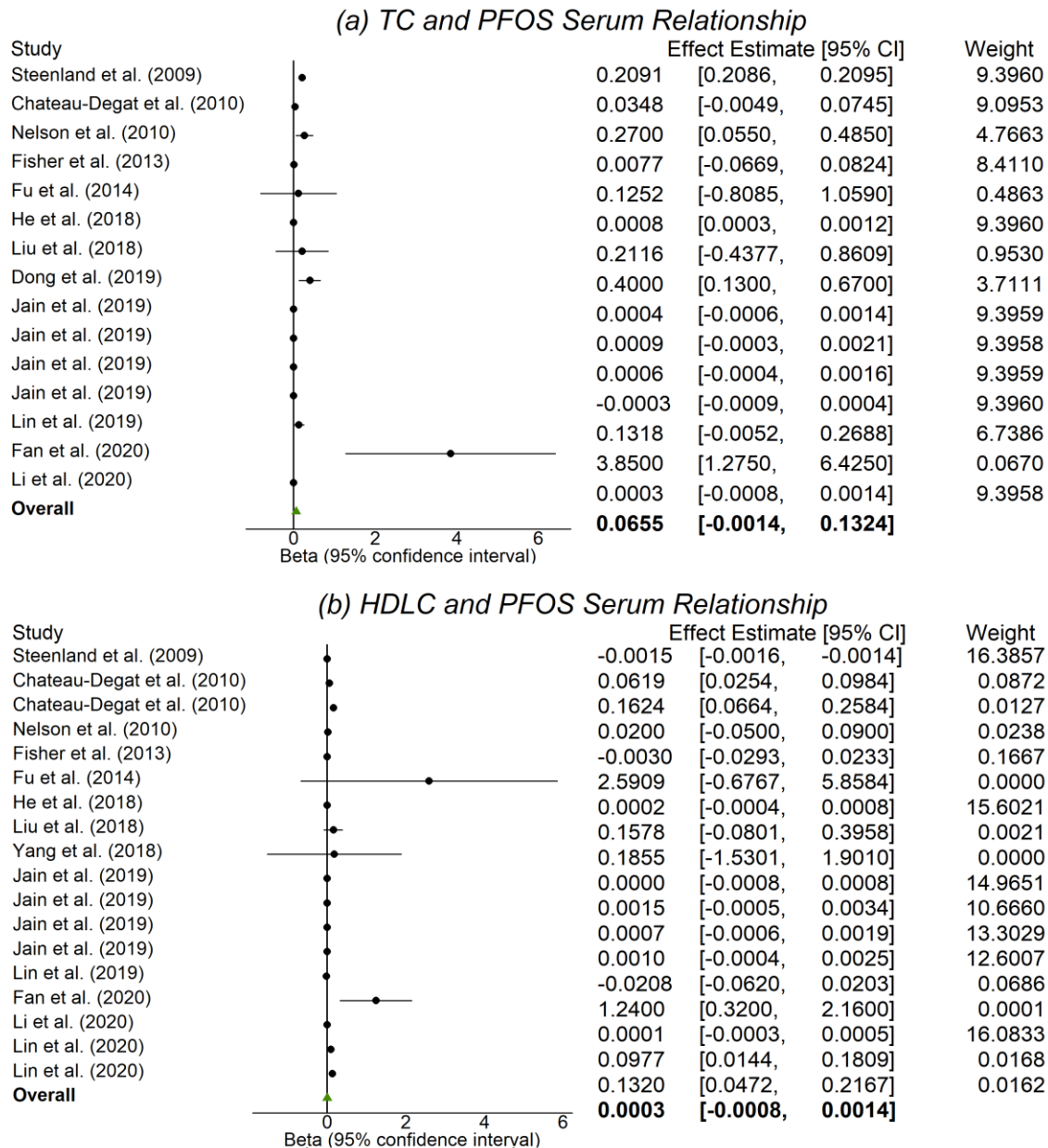
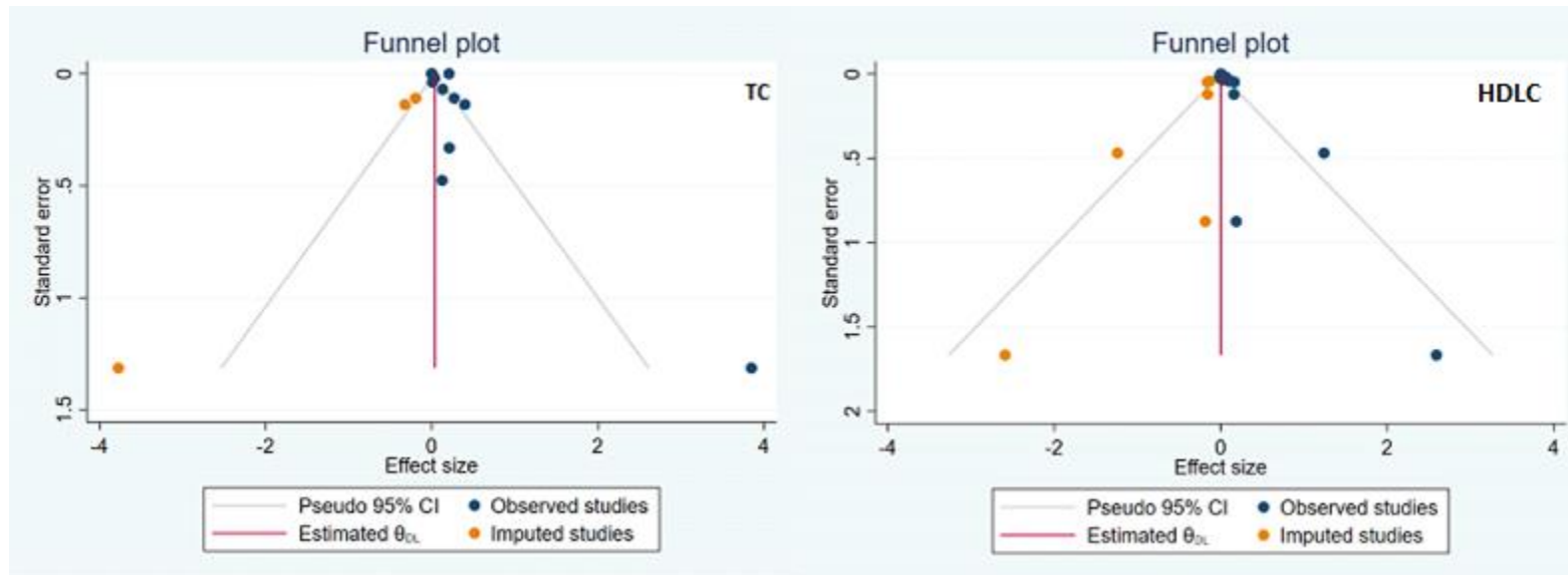


Figure A-9. Filled-in Funnel Plots to Evaluate Publication Bias of the PFOS and TC (Left) or HDLC (Right) Association.



Note: The funnel plot shows individual studies included in the analysis according to random-effect beta estimates (x-axis) and the standard error of each study-specific beta (y-axis). The red vertical line indicates the pooled estimate for all studies combined and the gray lines indicate pseudo 95% confidence limits around the pooled estimate. Number of observed studies: 12 (TC) and 13 (HDLC).

A.2.5 Limitations and Uncertainties

There are several limitations and uncertainties to consider when interpreting the findings from the meta-analyses. All of the studies included in the meta-analysis, except one (P.-I. D. Lin et al., 2019), are cross-sectional designs with various design or methodologic limitations. Although the cross-sectional nature of designs could raise concerns about reverse causality, measuring PFOA or PFOS and serum lipids concurrently was considered adequate in terms of exposure assessment timing. Given the long half-lives of PFOA and PFOS (with median half-lives of 2.7 and 3.5 years, respectively; Y. Li et al., 2018), current blood serum concentrations are expected to correlate well with past exposures. Several NHANES-based studies do not clearly report whether sampling weights were used in the analyses to account for the complex sampling design (Dong et al., 2019; He et al., 2018). Another limitation is that some of the NHANES-based studies used data from overlapping NHANES cycles. For example, Dong et al., 2019 used data from 2003–2014, while He et al. (2018) used data from 2003–2012; Jain et al. (2019) used data from 2005–2014; Fan et al. (2020) used data from 2011–2014; Liu et al. (2018) used data from 2013–2014; and Nelson et al. (2010) used data from 2003–2004. However, a sensitivity analysis excluding three of the six NHANES studies supported the main findings. Adding to the complexity, studies used various statistical models for estimating the associations of interest (including NHANES-based studies). Most studies provided measurements of PFOA and PFOS in serum, except in three studies that used measurements in plasma (Château-Degat et al., 2010; Fisher et al., 2013; P.-I. D. Lin et al., 2019). There is also a large degree of heterogeneity in the studies included in the meta-analysis, which remains even after restricting analyses to similar models.

There also are several limitations to the existing approaches for evaluating statistical heterogeneity and the potential for publication bias. In small meta-analyses, the I^2 statistic is imprecise and biased, and thus should be interpreted cautiously (von Hippel, 2015). The funnel plot asymmetry is a subjective assessment and is recommended only when at least 10 studies are included in the meta-analysis (J. P. T. Higgins et al., 2021). Furthermore, the Egger regression test and Begg's rank tests (Begg et al., 1994; M. Egger et al., 2008; Matthias Egger et al., 1997) may suffer from inflated type I error and limited power in certain situations, especially when there is a high degree of heterogeneity (L. Lin et al., 2018). Finally, the small number of studies reporting slopes from similar models limits the power of the meta-analysis.

Appendix B. CVD Model

This appendix provides details of the cardiovascular disease (CVD) model linking changes in total cholesterol to changes in incidence of first hard CVD events in populations exposed to perfluorooctanoic acid (PFOA)/perfluorooctanesulfonic acid (PFOS) through drinking water. Total cholesterol was linked to serum PFOA and serum PFOS, as described in Appendix A. First hard CVD events included in the model are non-fatal myocardial infarction (MI), non-fatal ischemic stroke (IS), and coronary heart disease (CHD) deaths. The model also captures post-acute CVD mortality experienced by the first non-fatal MI or IS survivors within 6 years of the initial event.

B.1 Model Overview and Notation

The CVD model is designed to estimate a time series of hard CVD event incidence for a population cohort characterized by sex, race/ethnicity, and age at the beginning of the evaluation period (i.e., 2023), and age- and sex-specific total cholesterol level time series estimated upstream. The first hard CVD event incidence estimates are generated using the Pooled Cohort Atherosclerotic Cardiovascular Disease (ASCVD) model (Goff et al., 2014), whose predictors include age, cholesterol levels, blood pressure, smoking status, and diabetes status. For those ages 40–80, the ASCVD model predicts the 10-year probability of a hard CVD event—non-fatal MI, fatal and non-fatal IS, or CHD death—to be experienced by a person without a prior history of MI, IS, congestive heart failure, percutaneous coronary intervention, coronary bypass surgery, or atrial fibrillation. EPA models post-acute CVD mortality for survivors of the first MI or IS at ages 45–65 using race/ethnicity- and sex-specific estimates at 1-year and 5-year follow-up from Thom et al. (2001). For survivors of the first MI or IS at age 66 or older, EPA models post-acute CVD mortality using estimates at 1- to 6-year follow-ups from S. Li et al. (2019).

The CVD model integrates the ASCVD model predictions and post-acute CVD mortality estimates in the series of recurrent calculations that produce a life table estimate for the population cohort of interest (e.g., non-Hispanic White females aged 70 years at the beginning of the evaluation period). For each public water system (PWS), EPA evaluates 1,048 population cohorts defined by a combination of birth year and age in or after 2023 (i.e., pairs of (2023,0), (2022,1), (2021,2), ... , (1934,89)⁸ and pairs of (2024,0), (2025,0), ... , (2065,0)⁹), sex (males and females), and race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, Other). In addition to the standard life table components, such as the annual number of all-cause survivors and deaths for all ages, for ages 40–89, the CVD model estimates the number of surviving persons with and without a history of hard CVD events, the number of persons experiencing hard CVD events at a given age, and deaths from CVD and non-CVD causes at a given age.

Figure B-1 **Error! Reference source not found.** summarizes the main types of CVD model calculations for a population cohort age 0 at the start of the evaluation period.¹⁰ The CVD model

⁸ While the ASCVD model applies to ages 40–80, it generates 10-year hard CVD event risk predictions. Therefore, it is possible to use the model predictions to carry incidence calculations out to age 89.

⁹ Those born after 2065 will not reach age 40 by the end of the analysis period in 2104. Because the ASCVD model predictions apply to ages 40–89, changes in CVD risk cannot be quantified for these cohorts.

¹⁰ This initial population cohort age is chosen because it allows for the illustration of the full set of calculation types used in the CVD model.

calculations are identical across the race/ethnicity and sex demographic subgroups but use subgroup-specific coefficients.¹¹ For cohorts born prior to or in 2023, the CVD model is initialized using the PWS-specific number of persons alive at the beginning of 2023. For cohorts born after 2023, the CVD model is initialized using the projected PWS-specific number of persons aged 0 years in the corresponding future year (i.e., 2024–2104). PWS- and sex, race/ethnicity-, and age-specific population estimation and projection details are included in Appendix C. Once the model is initialized, the following types of calculations occur for each year within the 95-year simulation period:

- Recurrent **standard life table calculations** that rely on the all-cause age-specific annual mortality rates to evaluate the number of deaths among persons of a specific integer age and the number of survivors to the beginning of the next integer age. These calculations are executed whenever the current cohort age is in the 0–39 range. They are represented by the green segments of the timeline shown in Figure B-1.
- Recurrent **life table calculations that separately track subpopulations with and without a history of hard CVD events**, including estimation of the number of annual CVD and non-CVD deaths (in either subpopulation), as well as the number of annual post-acute CVD deaths experienced by survivors of the first hard CVD events that occurred, at most, 5 years ago. These calculations are executed whenever the current cohort age is in the 40–89 range. These calculations are represented by the red segment of the timeline in Figure B-1. Figure B-2 further illustrates the year-specific calculations required for explicit tracking of subpopulations with and without a hard CVD event history. These calculations are also summarized in Table B-1.
- Recurrent **life table calculations tracking survivors of the first hard CVD event at ages 85–89 who attain ages 90–94** in order to estimate the number of annual post-acute CVD deaths that occur within 6 years of the initial event. These calculations are represented by the orange segment of the timeline in Figure B-1. Table B-1 **Error! Reference source not found.** and Figure B-2 provide additional details.

¹¹ There are different ASCVD model coefficients for non-Hispanic White and non-Hispanic Black males and females. The figure shows the generalized approach of the CVD model.

Figure B-1: Overview of Life Table Calculations in the CVD Model.

Note: The figure illustrates the model for population cohort age 0 at the beginning of the evaluation period (i.e., calendar year 2023). The model is initialized using an age 0 PWS-specific population (see Appendix C for PWS population estimation and projection details).

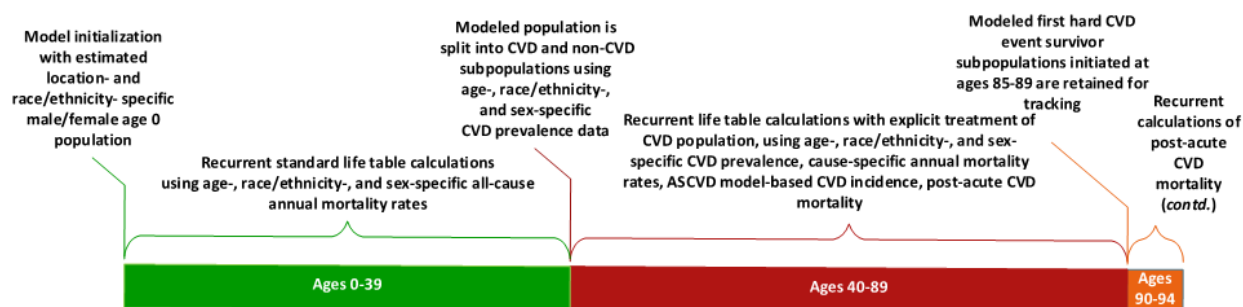


Table B-1 offers a more detailed summary of the year-specific calculations required for explicit tracking of subpopulations with and without a hard CVD event history illustrated in Figure B-2. Some of the calculation steps differ depending on whether the baseline or the treatment scenario is evaluated. In addition to providing estimates of the quantities of interest (e.g., the number of persons experiencing their first CVD event in a given year), the execution of the baseline scenario allows EPA to obtain age-specific baseline CVD mortality rates for the CVD subpopulation that are calibrated to the relevant baseline data (i.e., cause-specific mortality in the general population, CVD prevalence statistics, and baseline incidence of first hard CVD events). These calibration factors are not affected by differences between baseline and treatment scenario conditions. As shown in Table B-1 (Steps 6, 7, 9, and 10), EPA uses these rates to support CVD and non-CVD subpopulation calculations under treatment scenarios.

Figure B-2 and notes to Table B-1 provide additional information on the post-acute CVD mortality estimation. Each person included in the surviving current age-specific incident CVD subpopulation¹² (corresponding to the group F result in Figure B-2 or the Step 9 result in Table B-1) is tracked for 5 additional years to estimate the number of CVD deaths occurring in that timeframe. The recurrent estimates rely on age-specific non-CVD mortality, estimated based on CDC life table data and age- and sex-specific annual CVD mortality rates, and age- and post-acute CVD mortality, estimated based on Thom et al. (2001) and S. Li et al. (2019).

¹² For example, persons who experienced their first non-fatal MI or IS at age 70 and survived through the first post-event year.

Figure B-2: CVD Model Calculations Tracking CVD and Non-CVD Subpopulations for a Specific Current Age of Cohort.

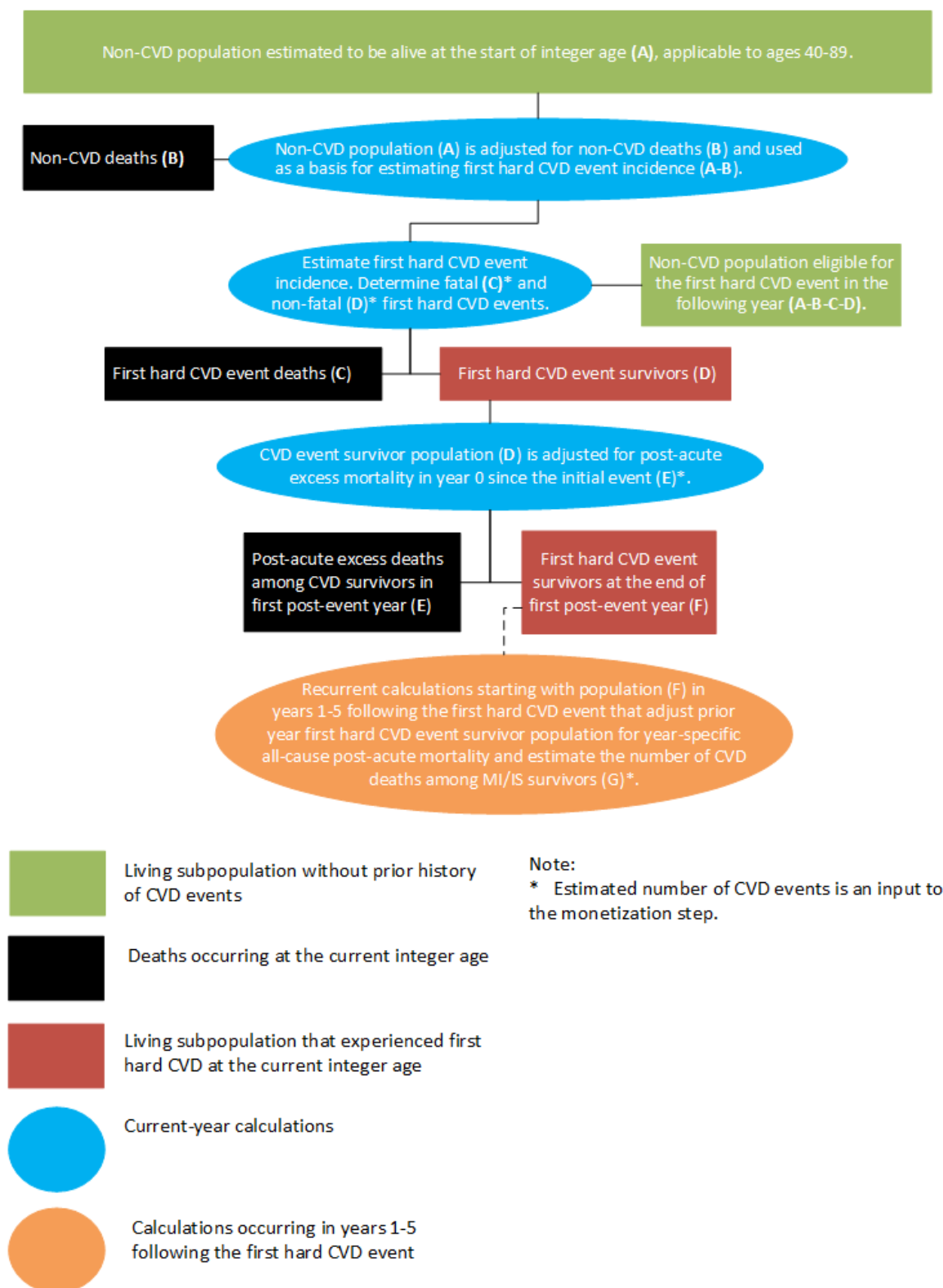


Table B-1: Year-Specific Calculations Required for Explicit Tracking of Subpopulations With and Without a Hard CVD Event History, Ages 40–89

Calculation Step	Baseline Calculations	Treatment Scenario Calculations
1. Estimate the sizes of subpopulations with and without a hard CVD event history alive at the start of the year.	National-level CVD prevalence statistics are used to allocate overall population alive at the start of the year.	The sizes of these subpopulations are computed internally, based on internal quantities estimated in the previous year (see Step 13).
2. Estimate the number of non-CVD deaths in subpopulations with and without a hard CVD event history.	Sex-, race/ethnicity-, and age-specific mortality rates for causes other than CVD are applied to the surviving subpopulation size (Step 1) to estimate the number of non-CVD deaths in the CVD subpopulation and non-CVD subpopulation.	
3. Estimate the <i>uncalibrated*</i> initial number of persons experiencing first hard CVD events .	The non-CVD subpopulation (Step 1) adjusted for non-CVD deaths (Step 2) is multiplied by the probability of the first hard CVD event generated by the ASCVD model.	
4. Estimate the <i>uncalibrated*</i> number of CVD deaths among persons experiencing first hard CVD events .	The national-level statistics (see section Error! Reference source not found.) and results in Thom et al. (2001) and S. Li et al. (2019) are used to estimate acute and post-acute CVD mortality** among persons experiencing their first hard CVD event in the current year (Step 3).	
5. Estimate the <i>uncalibrated*</i> end-of-year number of persons who experienced their first hard CVD event in the current year .	The initial number of persons experiencing their first hard CVD event (Step 3) is reduced by the number of CVD deaths in this population occurring within the same year (Step 4).	
6. Estimate the number of CVD deaths in the overall population.	Sex-, race/ethnicity-, and age-specific CVD mortality rates are applied to the overall surviving population size (Step 1) net of the total number of non-CVD deaths (Step 2).	The number of deaths among persons experiencing their first CVD event this year (Step 10) is added to the number of deaths in the start-of-the year CVD subpopulation (Step 1) multiplied by the corresponding baseline CVD death rate in the CVD subpopulation (Step 12 baseline).
7. Estimate the CVD event incidence consistent with the change in published baseline CVD prevalence .	Current-year CVD incidence is calculated as the difference in next year's CVD prevalence (Step 1 applied to the next year's population) and this CVD prevalence (Step 1) with an addition of CVD deaths (Step 6) and non-CVD deaths in the CVD population (Step 2).	Current-year CVD incidence is calculated as the product of the raw ASCVD model-based incidence adjusted for post-acute mortality (Step 4 + Step 5) and the CVD event incidence calibration factor (Step 8 baseline).

Table B-1: Year-Specific Calculations Required for Explicit Tracking of Subpopulations With and Without a Hard CVD Event History, Ages 40–89

Calculation Step	Baseline Calculations	Treatment Scenario Calculations
8. Estimate the CVD event incidence calibration factor .	Divide the reported prevalence and CVD mortality rate-based CVD event incidence (Step 7) by the raw ASCVD model-based incidence adjusted for post-acute mortality (Step 4 + Step 5).	Not applicable
9. Estimate the end-of-year number of persons who experienced their first hard CVD event in the current year .	Multiply the uncalibrated estimate of first hard CVD event survivors (Step 5) by the minimum of the calibration factor value (Step 8 baseline) and 1 (i.e., min (factor value,1)) to ensure that the ASCVD model-based CVD incidence does not exceed the reported statistics-based CVD incidence.	
10. Estimate the number of CVD deaths among persons experiencing a first hard CVD event .	Multiply the uncalibrated estimate of the number of deaths among persons experiencing their first hard CVD event this year (Step 4) by the minimum of the calibration factor value (Step 8) and 1 (i.e., min (factor value,1)) to ensure that the ASCVD model-based CVD incidence does not exceed the reported statistics-based CVD incidence.	
11. Estimate the number of post-acute CVD deaths that occur within 6 years of the initial non-fatal MI or IS .	The post-acute all-cause mortality in years 1–5 of the initial non-fatal MI/IS,** U.S. life table data, and U.S. general population CVD mortality data are combined with the number of first hard CVD event survivors (Step 9) to recurrently estimate CVD mortality.	
12. Estimate the baseline CVD mortality rate in the CVD subpopulation .	<p>Compute the baseline CVD mortality rate for subpopulations with a CVD event history, which is an important input for the treatment scenario life table calculations but is not reported in published statistics.</p> <p>This rate is estimated by subtracting the number of deaths among persons experiencing their first hard CVD event (Step 10) from the overall number of CVD deaths (Step 6) and dividing the result by the initial size of the CVD subpopulation (Step 1).</p>	Not applicable

Table B-1: Year-Specific Calculations Required for Explicit Tracking of Subpopulations With and Without a Hard CVD Event History, Ages 40–89

Calculation Step	Baseline Calculations	Treatment Scenario Calculations
13. Estimate next year’s CVD subpopulation and non-CVD subpopulation for the treatment scenario.	Not applicable	<p>The size of the next year’s subpopulation with a CVD event history is estimated by adding CVD incidence (Step 7) to the CVD population at the start of the year (Step 1) and subtracting the number of non-CVD deaths in the CVD subpopulation (Step 2) and the number of CVD deaths (Step 6).</p> <p>The size of the next year’s subpopulation without a CVD history is estimated by subtracting the CVD incidence (Step 7) and the number of non-CVD deaths in the non-CVD subpopulation (Step 2) from the non-CVD subpopulation in the beginning of the year (Step 1).</p>

Notes:

* Note that ASCVD-based estimates may be inconsistent with the recent CVD prevalence statistics. This is because (1) ASCVD events are a subset of all CVD events that may occur in the CVD population, and (2) the ASCVD model is not a perfect fit to the recent prevalence data. Therefore, EPA calibrates raw ASCVD model-based predictions to the baseline CVD prevalence and mortality statistics at baseline.

** The recurrent estimates rely on non-CVD mortality and post-acute CVD mortality specific to the year since the initial event based on Thom et al. (2001) and Li et al. (2019).

Abbreviations: ASCVD – atherosclerotic cardiovascular disease, CVD – cardiovascular disease

Table B-2 summarizes the data elements and notation of the CVD model. The CVD model elements fall into four categories: indices, data, quantities computed upstream, and internally computed quantities. Information sources and computational notes for the model elements identified as “data” are fully described in Appendix D. Total cholesterol change, $\Delta\tau_{b,a,s,t}$, is a birth year, age, sex, and calendar year-specific quantity computed upstream for the treatment scenarios as described in section 4 of the main document.¹³ Section B.2 describes the estimation of first hard CVD event incidence and post-acute CVD mortality, which are internally computed quantities. Derivation of the remaining internally computed quantities for the baseline life table is given in section B.3.1 and section B.3.2, while derivation of those quantities for the treatment scenario life table is given in section B.3.3.

¹³ Total cholesterol change for the baseline life table calculations is 0 by definition.

Table B-2: CVD Life Table Model Elements and Notation Summary

Model Element	Element Type	Definition
a	Index	Current integer age, $A = \{0, 1, 2, \dots, 94\}$
t	Index	Current calendar year, $t = 0$ marks the beginning evaluation period, $t = T$ marks the end of evaluation period
b	Index	Calendar birth year, $B = \{-T, \dots, 0, 1, \dots, T - 40\}$
s	Index	Sex, $S = \{\text{male}, \text{female}\}$
r	Index	Race/Ethnicity, $R = \{\text{non-Hispanic White}, \text{non-Hispanic Black}, \text{other}\}$
f	Index	First hard non-fatal CVD event type, $F = \{\text{non-fatal MI}, \text{non-fatal IS}\}$
p	Index	Population type: CVD – population with a history of hard CVD events; OTH – non-CVD population
c	Index	Cause of death: CVD – cardiovascular disease death; OTH – death from causes other than cardiovascular disease
k	Index	Number of years elapsed since first hard CVD event, $K = \{0, 1, 2, 3, 4, 5\}$
$l_{b,a,s,r,\max(0,b)}$	Data	Living population of age a , sex s , and race/ethnicity r , born in year b , at the beginning of the evaluation period for the cohort: $t = \max(0, b)$
$l_{b,a,s,r,t}$	Internally computed quantity	Living population born in year b , of sex s and race/ethnicity r , at the beginning of integer age a and calendar year t
$d_{b,a,s,r,t}$	Internally computed quantity	Number of all-cause deaths in population born in year b , of sex s and race/ethnicity r , at integer age a and calendar year t
$\pi_{a,s,r}$	Data	Prevalence rate of persons with past experience of hard CVD events at age a , sex s , and race/ethnicity r
$l_{b,a,s,r,t,p}$	Internally computed quantity	Living population born in year b , of type p , sex s , and race/ethnicity r , at the beginning of integer age a and calendar year t . Note that $l_{b,0,s,r,t,\text{CVD}} \equiv 0$, i.e., we assume that people who have just been born do have CVD history by definition.
$d_{b,a,s,r,t,p,c}$	Internally computed quantity	Number of deaths from cause c in population born in year b , of type p , sex s , and race/ethnicity r , throughout integer age a and calendar year t ; deaths from cardiovascular causes occur only in the CVD population (i.e., $d_{b,a,s,r,t,\text{OTH},\text{CVD}} \equiv 0$)
$q_{a,s,r}$	Data	General population probability of all-cause death at integer age a , sex s , race/ethnicity r
$q_{a,s,r,c}$	Data	General population probability of death from cause c at integer age a , sex s , race/ethnicity r
$\Delta\tau_{b,a,s,t}$	Quantity computed upstream	Modeled change in total cholesterol for population born in year b , of sex s , age a , in calendar year t ; this quantity is set to 0 for baseline calculations

Table B-2: CVD Life Table Model Elements and Notation Summary

Model Element	Element Type	Definition
$i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t})$	Internally computed quantity	Incidence rate of first hard CVD events for persons born in year b , of sex s and race/ethnicity r at age a and calendar year t ; this rate is computed using the ASCVD model that applies to ages 40–80 and produces the 10-year probability of the first hard CVD event
$\gamma_{a,s,r,f}$	Data	Share of first non-fatal hard CVD event type f among all first hard CVD events at age a , sex s , race/ethnicity r
$\rho_{b,a,s,r}$	Internally computed quantity	Rate of CVD deaths in CVD population born in year b , alive at the beginning of age a , for sex s and race/ethnicity r
$\mu_{a,s,r,f,k}$	Data	Probability of post-acute CVD death in age a , sex s , and race/ethnicity r CVD population who experienced first type f non-fatal hard CVD event k integer years ago
$x_{b,a,s,r,t}$	Internally computed quantity	Incident CVD population born in year b , of sex s and race/ethnicity r , at the beginning of integer age a and calendar year t
$\chi_{b,a,s,r,t}$	Internally computed quantity	Calibration factor for the incident CVD population born in year b , of sex s and race/ethnicity r , at the beginning of integer age a and calendar year t
$\tilde{n}_{b,a,s,r,f,t,0}$	Internally computed quantity	Uncalibrated number of living age a , sex s , and race/ethnicity r persons born in year b , whose first type f non-fatal hard CVD event occurred 0 years ago, corresponding to calendar year t
$n_{b,a,s,r,f,t,k}$	Internally computed quantity	Number of living age a , sex s , and race/ethnicity r persons born in year b , whose first type f non-fatal hard CVD event occurred k years ago, corresponding to calendar year t
$\tilde{m}_{b,a,s,r,t,0}$	Internally computed quantity	Uncalibrated number of CVD deaths among those born in year b , age a , sex s , and race/ethnicity r persons whose first hard CVD event occurred 0 years ago, corresponding to calendar year t
$m_{b,a,s,r,t,k}$	Internally computed quantity	Number of CVD deaths among those born in year b , age a , sex s , and race/ethnicity r persons whose first hard CVD event occurred k years ago, corresponding to calendar year t
$\Delta n_{b,a,s,r,f,t}$	Internally computed quantity	Difference between treatment scenario and baseline number of persons born in year b , of sex s and race/ethnicity r , whose first type f non-fatal hard CVD event occurred at age a , corresponding to calendar year t
$\Delta m_{b,a,s,r,t}$	Internally computed quantity	Difference between calendar year t treatment scenario and baseline number of CVD deaths among age a , sex s , and race/ethnicity r persons born in year b , who experienced their first hard CVD event during calendar years $t - 5, t - 4, \dots, t$
$\Delta N_{f,t}$	Internally computed quantity	Difference between treatment scenario and baseline number of persons whose first type f non-fatal hard CVD event occurred during calendar year t

Table B-2: CVD Life Table Model Elements and Notation Summary

Model Element	Element Type	Definition
ΔM_t	Internally computed quantity	Difference between treatment scenario and baseline number of year t CVD deaths among persons whose first hard CVD event occurred during calendar years $t - 5, t - 4, \dots t$

B.2 Hard CVD Event Incidence Estimation

In this section, EPA describes the process for estimating the probability of the first hard CVD event $i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t})$ using the ASCVD model (section B.2.1); the prevalence of persons with a history of hard CVD events $\pi_{a,s,r}$ (section B.2.2); the distribution of first hard CVD events by type, including the share of non-fatal first hard CVD events $\gamma_{a,s,r,f}$ (section B.2.3); and post-acute CVD mortality rates $\mu_{a,s,r,f,k}$ within 6 years of the initial event (section B.2.4).

B.2.1 Probability of the First Hard CVD Event

The first hard CVD event incidence estimates are generated by the Pooled Cohort ASCVD model (Goff et al., 2014). The ASCVD model is commonly used in clinical practice to estimate CVD risk for those aged 40–80 years. The ASCVD model predicts the 10-year probability of a hard CVD event—fatal and non-fatal MI, fatal and non-fatal IS, or CHD death—to be experienced by a person without a prior history of MI, IS, congestive heart failure, percutaneous coronary intervention, coronary bypass surgery, or atrial fibrillation.

Four large longitudinal community-based epidemiologic cohort studies have been combined to develop a geographically and racially diverse dataset used for the ASCVD model estimation: (1) the Atherosclerosis Risk in Communities Study (Williams, 1989), (2) the Cardiovascular Health Study (Fried et al., 1991), (3) the Coronary Artery Risk Development in Young Adults Study (Friedman et al., 1988), and (4) the Framingham Original and Offspring Cohort Study (Mahmood et al., 2014). Note that there are several other studies whose design is similar to the one used in Goff et al. (2014), including D’Agostino et al. (2001), D’Agostino et al. (2000), D’Agostino et al. (2008), D’Agostino et al. (1994), Pencina et al. (2009), Pencina et al. (2011), Wilson et al. (1998), and Uno et al. (2011). Except for Uno et al. (2011), who also used the Breast Cancer Survival Study (Chang et al., 2005), all of these studies used the Framingham cohort study data that are not as diverse as the data used to estimate the ASCVD model.

Table B-3 shows the ASCVD model coefficient estimates used in the analysis. The predictors of the ASCVD model include age, total cholesterol (TC) and high-density lipoprotein cholesterol (HDLC)¹⁴ concentrations, systolic blood pressure (BP), current smoking, diagnosed diabetes, and whether the subject is undergoing treatment for high BP. The model has been fit separately to four population subgroups: non-Hispanic White females, non-Hispanic Black females, non-

¹⁴ As discussed in Appendix A, the literature supports a non-zero upper bound on the relationship between serum PFOS and HDLC. Because the ASCVD model includes HDLC as one of the predictors, it is possible to account for the effects of PFAS on the CVD risk via changes in HDLC, in addition to changes in total cholesterol. This modeling step has not been implemented for the current analysis. However, a sensitivity analysis that evaluates CVD risk impacts of PFAS through both total cholesterol and HDLC links could be implemented in the future.

Hispanic White males, and non-Hispanic Black males. EPA applied sex-specific model coefficients for non-Hispanic Blacks to estimate CVD risk in Hispanic and non-Hispanic other race population subgroups based on validation of the ASCVD model against published statistics as described in section B.4.

Table B-3: ASCVD Model Coefficients

Variable Name	Model Coefficient			
	Non-Hispanic White Females	Non-Hispanic Black Females*	Non-Hispanic White Males	Non-Hispanic Black Males*
Ln Age (y)	-29.799	17.114	12.344	2.469
Ln Age, squared	4.884	–	–	–
Ln Total Cholesterol (mg/dL)	13.54	0.94	11.853	0.302
Ln Age × Ln Total Cholesterol	-3.114	–	-2.664	–
Ln HDL-C (mg/dL)	-13.578	-18.92	-7.99	-0.307
Ln Age × Ln HDL-C	3.149	4.475	1.769	–
Ln Treated Systolic BP (mm Hg)	2.019	29.291	1.797	1.916
Ln Age x Ln Treated Systolic BP	–	-6.432	–	–
Ln Untreated Systolic BP (mm Hg)	1.957	27.82	1.764	1.809
Ln Age x Ln Untreated Systolic BP	–	-6.087	–	–
Current Smoker (1=Yes, 0=No)	7.574	0.691	7.837	0.549
Ln Age × Current Smoker	-1.665	–	-1.795	–
Diabetes (1=Yes, 0=No)	0.661	0.874	0.658	0.645
Mean (Coefficient × Value), $\bar{x}_{s,r}'\beta_{s,r}$	-29.18	86.61	61.18	19.54
ASCVD Baseline Survival, $S_{s,r}$	0.9665	0.9533	0.9144	0.8954

Note: * Based on the results of ASCVD model validation exercises (section B.4), the models for non-Hispanic Black males and females are applied to other ethnic groups.

Source: Goff et al. (2014), Table A

In order to be used for risk estimation, the ASCVD model needs to be parameterized using values of the predictors shown in Table B-3 that are appropriate for the current age, sex, and race/ethnicity of the cohort being evaluated. As shown in Table B-2, current age, sex, and race/ethnicity are easily accessible indices of the CVD model. In turn, baseline values for the other ASCVD model predictors come from several public health surveys implemented by the Centers for Disease Control and Prevention, as detailed in Appendix D.

To compute the 10-year probability of the first hard CVD event for a birth year b , sex s and race/ethnicity r cohort at age a , EPA uses the ASCVD risk equation (Goff et al., 2014, Table 5) adjusted to express the type of scenario being evaluated (i.e., baseline or treatment scenario):

$$R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t}) = 1 - S_{s,r}^{\exp([\beta_{\tau,s,r} + \beta_{a\tau,s,r} \cdot \ln(a)] \cdot \ln(\tau_{a,s,r} + \Delta\tau_{b,a,s,t}) + x_{-\tau,a,s,r}' \beta_{-\tau,s,r} - \bar{x}_{s,r}' \beta_{s,r})} \quad (\text{Eq. B-1})$$

where

$R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t})$	probability of the first hard CVD event to occur between years t and $t + 9$ for a birth year b , sex s / race/ethnicity r person whose age at time t is a . $R_{b,a,s,r,t:t+9}(0)$ represents baseline 10-year first hard CVD event risk, whereas $R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t})$ expresses treatment scenario risk consistent with a birth year b -, age a -, sex s -, calendar year t -specific change in the baseline total cholesterol level $\Delta\tau_{b,a,s,t}$;
$S_{s,r}$	ASCVD baseline CVD event-free survival rate at 10 years, consistent with the sex s and race/ethnicity r of the cohort being evaluated (see parameter estimates in Table B-3);
$\tau_{a,s,r}$	baseline input value for the total cholesterol consistent with the current age a , sex s , and race/ethnicity r of the cohort being evaluated (see Appendix D);
$\beta_{\tau,s,r}$	ASCVD model coefficient for the log-total cholesterol predictor, consistent with the sex s and race/ethnicity r of the cohort being evaluated (see parameter estimates in Table B-3);
$\beta_{a\tau,s,r}$	ASCVD model coefficient for the interaction between log-current age and log-total cholesterol predictor, consistent with the sex s and race/ethnicity r of the cohort being evaluated (see parameter estimates in Table B-3);
$x_{-\tau,a,s,r}' \beta_{-\tau,s,r}$	inner product of the ASCVD model coefficient vector (excluding total cholesterol-related coefficients) and a vector of baseline input values (excluding total cholesterol-related inputs), consistent with the current age a , sex s , and race/ethnicity r of the cohort being evaluated (see parameter estimates in Table B-3 and Appendix D); and
$\bar{x}_{s,r}' \beta_{s,r}$	inner product of the ASCVD model coefficient vector and a vector of average input values in the ASCVD estimation dataset (see parameter estimates in Table B-3).

To obtain the annual probability of the first hard CVD event, EPA adjusts $R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t})$ as follows:

$$i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t}) = 1 - \left(1 - R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t})\right)^{\frac{1}{10}} \quad (\text{Eq. B-2})$$

where

$i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t})$ probability of the first hard CVD event to occur in year t for a birth year b , sex s / race/ethnicity r person whose age at time t is a ; and

$R_{b,a,s,r,t:t+9}(\Delta\tau_{b,a,s,t})$ probability of the first hard CVD event to occur between years t and $t + 9$ for a birth year b , sex s / race/ethnicity r person whose age at time t is a .

B.2.2 Prevalence of Past Hard CVD Events

Because the population evaluated for the first hard CVD event estimation excludes those with a history of hard CVD events, model inputs require information on the baseline prevalence of the past hard CVD event history in the U.S. population. EPA used the Medical Expenditure Panel Survey (MEPS) 2010–2017 data to estimate the prevalence of persons with a prior experience of hard CVD events, including MI, stroke, and other acute CHD events. MEPS is a nationally representative survey of the U.S. civilian non-institutionalized population implemented by the Agency for Healthcare Research and Quality (AHRQ). The survey has an overlapping panel design, tracking individuals for, at most, two years and interviewing participants, at most, six times. MEPS collects demographic, socioeconomic, and health status information on the first interview and in each subsequent interview asks about medical events experienced between the current and the previous interview (generally 4–5 months), as well as changes in employment status, health insurance coverage, and so forth. Appendix D provides additional information on MEPS public use files that have been used in this analysis.

The prevalence of persons with a prior experience of hard CVD events has been estimated by dividing the number person-years in MEPS interview rounds with a reported history of MI, stroke, or other CHD by the total number of person-years in subpopulations defined by sex and round-specific age. The estimated ratios have been adjusted for MEPS complex survey design.

Table B-4 shows the resulting estimates of sex-, race/ethnicity-, and age category-specific prevalence of persons with prior experience of hard CVD events, along with 95% confidence intervals that reflect sampling uncertainty. Compared with the prevalence estimates for females, the estimated prevalence is higher for males in all age categories and for all CVD event categories. Among adults aged 65 or older, estimated MI, other CHD, and overall prevalence is highest for non-Hispanic White males, while stroke prevalence is highest among non-Hispanic Black males. Regardless of the age category, the estimated prevalence of an MI history is higher for males, while the prevalence of a stroke history is higher for females. The prevalence of other CHD event history is approximately three to 10 times higher compared with the prevalence of an MI or stroke history.

Table B-4: Estimated Past Hard CVD Event Prevalence per 100,000

Sex	Age (years)	Race/Ethnicity	MI	Stroke	Other CHD	Overall
Males	18–44	NH White	632 (410–855)	495 (317–673)	5,709 (5,072–6,346)	6,292 (5,620–6,965)
	45–64	NH White	5,099 (4,569–5,629)	3,314 (2,804–3,823)	15,439 (14,523–16,355)	17,963 (16,930–18,995)
	65 or older	NH White	16,477 (15,088–17,865)	11,002 (9,956–12,047)	41,600 (40,040–43,161)	47,465 (45,831–49,099)
Males	18–44	NH Black	436 (146–726)	614 (304–924)	3,886 (2,998–4,773)	4,667 (3,651–5,684)
	45–64	NH Black	4,786 (3,928–5,644)	5,316 (4,222–6,409)	12,261 (10,801–13,720)	16,590 (14,898–18,282)
	65 or older	NH Black	13,768 (11,218–16,319)	18,908 (16,185–21,631)	30,307 (26,724–33,891)	42,090 (38,368–45,812)
Males	18–44	Hispanic	480 (293–667)	180 (75–285)	3,065 (2,479–3,651)	3,417 (2,816–4,019)
	45–64	Hispanic	4,299 (3,383–5,214)	3,010 (2,225–3,796)	9,979 (8,640–11,318)	12,584 (11,045–14,124)
	65 or older	Hispanic	14,071 (11,569–16,573)	8,254 (6,031–10,477)	25,866 (22,420–29,313)	30,548 (26,960–34,136)
Males	18–44	NH Other	347 (122–572)	342 (75–610)	3,262 (2,330–4,194)	3,669 (2,695–4,643)
	45–64	NH Other	4,338 (3,012–5,665)	2,693 (1,791–3,595)	11,339 (9,033–13,645)	13,638 (11,118–16,158)
	65 or older	Other	12,256 (9,167–15,344)	12,354 (8,911–15,798)	30,516 (25,051–35,982)	36,932 (31,240–42,624)
Females	18–44	NH White	439 (278–600)	830 (608–1,052)	6,262 (5,528–6,997)	6,954 (6,223–7,685)
	45–64	NH White	2,199 (1,841–2,557)	3,127 (2,595–3,659)	15,496 (14,522–16,469)	17,925 (16,791–19,059)
	65 or older	NH White	7,510 (6,686–8,335)	10,055 (9,098–11,011)	31,861 (30,278–33,445)	37,538 (35,913–39,162)
Females	18–44	NH Black	393 (204–582)	1,092 (783–1,402)	4,628 (3,917–5,338)	5,612 (4,847–6,378)
	45–64	NH Black	3,484 (2,808–4,160)	6,491 (5,640–7,343)	15,292 (13,915–16,670)	19,596 (17,981–21,210)

Table B-4: Estimated Past Hard CVD Event Prevalence per 100,000

Sex	Age (years)	Race/Ethnicity	MI	Stroke	Other CHD	Overall
Males	65 or older	NH Black	8,803 (7,130–10,476)	14,188 (12,304–16,071)	29,296 (26,441–32,151)	38,073 (35,102–41,045)
	18–44	Hispanic	313 (171–454)	717 (469–965)	3,690 (3,182–4,199)	4,363 (3,808–4,918)
	45–64	Hispanic	2,597 (1,947–3,248)	3,627 (2,864–4,391)	10,335 (9,066–11,604)	12,777 (11,361–14,193)
	65 or older	Hispanic	7,513 (5,953–9,073)	9,469 (7,385–11,554)	23,149 (20,350–25,948)	29,186 (26,206–32,167)
Females	18–44	NH Other	722 (123–1,320)	383 (90–675)	4,569 (3,181–5,957)	4,884 (3,502–6,266)
	45–64	NH Other	1,292 (710–1,874)	2,770 (1,679–3,860)	11,098 (8,978–13,218)	13,148 (10,758–15,538)
	65 or older	NH Other	4,150 (2,557–5,742)	7,321 (5,054–9,589)	19,001 (15,308–22,694)	23,463 (19,638–27,288)

Abbreviations: MI – myocardial infarction (ICD9=410 or MIDX=1); NH – non-Hispanic; Other CHD – other coronary heart disease (ICD9=413,414,427,428 or CHDDX=1, ANGI DX=1, OHRTDX=1); Stroke (ICD9=433,434,435,436 or STRKDX=1); 95% confidence interval shown in parentheses below the point estimate.

Source: EPA analysis based on MEPS, 2010–2017

B.2.3 Distribution of Fatal and Non-Fatal First Hard CVD Events

The ASCVD model predicts the risk of a composite hard CVD event (i.e., MI, IS, or CHD death). However, modeling requires separate tracking of morbidity and mortality for life table calculation purposes. In addition, acute-phase mortality and morbidity valuation depends on the endpoint (i.e., MI or IS). Therefore, EPA used MEPS 2010–2017 data to estimate the distribution of first hard CVD events by type of condition (i.e., MI, stroke, and other CHD). EPA estimated the incidence of first hard CVD events by dividing the number of person-years in MEPS interview rounds with reported new occurrences of MI, stroke, or other CHD by the number of person-years in MEPS interview rounds without resorted prior experience of CVD events, in subpopulations defined by race/ethnicity, sex and round-specific age. EPA adjusted the estimated ratios for MEPS complex survey design. Distribution of CVD events by condition type was calculated based on the estimated condition-specific incidence rates.

Table B-5 shows the resulting estimates of sex-, race/ethnicity-, and age category-specific first hard CVD event incidence, along with 95% confidence intervals that reflect sampling uncertainty. The table also shows the distribution of first hard CVD events by event type. In males, 15% to 17% of first hard CVD events are MIs, whereas 13% to 20% of first hard CVD events are strokes. In females, 8% to 12% of first hard CVD events are MIs, whereas 17% to 28% of first hard CVD events are strokes. The shares of MIs and strokes increase with age for

both sexes. Among adults aged 65 or older, estimated MI, stroke, other CHD, and overall incidence are highest for non-Hispanic White males and females.

Table B-5: Estimated First Hard CVD Event Incidence and Distribution by CVD Event Type

Sex	Age (years)	Race/ Ethnicity	MI	Stroke	Other CHD	Overall
Males	18–44	NH White	82 (29–135)	57 (3–110)	454 (299–609)	540 (375–705)
	45–64	NH White	356 (225–486)	333 (194–471)	1,536 (1,213–1,859)	2,048 (1,678–2,417)
	65 or older	NH White	1,326 (679–1,973)	2,001 (1,248–2,754)	6,233 (5,035–7,431)	8,125 (6,651–9,598)
Males	18–44	NH Black	23 (–3–49)	81 (4–159)	363 (156–570)	447 (227–668)
	45–64	NH Black	235 (64–407)	805 (399–1,211)	1,039 (676–1,401)	1,862 (1,339–2,385)
	65 or older	NH Black	319 (–1–639)	765 (76–1,454)	2,332 (1,217–3,447)	3,273 (1,926–4,621)
Males	18–44	Hispanic	52 (6–99)	40 (–4–83)	135 (55–214)	212 (111–313)
	45–64	Hispanic	276 (72–479)	421 (2–839)	735 (419–1,052)	1,142 (625–1,659)
	65 or older	Hispanic	951 (285–1,618)	816 (349–1,283)	2,747 (1,432–4,061)	3,915 (2,440–5,390)
Males	18–44	NH Other	72 (–70–215)	85 (–54–223)	121 (35–207)	278 (63–493)
	45–64	NH Other	830 (171–1,489)	548 (39–1,057)	1,513 (643–2,383)	2,537 (1,356–3,718)
	65 or older	NH Other	665 (–14–1,343)	1,232 (431–2,033)	2,940 (1,496–4,383)	4,251 (2,506–5,997)
Females	18–44	NH White	56 (–21–134)	135 (54–216)	492 (317–668)	646 (437–856)
	45–64	NH White	140 (56–225)	407 (193–620)	1,423 (1,109–1,737)	1,865 (1,490–2,240)
	65 or older	NH White	831 (533–1,130)	2,102 (1,498–2,705)	4,271 (3,461–5,081)	6,294 (5,358–7,231)
Females	18–44	NH Black	96 (1–191)	57 (5–108)	487 (279–695)	597 (360–834)
	45–64	NH Black	196 (74–318)	530 (247–812)	1,168 (793–1,543)	1,754 (1,285–2,223)
	65 or older	NH Black	382 (8–756)	1,607 (762–2,453)	3,383 (2,221–4,545)	4,546 (3,179–5,913)

Table B-5: Estimated First Hard CVD Event Incidence and Distribution by CVD Event Type

Sex	Age (years)	Race/ Ethnicity	MI	Stroke	Other CHD	Overall
Females	18–44	Hispanic	38	78	308	392
			(-24–100)	(25–131)	(130–487)	(190–595)
	45–64	Hispanic	145	308	664	1,065
			(33–257)	(76–541)	(393–936)	(699–1,432)
	65 or older	Hispanic	992	1,321	2,610	4,456
			(215–1,768)	(611–2,031)	(1,670–3,550)	(3,348–5,564)
Females	18–44	NH Other	47		315	315
			(-46–141)	Omitted	(42–589)	(42–589)
	45–64	NH Other	201	399	759	1,297
			(-6–409)	(74–724)	(259–1,259)	(627–1,967)
	65 or older	NH Other	576	1,328	2,689	4,349
			(-43–1,195)	(381–2,276)	(1,234–4,144)	(2,463–6,234)

Abbreviations: MI – myocardial infarction (ICD9=410 or MIDX=1), NH – non-Hispanic, Stroke (ICD9=433,434,435,436 or STRKDX=1), Other CHD – other coronary heart disease (ICD9=413,414,427,428 or CHDDX=1, ANGIDX=1, OHRTDX=1); 95% confidence interval shown in parentheses below the point estimate.

The ASCVD model predicts the risk of first MI (fatal and non-fatal), IS (fatal and non-fatal), or other fatal CHD within the next 10 years. Notably, other non-fatal CHD events are not included among the CVD event types predicted by the ASCVD model (Goff et al., 2014). Because MEPS data do not have sufficient information to estimate acute-phase CVD event mortality, EPA used AHRQ’s Healthcare Cost and Utilization Project (HCUP) data on hospital mortality to allocate CVD events into fatal and non-fatal categories. Appendix D provides additional information on the in-hospital mortality data.

Table B-6 shows sex- and age category-specific probability of in-hospital CVD event death based on HCUP 2017 inpatient data (Agency for Healthcare Research and Quality, 2017). Probability of an in-hospital death is highest for MI events (4.64%), followed by IS events (4.01%), and then other CHD events (1.07%). This probability grows with age across all CVD event types and is higher for females when compared with males.

Table B-6: Probability of Hospital Death for a Hard CVD Event

Category	MI (%)	IS (%)	Other CHD (%)
Overall	4.65	4.01	1.07
Age (years)			
18–44	1.43	1.91	0
45–64	2.60	2.46	0.67
65–84	5.42	3.88	1.23
85 or older	9.80	7.29	3.14
Sex			
Males	4.41	3.71	1.01
Females	5.04	4.30	1.20

Abbreviations: IS – ischemic stroke (ICD10=I63), MI – myocardial infarction (ICD10=I21), Other CHD – other coronary heart disease (ICD10=I20, I22-I25)

Source: HCUP 2017 (Agency for Healthcare Research and Quality, 2017)

EPA combined estimates in Table B-5 and Table B-6 to derive the ASCVD event distribution over the following event types: non-fatal MI, non-fatal IS, and fatal CVD events (i.e., fatal MI, fatal IS, and other fatal CHD events). Table B-7 shows the final sex-, race/ethnicity-, and age category-specific estimates of the ASCVD event distribution needed as the CVD model input. For males, the share of non-fatal MI events is 22% to 58%, the share of non-fatal IS events is 39% to 77%, and the share of fatal CVD events is 2% to 13%. For females, the share of non-fatal MI events is 16% to 62%, the share of non-fatal IS events is 36% to 76%, and the share of fatal CVD events is 2% to 14%. The shares of non-fatal MI decrease with age, whereas the share of fatal CVD events increase with age. Shares of non-fatal MI are generally highest among non-Hispanic White males, while shares of non-fatal IS are highest among non-Hispanic Black males. Among females, shares of non-fatal IS are highest for non-Hispanic White females aged 18–64 years, with higher shares among non-Hispanic Black females aged 65 or older. Among females aged 65 or older, shares of non-fatal MI are highest in the Hispanic population.

Table B-7: Estimated Distribution of Fatal and Non-Fatal First Hard CVD Events

Sex	Age (years)	Race/Ethnicity	Non-Fatal MI (%)	Non-Fatal IS (%)	Fatal CVD Event (%)
Males	18–44	NH White	58	40	1.5
	45–64	NH White	50	47	3.7
	65–84	NH White	37	57	6.2
	85 or older	NH White	34	53	13
Males	18–44	NH Black	22	77	1.7
	45–64	NH Black	22	75	2.9
	65–84	NH Black	27	66	6.4
	85 or older	NH Black	25	62	13

Table B-7: Estimated Distribution of Fatal and Non-Fatal First Hard CVD Events

Sex	Age (years)	Race/Ethnicity	Non-Fatal MI (%)	Non-Fatal IS (%)	Fatal CVD Event (%)
Males	18–44	Hispanic	56	42	1.5
	45–64	Hispanic	38	59	3
	65–84	Hispanic	50	44	6.1
	85 or older	Hispanic	47	41	12
Males	18–44	NH Other	46	53	1.6
	45–64	NH Other	58	39	3.1
	65–84	NH Other	33	62	5.8
	85 or older	NH Other	30	58	12
Females	18–44	NH White	29	69	1.9
	45–64	NH White	24	71	4.6
	65–84	NH White	26	67	6.5
	85 or older	NH White	24	63	13
Females	18–44	NH Black	62	36	1.7
	45–64	NH Black	26	70	3.9
	65–84	NH Black	18	76	6.7
	85 or older	NH Black	16	70	14
Females	18–44	Hispanic	32	66	1.9
	45–64	Hispanic	31	65	3.8
	65–84	Hispanic	40	54	6.4
	85 or older	Hispanic	37	51	12
Females	18–44	NH Other	45	53	1.8
	45–64	NH Other	32	64	3.6
	65–84	NH Other	28	66	6.5
	85 or older	NH Other	26	61	13

Abbreviations: Fatal CVD – includes fatal MI, fatal IS, and fatal other coronary heart disease events; IS – ischemic stroke; MI – myocardial infarction; NH – non-Hispanic

B.2.4 Post-Acute CVD Mortality

Persons who have experienced non-fatal MI and non-fatal IS events have elevated post-acute CVD mortality and morbidity (Roger et al., 2012). EPA identified four studies that examined risk factors for secondary hard CVD events. These studies differ in terms of outcomes tracked (e.g., recurrent MI, recurrent IS, angina, heart failure, CVD, and all-cause death), conditioning event definition (e.g., MI, IS, CHD), and the length of follow-up for which statistics are reported (e.g., 1-year follow-up, 5-year follow-up). The data used to estimate the risks of secondary CVD events differ with respect to average age, sex, and share of individuals who are White among the participants:

- Data used in Kannel et al. (1999) and D’Agostino et al. (2000) come from the Framingham Heart Survey (Mahmood et al., 2014) and represent White males and females approximately age 60.

- Data used in Thom et al. (2001) are from the pooled Atherosclerosis Risk in Communities Study (Williams, 1989), Cardiovascular Health Study (Fried et al., 1991), and Framingham Original and Offspring Cohort Study (Mahmood et al., 2014). This pooled dataset offers representation for Black males and females, in addition to White males and females, and captures persons aged 45 or older.
- Beatty et al. (2015) used two predominantly White male datasets developed based on the Heart and Soul Study (Whooley et al., 2008) and the PEACE trial (PEACE Trial Investigators, 2004), capturing persons aged 67 years and 64 years, on average, respectively.
- S. Li et al. (2019) used data for 2008 and 2012 and two types of conditioning events (i.e., MI and IS) to assess the risk of secondary events in four large Medicare cohorts: survivors of the first MI in 2008, survivors of the first IS in 2008, survivors of the first MI in 2012, and survivors of the first IS in 2012.¹⁵ These data represent older populations (age 80, on average) and are not limited to a particular race/ethnicity or sex.

Of the studies that assessed risk factors for secondary hard CVD events, only three focused on developing a risk prediction model (Beatty et al., 2015; D'Agostino et al., 2000; Kannel et al., 1999) and only two have changes in cholesterol levels as a primary predictor (Beatty et al., 2015; D'Agostino et al., 2000). Furthermore, in these two studies, TC level does not appear to significantly increase the risk of recurrent CVD events. Beatty et al. (2015) concluded that precautionary measures and medication taken by patients who had suffered from a primary CVD event may decrease the initial risk factors (i.e., TC) and may be a reason for the lack of correlation between secondary CVD events and TC.

In sum, studies focusing on secondary CVD events point to an elevated risk of these events among survivors of the first hard CVD event, but do not support the link between these risks and TC levels. Therefore, the CVD model relies on the same secondary hard CVD event rates to estimate secondary hard CVD event incidence under baseline and treatment scenarios. Specifically, EPA focuses on post-acute CVD mortality as the secondary event of interest, because other non-fatal secondary CVD events are captured in the available unit values for first non-fatal MI and IS (see, e.g., O'Sullivan et al., 2011). EPA selected estimates in Thom et al. (2001) to model post-acute CVD mortality for survivors of MI or IS at ages 40–65, because Thom et al. (2001) is the only study that analyzed this age group. EPA selected estimates in S. Li et al. (2019) to model post-acute CVD mortality for survivors of MI or IS at ages 66–89, because cohorts analyzed in S. Li et al. (2019) are the largest and most representative of the U.S. population compared with the cohorts analyzed by other studies.

B.2.4.1 Survivors of the first hard CVD event at ages 40–65

EPA used estimates of all-cause post-acute mortality for MI survivors at the 1- and 5-year follow-ups from Thom et al. (2001) to model post-acute CVD mortality for survivors of non-fatal MI and non-fatal IS events at ages 45–65. While EPA was unable to identify comparable post-acute mortality statistics for non-fatal IS, an analysis of the Medicare population by S. Li et al.

¹⁵ Note that relative to other studies with sample sizes of, at most, 10,000, the sizes of these cohorts are 20,000, on average.

(2019) suggests that post-acute MI mortality is a reasonable approximation for post-acute IS mortality.¹⁶

Table B-8 shows estimated all-cause probability of death following first non-fatal MI by age category, race/ethnicity, and sex from Thom et al. (2001), as reported in Roger et al. (2012). These estimates are based on the analysis of pooled data from the Atherosclerosis Risk in Communities Study (Williams, 1989), the Cardiovascular Health Study (Fried et al., 1991), and the Framingham Original and Offspring Cohort Study (Mahmood et al., 2014). The estimates are available only for non-Hispanic Whites and non-Hispanic Blacks.

Table B-8: Post-Acute All-Cause Mortality After the First Myocardial Infarction				
Age Group (years)	Race/Ethnicity	Follow-up Period (years)	Probability of All-Cause Death (%)	
			Males	Females
45–64	Non-Hispanic White	1	5	9
45–64	Non-Hispanic Black	1	14	8
65 or older	Non-Hispanic White	1	25	30
65 or older	Non-Hispanic Black	1	25	30
45–64	Non-Hispanic White	5	11	18
45–64	Non-Hispanic Black	5	22	28
65 or older	Non-Hispanic White	5	46	53
65 or older	Non-Hispanic Black	5	54	58

Abbreviations: MI – myocardial infarction (ICD9=410; ICD10=I21)

Source: Thom et al. (2001)

Table B-9 shows estimated probabilities of post-acute CVD mortality after the first MI. EPA derived these probabilities by adjusting all-cause post-acute mortality probabilities reported in Table B-8 for the ages 45–64 group¹⁷ to exclude the probability of death from non-CVD causes. Appendix D provides details on an estimation of integer age-, race/ethnicity- and sex-specific probability of death from non-CVD causes based on the U.S. Life Tables, 2017 (Arias et al., 2019) and CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020). The last two columns of Table B-9 show annual race/ethnicity- and sex-specific post-acute CVD death probabilities used by the CVD model in estimation of secondary mortality in years 1–5 following the first non-fatal MI or IS that occurred at ages 45–65. EPA used post-acute mortality data for non-Hispanic Whites to estimate mortality effects for the other race/ethnicity groups.

¹⁶ For those aged 65 or older, S. Li et al. (2019) have estimated the probability of death within 1 year after a non-fatal IS to be 32.07% and the probability of death within 1 year after a non-fatal MI to be 32.09%.

¹⁷ EPA applies post-acute mortality probabilities estimated for ages 45–64 to the survivors of first MI or IS, ages 45–65, because the magnitude of the annual death probability at age 65 is closer to the average annual death probability for ages 45–64 than to the average annual death probability for ages 66–99.

Table B-9: Post-Acute Mortality After the First Myocardial Infarction

Integer Year Since First MI ^a	All-Cause Death Probability (%)		Non-CVD Death Probability (%) ^b		CVD Death Probability (%) ^c	
	Males	Females	Males	Females	Males	Females
All Races/Ethnicities^d						
0	5.6	8.8	0.56	0.38	5.0	8.4
1	1.5	2.7	0.60	0.41	0.93	2.3
2	1.5	2.7	0.65	0.44	0.88	2.3
3	1.5	2.7	0.70	0.48	0.83	2.3
4	1.5	2.7	0.75	0.51	0.78	2.2
Non-Hispanic White^e						
0	5.0	9.0	–	–	4.5	8.6
1	1.5	2.3	–	–	0.91	1.9
2	1.5	2.3	–	–	0.86	1.9
3	1.5	2.3	–	–	0.82	1.9
4	1.5	2.3	–	–	0.76	1.8
Non-Hispanic Black						
0	14	8.0	–	–	12	7.7
1	2.0	5.0	–	–	1.2	4.3
2	2.0	5.0	–	–	1.1	4.2
3	2.0	5.0	–	–	1.1	4.1
4	2.0	5.0	–	–	1.0	4.1

Notes:

^a Post-acute death probabilities at 1- and 5-year follow-ups in Table B-8 are converted to the integer year-specific post-acute death probabilities by assuming that the annual death probabilities in years 1–4 are identical. This assumption is supported by data in S. Li et al. (2019), who report post-acute death probabilities at 1-, 2-, 3-, 4-, 5-, and 6-year follow-ups.

^b Reported annual probability of non-CVD death is a weighted average of life table age-specific probabilities for ages 45–64. The weights are the sex-specific age distribution of the first MI survivor population, estimated using MEPS 2010–2017 data.

^c For all race/ethnicity categories, CVD death probability is the difference between all-cause death probability and non-CVD death probability. For the non-Hispanic White and non-Hispanic Black race/ethnicity categories, EPA obtained the estimates by multiplying the corresponding all-cause post-acute death probability with the all-race/ethnicity ratio of post-acute CVD death probability to all-cause post-acute death probability.

^d Race/Ethnicity-specific data for the ages 45–64 group in Table B-8 are pooled using a sex-specific race/ethnicity distribution of the first MI survivor population, estimated using MEPS 2010–2017 data.

^e Post-acute CVD death probability for non-Hispanic Whites is used to estimate mortality effects for the other race/ethnicity groups.

Abbreviations: CVD – cardiovascular disease, MEPS – Medical Expenditure Panel Survey, MI – myocardial infarction (ICD9=410; ICD10=I21)

Sources: Thom et al. (2001); U.S. Life Tables, 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020)

B.2.4.2 Survivors of the first hard CVD event at ages 66–89

EPA used the results in S. Li et al. (2019) to estimate the number of post-acute CVD deaths for survivors of the first MI and IS events, aged 66 years or older at the time of the initial event. Table B-10 summarizes the key results in S. Li et al. (2019) that are used to parameterize the CVD model and the results of adjustments that EPA made to incorporate CVD mortality information in the model. First, EPA estimated CVD death probabilities by subtracting non-CVD death probabilities from all-cause post-acute mortality probabilities reported in S. Li et al. (2019). Appendix D describes the derivation of the sex- and age-specific non-CVD mortality rates from U.S. Life Tables, 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020); and U.S. Life Tables Eliminating Certain Causes of Death, 1999–2000 (Arias et al., 2013). EPA has averaged age- and sex-specific non-CVD death probabilities for those age 66 or older using the demographic characteristics of the MI and IS cohorts analyzed by S. Li et al. (2019). Second, EPA calculated CVD mortality probability as the difference between the all-cause death probability and the non-CVD death probability. Third, EPA calculated CVD mortality rate multipliers as a ratio of CVD mortality probability to the non-CVD death probability. EPA combined these multipliers (reported in Table B-10 for MI and IS survivors) with age-, sex-, and race/ethnicity-specific non-CVD death rates to obtain post-acute CVD mortality rates for each cohort included in the analysis.

Table B-10: Post-Acute CVD Mortality Following the First Myocardial Infarction and First Ischemic Stroke in the Population Aged 66 Years or Older

Follow-up Period (years)	MI Survivors				IS Survivors			
	All-Cause Death Probability (%) ^a	Non-CVD Death Probability (%) ^b	CVD Death Probability (%) ^c	CVD Mortality Rate Multiplier ^d	All-Cause Death Probability (%) ^a	Non-CVD Death Probability (%) ^b	CVD Death Probability (%) ^c	CVD Mortality Rate Multiplier ^d
0	32	4.3	27	6.4	32	4.5	28	6.1
1	16	4.6	11	2.5	15	4.8	9.9	2.07
2	15	4.9	9.6	1.9	16	5.2	10	2.1
3	14	5.2	9.04	1.7	15	5.5	9.8	1.8
4	14	5.6	8.6	1.5	15	5.9	8.9	1.5
5	14	5.9	8.04	1.4	14	6.2	8.03	1.3

Notes:

^a For MI, the follow-up year specific all-cause death probability is from S. Li et al. (2019) reported data for the 2008 MI survivor cohort (N=26,46). For IS, the follow-up year specific all-cause death probability is from S. Li et al. (2019) reported data for the 2008 IS survivor cohort (N=17,566).

^b Non-CVD annual mortality rate is based on U.S. Life Tables 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020); and U.S. Life Tables Eliminating Certain Causes of Death, 1999–2000 (Arias et al., 2013) for those age 66 or older. See Appendix D for details. The annual age- and sex-specific death probabilities were averaged using S. Li et al. (2019) MI/IS survivor cohort demographic characteristics.

^c Post-acute CVD death probability rate is estimated by subtracting the non-CVD annual death probability from the all-cause post-acute death probability.

^d The CVD mortality rate multiplier is defined as the difference between all-cause death probability and non-CVD death probability divided by the non-CVD death probability. The CVD model combines the baseline rate multiplier with race/ethnicity-, age-, and sex-specific non-CVD baseline death rates to obtain mortality rates that are appropriate for the race/ethnicity, age, and sex of each cohort included in the analysis.

Abbreviations: CVD – cardiovascular disease, IS – ischemic stroke (ICD9=433, 434; ICD10=I63), MI – myocardial infarction (ICD9=410; ICD10=I21)

Sources: Li et al. (2019); U.S. Life Tables, 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020); U.S. Life Tables Eliminating Certain Causes of Death, 1999–2000 (Arias et al., 2013).

B.3 Detailed CVD Model Calculations

Table B-11 provides a guide to sections containing the recurrent CVD model calculations applicable under conditions defined by initial cohort age, current cohort age, and estimation type. Estimation types include baseline estimation, treatment scenario estimation, and risk reduction estimation. Note that standard life table calculations for current cohort ages 0–39 in section B.3.1 apply to both the baseline and treatment scenario estimation types. The CVD risk reduction estimation equations in section B.3.5 apply to ages 40–99, for which the model explicitly estimates the number of first hard CVD events and the number of post-acute CVD deaths for survivors of the first hard CVD event.

Table B-11: A Mapping of CVD Model Calculations by Initial Cohort Age, Current Cohort Age, and Estimation Type

Initial Cohort Age (years)	Current Cohort Age (years)			
	0–39	40–65	66–89	90–94
Baseline Estimation				
0–39	Section B.3.1	Section B.3.2, Section B.3.4	Section B.3.2, Section B.3.4	Section B.3.4
40–80	–	Section B.3.2, Section B.3.4	Section B.3.2, Section B.3.4	Section B.3.4
81–99	–	Section B.3.2, Section B.3.4	Section B.3.2, Section B.3.4	Section B.3.4
Treatment Scenario Estimation				
0–39	Section B.3.1	Section B.3.3, Section B.3.4	Section B.3.3, Section B.3.4	Section B.3.4
40–80	–	Section B.3.3, Section B.3.4	Section B.3.3, Section B.3.4	Section B.3.4
81–99	–	Section B.3.3, Section B.3.4	Section B.3.3, Section B.3.4	Section B.3.4
Risk Reduction Estimation				
0–39	–	Section B.3.5	Section B.3.5	Section B.3.5
40–80	–	Section B.3.5	Section B.3.5	Section B.3.5
81–99	–	Section B.3.5	Section B.3.5	Section B.3.5

B.3.1 Baseline Recurrent Calculations Without Explicit Treatment of the CVD Population

The number of deaths occurring in year t is estimated using the number of persons alive at the start of the year, $l_{b,a,s,r,t}$, and all-cause annual probability of death, $q_{a,s,r}$:

$$d_{b,a,s,r,t} = q_{a,s,r} \cdot l_{b,a,s,r,t} \quad (\text{Eq. B-3})$$

The number of persons surviving to the start of the next year is calculated as the difference between the number of persons alive at the start of the year, $l_{b,a,s,r,t}$, and the number of deaths estimated to occur during the year, $d_{b,a,s,r,t}$:

$$l_{b,a+1,s,r,t+1} = l_{b,a,s,r,t} - d_{b,a,s,r,t} \quad (\text{Eq. B-4})$$

B.3.2 Baseline Recurrent Calculations with Explicit Treatment of the CVD Population

The population of persons alive at the start of year t , $l_{b,a,s,r,t}$, is split into CVD and non-CVD subpopulations using externally estimated age-, race/ethnicity-, and sex-specific CVD prevalence, $\pi_{a,s,r}$:

$$l_{b,a,s,r,t,CVD} = \pi_{a,s,r} \cdot l_{b,a,s,r,t} \quad (\text{Eq. B-5})$$

$$l_{b,a,s,r,t,OTH} = (1 - \pi_{a,s,r}) \cdot l_{b,a,s,r,t} \quad (\text{Eq. B-6})$$

The year t number of non-CVD deaths in the CVD and non-CVD subpopulations is estimated by applying the annual age-, race/ethnicity-, and sex-specific probability of non-CVD death, $q_{a,s,r,OTH}$, to the number of persons alive at the start of the year in each subpopulation ($l_{b,a,s,r,t,CVD}$ and $l_{b,a,s,r,t,OTH}$), respectively:

$$d_{b,a,s,r,t,CVD,OTH} = q_{a,s,r,OTH} \cdot l_{b,a,s,r,t,CVD} \quad (\text{Eq. B-7})$$

$$d_{b,a,s,r,t,OTH,OTH} = q_{a,s,r,OTH} \cdot l_{b,a,s,r,t,OTH} \quad (\text{Eq. B-8})$$

The year t number of CVD deaths in the CVD subpopulation is estimated by applying the annual CVD death probability, $q_{a,s,r,CVD}$, to the total population alive at the start of the year, $l_{b,a,s,r,t}$, net of deaths from other causes, $q_{a,s,r,OTH}$, estimated to occur during the year:

$$d_{b,a,s,r,t,CVD,CVD} = q_{a,s,r,CVD} \cdot (1 - q_{a,s,r,OTH}) \cdot l_{b,a,s,r,t} \quad (\text{Eq. B-9})$$

The number of persons surviving to the start of the next year is estimated as:

$$l_{b,a+1,s,r,t+1} = l_{b,a,s,r,t} - d_{b,a,s,r,t,CVD,CVD} - d_{b,a,s,r,t,OTH,OTH} - d_{b,a,s,r,t,CVD,OTH} \quad (\text{Eq. B-10})$$

The *uncalibrated* number of persons experiencing their first hard CVD event in year t , $i_{b,a,s,r,t}$, is estimated by applying the baseline annual probability of first hard CVD event to the start-of-the-year number of persons in the non-CVD subpopulation, $l_{b,a,s,r,t,OTH}$, net of non-CVD deaths, $d_{b,a,s,r,t,OTH,OTH}$. Notably, because the ASCVD model applies to ages 40–80 and because it predicts a 10-year probability of the first hard CVD event, EPA applies a constant baseline annual probability of first hard CVD event estimated at age 80 to those currently aged 81–89 years.¹⁸ Finally, EPA uses the externally estimated share of non-fatal first hard CVD events, $\gamma_{a,s,r,f}$, and same-year post-acute CVD mortality probability, $\mu_{a,s,r,f,0}$, to compute the number of persons surviving their first hard type f CVD event in year t :

$$\tilde{n}_{b,a,s,r,f,t,0} = \begin{cases} (1 - \mu_{a,s,r,f,0}) \cdot \gamma_{a,s,r,f} \cdot i_{b,a,s,r,t}(0) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a \leq 80 \\ (1 - \mu_{a,s,r,f,0}) \cdot \gamma_{a,s,r,f} \cdot i_{b,80,s,r,t}(0) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a > 80 \end{cases} \quad (\text{Eq. B-11})$$

¹⁸ Because the ASCVD model applies only to those aged 40 or older, EPA does not estimate the number of first hard CVD events for those currently younger than age 40.

EPA uses the externally estimated share of fatal first hard CVD events, $1 - \sum_{f \in F} \gamma_{a,s,r,f}$, and same-year post-acute CVD mortality probability, $\mu_{a,s,r,f,0}$, to compute the *uncalibrated* number of year t deaths in the incident CVD population at baseline:

$$\begin{aligned} \tilde{m}_{b,a,s,r,t,0} = & \\ & \begin{cases} [1 + \sum_{f \in F} (\mu_{a,s,r,f,0} - 1) \cdot \gamma_{a,s,r,f}] \cdot i_{b,a,s,r,t}(0) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a \leq 80 \\ [1 + \sum_{f \in F} (\mu_{a,s,r,f,0} - 1) \cdot \gamma_{a,s,r,f}] \cdot i_{b,80,s,r,t}(0) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a > 80 \end{cases} \end{aligned} \quad (\text{Eq. B-12})$$

For calibration purposes, EPA calculated the incident CVD population size, $x_{b,a,s,r,t}$, that is consistent with the reported CVD prevalence rates, $\pi_{a,s,r}$, and $\pi_{a+1,s,r}$, and cause-specific mortality rates, $q_{a,s,r,CVD}$ and $q_{a,s,r,OTH}$:

$$x_{b,a,s,r,t} = \pi_{a+1,s,r} l_{b,a+1,s,r,t+1} - l_{b,a,s,r,t,CVD} + d_{b,a,s,r,t,CVD,CVD} + d_{b,a,s,r,t,CVD,OTH} \quad (\text{Eq. B-13})$$

EPA used the incident CVD population size to estimate a calibration factor for scaling raw ASCVD model-based results:

$$\chi_{b,a,s,r,t} = \frac{x_{b,a,s,r,t}}{\sum_{f \in F} \tilde{n}_{b,a,s,r,f,t,0} + \tilde{m}_{b,a,s,r,t,0}} \quad (\text{Eq. B-14})$$

Using the estimated calibration factor, EPA adjusted the raw number of persons surviving their first hard type f CVD event in year t , $\tilde{n}_{b,a,s,r,f,t,0}$, and the raw number of year t deaths in the incident CVD population at baseline, $\tilde{m}_{b,a,s,r,t,0}$, to ensure that EPA does not project a larger number of incident events than is consistent with the CVD prevalence statistics and mortality rates:

$$n_{b,a,s,r,f,t,0} = \min(1, \chi_{b,a,s,r,t}) \cdot \tilde{n}_{b,a,s,r,f,t,0} \quad (\text{Eq. B-15})$$

$$m_{b,a,s,r,t,0} = \min(1, \chi_{b,a,s,r,t}) \cdot \tilde{m}_{b,a,s,r,t,0} \quad (\text{Eq. B-16})$$

Finally, EPA uses the overall number of year t CVD deaths, $d_{b,a,s,r,t,CVD,CVD}$, net of the number of deaths in the incident CVD population, $m_{b,a,s,r,t,0}$, and the size of CVD population alive at the start of the year, $l_{b,a,s,r,t,CVD}$, to estimate the baseline CVD death rate in the prevalent CVD population. This quantity is needed to support treatment scenario estimation:

$$\rho_{b,a,s,r} = (d_{b,a,s,r,t,CVD,CVD} - m_{b,a,s,r,t,0}) / l_{b,a,s,r,t,CVD} \quad (\text{Eq. B-17})$$

B.3.3 Treatment Scenario Recurrent Calculations with Explicit Treatment of the CVD Population

If current cohort age a is equal to the initial cohort age, the sizes of CVD and non-CVD subpopulations at the start of year 0 are calculated using externally estimated CVD prevalence, $\pi_{a,s,r}$, and the initial population size, $l_{b,a,s,r,t}$. If, however, the current cohort age a is greater than the initial cohort age, then the sizes of CVD and non-CVD subpopulations at the start of year t are the same as the end-of-year $t - 1$ CVD and non-CVD subpopulation sizes. That is, the CVD and non-CVD populations are computed in a recurrent manner.

$$l_{b,a,s,r,t,CVD} = \begin{cases} \pi_{a,s,r} \cdot l_{b,a,s,r,t} & \text{if } a = \max(a - t, 40) \\ l_{b,a-1,s,r,t-1,CVD} & \text{if } a > \max(a - t, 40) \end{cases} \quad (\text{Eq. B-18})$$

$$l_{b,a,s,r,t,OTH} = \begin{cases} (1 - \pi_{a,s,r}) \cdot l_{b,a,s,r,t} & \text{if } a = \max(a - t, 40) \\ l_{b,a-1,s,r,t-1,OTH} & \text{if } a > \max(a - t, 40) \end{cases} \quad (\text{Eq. B-19})$$

The year t number of non-CVD deaths in CVD and non-CVD subpopulations is estimated by applying the annual age-, race/ethnicity-, and sex-specific probability of non-CVD death, $q_{a,s,r,OTH}$, to the number of persons alive at the start of the year in each subpopulation, respectively:

$$d_{b,a,s,r,t,CVD,OTH} = q_{a,s,r,OTH} \cdot l_{b,a,s,r,t,CVD} \quad (\text{Eq. B-20})$$

$$d_{b,a,s,r,t,OTH,OTH} = q_{a,s,r,OTH} \cdot l_{b,a,s,r,t,OTH} \quad (\text{Eq. B-21})$$

The uncalibrated number of fatal and non-fatal first hard CVD events under the treatment scenario is estimated using the same equations (i.e., (Eq. B-11 and (Eq. B-12) as the ones used for the baseline scenario, except for the non-zero difference between treatment scenario and baseline total cholesterol $\Delta\tau_{b,a,s,t}$:

$$\tilde{n}_{b,a,s,r,f,t,0} = \begin{cases} (1 - \mu_{a,s,r,f,0}) \cdot \gamma_{a,s,r,f} \cdot i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t}) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a \leq 80 \\ (1 - \mu_{a,s,r,f,0}) \cdot \gamma_{a,s,r,f} \cdot i_{b,80,s,r,t}(\Delta\tau_{b,a,s,t}) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a > 80 \end{cases} \quad (\text{Eq. B-22})$$

$$\begin{aligned} \tilde{m}_{b,a,s,r,t,0} = & \\ & \begin{cases} [1 + \sum_{f \in F} (\mu_{a,s,r,f,0} - 1) \cdot \gamma_{a,s,r,f}] \cdot i_{b,a,s,r,t}(\Delta\tau_{b,a,s,t}) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a \leq 80 \\ [1 + \sum_{f \in F} (\mu_{a,s,r,f,0} - 1) \cdot \gamma_{a,s,r,f}] \cdot i_{b,80,s,r,t}(\Delta\tau_{b,a,s,t}) \cdot (l_{b,a,s,r,t,OTH} - d_{b,a,s,r,t,OTH,OTH}) & \text{if } a > 80 \end{cases} \end{aligned} \quad (\text{Eq. B-23})$$

These estimates are used in combination with the baseline calibration factor, $\chi_{b,a,s,r,t}$, and EPA-estimated treatment scenario incident CVD population size, $x_{b,a,s,r,t}$:

$$x_{b,a,s,r,t} = \chi_{b,a,s,r,t} (\sum_{f \in F} \tilde{n}_{b,a,s,r,f,t,0} + \tilde{m}_{b,a,s,r,t,0}) \quad (\text{Eq. B-24})$$

Using the estimated baseline calibration factor, $\chi_{b,a,s,r,t}$, EPA adjusted the raw number of persons surviving their first hard type f CVD event in year t , $\tilde{n}_{b,a,s,r,f,t,0}$, and the raw number of year t deaths in the incident CVD population, $\tilde{m}_{b,a,s,r,t,0}$:

$$n_{b,a,s,r,f,t,0} = \min(1, \chi_{b,a,s,r,t}) \cdot \tilde{n}_{b,a,s,r,f,t,0} \quad (\text{Eq. B-25})$$

$$m_{b,a,s,r,t,0} = \min(1, \chi_{b,a,s,r,t}) \cdot \tilde{m}_{b,a,s,r,t,0} \quad (\text{Eq. B-26})$$

The number of CVD deaths at age a during year t is estimated as the sum of the number of deaths among those whose CVD event history began before age a , $\rho_{b,a,s,r} \cdot l_{b,a,s,r,t}$, and the number of deaths among those who experienced their first CVD event at age a , $m_{b,a,s,r,t,0}$. The number of deaths among those whose CVD event history began before age a is the product of the baseline CVD death rate in the CVD subpopulation, $\rho_{b,a,s,r}$, and the size of the CVD subpopulation at the start of year t , $l_{b,a,s,r,t}$:

$$d_{b,a,s,r,t,CVD,CVD} = \rho_{b,a,s,r} \cdot l_{b,a,s,r,t} + m_{b,a,s,r,t,0} \quad (\text{Eq. B-27})$$

Finally, the following recurrent equations are used to compute the sizes of total, CVD, and non-CVD populations surviving through to the beginning of year $t + 1$:

$$l_{b,a+1,s,r,t+1} = l_{b,a,s,r,t} - d_{b,a,s,r,t,CVD,CVD} - d_{b,a,s,r,t,OTH,OTH} - d_{b,a,s,r,t,CVD,OTH} \quad (\text{Eq. B-28})$$

$$l_{b,a+1,s,r,t+1,CVD} = l_{b,a,s,r,t,CVD} + x_{b,a,s,r,t} - d_{b,a,s,r,t,CVD,CVD} - d_{b,a,s,r,t,CVD,OTH} \quad (\text{Eq. B-29})$$

$$l_{b,a+1,s,r,t+1,OTH} = l_{b,a,s,r,t,OTH} - x_{b,a,s,r,t} - d_{b,a,s,r,t,OTH,OTH} \quad (\text{Eq. B-30})$$

B.3.4 Recurrent Estimation of Post-Acute CVD Mortality

Survivors of the first type f non-fatal hard CVD event at age a in year t , $n_{b,a,s,r,f,t,0}$, are followed for 5 future years (i.e., $k = 1, 2, 3, 4, 5$) to evaluate post-acute CVD mortality.

EPA estimates the number of post-acute CVD deaths among survivors of a first hard CVD event in year k since the initial event at age a , $m_{b,a+k,s,r,t+k,k}$, by (1) adjusting the number of those who survived $k - 1$ years after the initial event, $n_{b,a+k-1,s,r,f,t+k-1,k-1}$, for non-CVD mortality using externally estimated non-CVD mortality rate, $q_{a+k,s,r,OTH}$; (2) multiplying the result by externally estimated post-acute CVD mortality rate, $\mu_{a+k,s,r,f,k}$; and (3) summing over the first hard CVD event type f :

$$m_{b,a+k,s,r,t+k,k} = \sum_{f \in F} [\mu_{a+k,s,r,f,k} \cdot (1 - q_{a+k,s,r,OTH}) \cdot n_{b,a+k-1,s,r,f,t+k-1,k-1}] \quad (\text{Eq. B-31})$$

EPA estimates the number of survivors of type f first hard CVD event in year k since the initial event at age a , $n_{b,a+k,s,r,f,t+k,k}$, by adjusting the number of those who survived $k - 1$ years after the initial event, $n_{b,a+k-1,s,r,f,t+k-1,k-1}$, for mortality using externally estimated non-CVD mortality rate, $q_{a+k,s,r,OTH}$, and post-acute CVD mortality using rate, $\mu_{a+k,s,r,f,k}$:

$$n_{b,a+k,s,r,f,t+k,k} = (1 - \mu_{a+k,s,r,f,k}) \cdot (1 - q_{a+k,s,r,OTH}) \cdot n_{b,a+k-1,s,r,f,t+k-1,k-1} \quad (\text{Eq. B-32})$$

B.3.5 Risk Reduction Calculations

Assuming that the treatment scenario is associated with a lower incidence of first hard CVD events (via lower total cholesterol levels due to lower serum PFAS), at the end of time period t , the number of avoided type f non-fatal first hard CVD events in the sex s and race/ethnicity r cohort born in year b and currently age a is estimated as:

$$\Delta n_{b,a,s,r,f,t} = n_{b,a,s,r,f,t}^{Treatment\ Scenario} - n_{b,a,s,r,f,t,0}^{Baseline} \quad (\text{Eq. B-33})$$

The number of avoided year t CVD deaths in the first hard CVD population in the sex s and race/ethnicity r cohort born in year b and currently age a years is:

$$\Delta m_{b,a,s,r,t} = \sum_{k=0}^5 (m_{b,a,s,r,t,k}^{Treatment\ Scenario} - m_{b,a,s,r,t,k}^{Baseline}) \quad (\text{Eq. B-34})$$

Total number of avoided type f non-fatal first hard CVD events in year t is:

$$\Delta N_{f,t} = \sum_{a \in A, b \in B} \sum_{s \in S} \sum_{r \in R} \Delta n_{b,a,s,r,f,t} \quad (\text{Eq. B-35})$$

Total number of avoided CVD deaths in the first hard CVD population in year t is:

$$\Delta M_t = \sum_{a \in A, b \in B} \sum_{s \in S} \sum_{r \in R} \Delta m_{b,a,s,r,t} \quad (\text{Eq. B-36})$$

B.4 ASCVD Model Validation

EPA generated life table CVD model results for race/ethnicity subpopulations under different assumptions regarding the applicability of ASCVD coefficients for non-Hispanic Whites and non-Hispanic Blacks to Hispanic and non-Hispanic other subpopulations. CVD model inputs are summarized in Appendix D, Table D-1. The size of each subpopulation cohort was estimated using the 2020 U.S. population size and nationally representative age / sex / race/ethnicity distribution from the American Community Survey, 2017 (U.S. Census Bureau, 2017). EPA evaluated the alignment among age-, sex-, and race/ethnicity-specific CVD incidence prediction using the ASCVD model and age-, sex-, and race/ethnicity-specific CVD incidence prediction calculated by the CVD model on the basis of race-, sex-, and age-specific prevalence of persons with a history of CVD events based on MEPS 2010–2017 (see section B.2.2); U.S. Life Tables, 2017 (Arias et al., 2019); and CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020).

For each race/ethnicity, sex, and age combination, EPA first computed the ratio of CVD incidence based on reported data and incidence based on the ASCVD model. EPA then computed the absolute value of the deviation of this ratio from 1 and averaged the results over age using population weights for each sex and race/ethnicity subpopulation. Table B-12 reports the resulting alignment metrics for each combination of subpopulation and ASCVD model coefficient set. Results show that the ASCVD model coefficients for the non-Hispanic Black model are more consistent with data on CVD prevalence and mortality for Hispanic and non-Hispanic other race subpopulations than the ASCVD model coefficients for the non-Hispanic White model.

Table B-12: Summary of ASCVD Model Validation

Sex	Race/Ethnicity	Alignment of ASCVD Model Predictions With Prevalence and Mortality Statistics*	
		ASCVD Model Coefficients Estimated in Non-Hispanic White Sample	ASCVD Model Coefficients Estimated in Non-Hispanic Black Sample
Males	Non-Hispanic White	0.64	–
	Non-Hispanic Black	–	0.22
	Hispanic	0.44	0.23
	Non-Hispanic Other	0.57	0.18
Females	Non-Hispanic White	2.00	–
	Non-Hispanic Black	–	1.37
	Hispanic	1.53	0.90
	Non-Hispanic Other	1.44	1.07

Note: * Alignment is represented by the population-weighted absolute value of age-specific $|R - 1|$ within each sex and race/ethnicity subpopulation, where R is the race/ethnicity-, age-, and sex-specific ratio of CVD incidence computed from reported data and incidence computed from the ASCVD model.

Appendix C. Affected Population and Population Projection

This appendix describes the data sources used to evaluate the population potentially affected by human health risk reductions due to reductions in drinking water exposure to per- and polyfluoroalkyl substances (PFAS). Table C-1 describes the three main data elements used to assess the affected population, in this illustrative example, the 100,000 people served by the hypothetical PWS. These elements include the Safe Drinking Water Information System (SDWIS) 2020 quarter 4 (Q4) dataset, the Woods & Poole population projections dataset, and EPA's extrapolation of the Woods & Poole population estimates throughout the period of analysis (U.S. Environmental Protection Agency (2020); Woods & Poole Economics Inc., 2021).

The EPA SDWIS dataset provides information reported by states on drinking water systems, as required by the Safe Drinking Water Act. The dataset generally includes information on system name, identification number (public water system [PWS] ID), the cities or counties served, the number of people served, the type of system (community, transient, or non-transient), whether the system operates year-round or seasonally, and characteristics of the system's source water.

The Woods & Poole database includes detailed population data by age, sex, race, and ethnicity from 1990 to 2050. The database authors used regional projection methods that are revised annually to reflect new computational techniques, assumptions, and sources of regional demographic information. The demographic projections stem from a comprehensive historical county database and integrated projection methods. The projection for each county in the United States is performed simultaneously so that changes in one county affect growth or decline in other counties. Future county-level migration patterns for population by age, sex, and race/ethnicity are estimated based on employment opportunities and historical population growth in a particular location. The projection methods assume that working age individuals and their families migrate in response to employment opportunities, whereas migration patterns for populations age 65 and older and for college or military populations are generally based on historical net migration. The Woods & Poole projections do not include 2020 U.S. Census data or the impacts of the COVID-19 pandemic.¹⁹

¹⁹ 2020 Census population data were not released during the first half of 2021.

Table C-1: Summary of Inputs and Data Sources Used to Estimate Affected Population

Data Element	Modeled Variability	Data Source	Notes
Initial Total Population	Location: PWS	SDWIS 2020 (U.S. EPA, 2020)	Public water system inventory from EPA's SDWIS Q4 in 2020. EPA uses the SDWIS 2020 population data as the initial total population per PWS.
Percentage of Population in a Demographic Population Subgroup	Age: integer ages 0–84, 85+ Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, Hispanic, other Location: U.S. counties	Woods & Poole 2021 (Woods & Poole Economics Inc., 2021)	The original data source contains total population by race/ethnicity, sex, and single year age group from 0–84 and 85+ through 2050. EPA extrapolated population projections through 2104 using exponential smoothing techniques.
Annual Infant Population	Age: 0 Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, Hispanic, other Location: U.S. counties	Woods & Poole 2021 (Woods & Poole Economics Inc., 2021)	The original data source contains the total population for infants (age 0) by race/ethnicity and sex through 2050. EPA extrapolated population projections through 2104 using exponential smoothing techniques.

Abbreviations: PWS – public water system, SDWIS – Safe Drinking Water Information System

C.1 SDWIS County FIPS Code Mapping

In order to link population projections from Woods & Poole to PWS locations, EPA developed a crosswalk between SDWIS county names and the county Federal Information Processing Standards (FIPS) codes included in the Woods & Poole database. The SDWIS 2020 Q4 dataset does not include FIPS codes, but rather reports county or city names. The Woods & Poole database includes standard FIPS codes, as well as special county definitions that combine several locations and report a single FIPS identifier for that location (e.g., Fairfax, Fairfax City, and Falls Church City, VA, are combined in a single special county FIPS code: 51919).²⁰

²⁰ Special county definitions are provided in Table 5 of Woods & Poole Economics Inc. (2021) technical documentation.

C.2 Extrapolation of Woods & Poole (2021)

EPA used two methods to extrapolate Woods & Poole population projections from 2051 to 2104:

1. Exponential smoothing via the “forecast” package in R (Hyndman et al., 2021); and
2. Compound annual growth rate.

EPA’s primary method for extrapolating Woods & Poole population projections was method 1. Extrapolations based on method 1 produce population forecasts using exponential smoothing techniques, which are based on weighted averages of past observations (Hyndman et al., 2008). More recent observations (in this case, observations closer to 2050) have a higher associated weight. The model estimation includes three major components: error, trend, and seasonality. Because the Woods & Poole estimates are annual, EPA specified no seasonality but allowed model error and trend types to vary depending on the historical data from 1990 to 2050. Errors can be additive or multiplicative and trends can be additive, multiplicative, additive-damped, or multiplicative-damped. Damped trends introduce a parameter that “dampens” the trend to a relatively flat line in the future. The ultimate functional specification was chosen automatically by the forecast function based on the Akaike information criterion (AIC) of each specification.²¹ EPA relied on method 1 for the majority of the strata-specific population extrapolations.

EPA employed method 2 in circumstances where

- the exponential smoothing function was unable to identify a trend among 1990 to 2050 population estimates from Woods & Poole; and
- the 2104 estimate obtained from exponential smoothing met the following conditions:
 - a. The ratio of the 2104 estimate obtained from exponential smoothing to the 2050 Woods & Poole estimate was greater than the 90th percentile 2104:2050 ratio per race/ethnicity and age group; and
 - i. the magnitude of the compound annual growth rate from the 2050 Woods & Poole estimate to the 2104 estimate obtained from exponential smoothing was greater than five times larger than the compound annual growth rate during 2020 to 2050; or
 - ii. the 2104 estimate obtained from exponential smoothing was more than two orders of magnitude larger than the average of the 2020 and 2050 Woods & Poole Economics Inc. (2021) population estimates.

Method 2 calculates the compound annual growth rate of Woods & Poole estimates from 2041 to 2050 and applies the calculated growth rate to future years. For example, to estimate the 2055 population in a given stratum, EPA multiplied the population estimate for 2050 by the compound annual growth rate raised to the power of 5 (i.e., the number of years between 2050 and 2055). EPA specified a compound annual growth rate cutoff at the 75th percentile value per race, ethnicity, sex, and 10-year age group.

²¹ AIC is a statistical method for comparing different possible models and determining which is the best fit for the data.

C.3 Projecting the 2020 SDWIS Population

Based on the Woods & Poole dataset 2020 population, EPA estimated the proportion of the total SDWIS 2020 population served that falls into each race/ethnicity, sex, and age strata for each county associated with PWSs included in the CVD risk reduction analysis. Based on the extrapolated Woods & Poole dataset, EPA calculated age-, sex-, race/ethnicity-, and county-specific growth rates with respect to the 2020 population for 2021 to 2104. To estimate the strata-specific SDWIS populations per PWS from 2021 to 2104, EPA applied these growth rates to the strata-specific 2020 populations.

C.4 Incorporating Projections into the CVD Model

The CVD model tracks PWS populations from 2023 to 2104 based on the initial SDWIS 2020 Q4 population served. Based on Woods & Poole estimates of the 2020 population, EPA estimated the proportion of total county-level populations that are made up of each race/ethnicity, age, and sex. To estimate the annual number of persons born during 2024 to 2104, EPA applied the age 0 sex-, race/ethnicity-, and county-specific population growth rates obtained from the extrapolated Woods & Poole dataset to the proportioned SDWIS 2020 populations.

Appendix D. CVD Model Inputs

This appendix describes the external data sources used to support the modeling of human health risk reductions due to reductions in drinking water exposure to per- and polyfluoroalkyl substances (PFAS).

D.1 CVD Model Inputs

Table D-1 summarizes the inputs and data sources used in the cardiovascular disease model, including survey health data, model coefficients, Centers for Disease Control and Prevention life tables, hospitalization data, and mortality incidence data.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model			
Data Element	Modeled Variability	Data Source	Notes
Percentage of population with high blood pressure	Age: age groups 40–59, 60+ Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic	NHANES 2011–2014 (Centers for Disease Control and Prevention, 2013a, 2015a, 2015b, 2015c)	EPA used the percentage of population with high blood pressure in 10-year age groups to estimate the number of exposed individuals with high blood pressure who are exposed to PFOA/PFOS in drinking water. EPA applied high blood pressure prevalence data for those aged 40–59 years to both the 40–49 and 50–59 age groups, and data for those age 60+ were applied to both the 60–69 and 70–79 age groups.
Percentage of population receiving blood pressure treatment	Age: age groups 40–59, 60+ Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic	NHANES 2011–2014 (Centers for Disease Control and Prevention, 2013a, 2015a, 2015b, 2015c)	To determine the percentage of the population with controlled high blood pressure, the percentage of the populations per age group and sex who have high blood pressure was multiplied by the percentage of the populations per age group and sex who received treatment for high blood pressure. Treated blood pressure prevalence data for those aged 40–59 years were applied to both the 40–49 and 50–59 age groups, and data for those aged 60+ years were applied to both the 60–69 and 70–79 age groups.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
Treated, untreated, and normal systolic blood pressure measurements	Age: age groups 40–59, 60+ Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic Treatment status: treated, untreated, normal	NHANES 2001–2008 (Centers for Disease Control and Prevention, 2004, 2005, 2007, 2009a, 2009b, 2009c, 2009d, 2011)	EPA applied average systolic blood pressure measurements for those aged 40–59 years to populations aged 40–59 years in the model and applied average systolic blood pressure measurements for those aged 60+ years to populations over the age of 60 in the model.
Baseline total cholesterol level	Age: 10-year age groups (ages 40–79) Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic	NHANES 2011–2016 (Centers for Disease Control and Prevention, 2013c, 2015a, 2015c, 2016b, 2017b, 2017c)	The total cholesterol NHANES datasets from 2011–2016 were combined with corresponding respondent-specific demographic profile datasets to summarize weighted average total cholesterol levels in mg/dL for each age-, sex-, and race-specific stratum.
Baseline high density lipoprotein cholesterol level (HDLc)	Age: 10-year age groups (ages 40–79) Sex: males, females Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic	NHANES 2011–2016 (Centers for Disease Control and Prevention, 2013b, 2015a, 2015c, 2016a, 2017a, 2017c)	The HDLC NHANES datasets from 2011–2016 were combined with corresponding respondent-specific demographic profile datasets to summarize weighted average HDLC levels in mg/dL for each age-, sex-, and race-specific stratum.
Smoking prevalence	Age: age groups 45–64, 65+ Sex: males, females Smoking status: fraction of smokers	NCHS 2018 (Centers for Disease Control and Prevention, 2019)	The percentage of smokers and non-smokers in each stratum were used as inputs in the ASCVD model, providing results similar to using binary variables representing that an individual is either a smoker or a non-smoker and further stratifying the sample. Smoking prevalence data for those aged 45–64 years were applied to the 40–49 and 50–59 age groups, and data for those aged 65+ years were applied to the 60–69 and 70–79 age groups.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
Diabetes prevalence	Age: age groups 45–54, 55–64, 65+ Sex: males, females Diabetes status: fraction of diabetics	NCHS 2017 (Centers for Disease Control and Prevention, 2018)	The percentage of the population with and without diabetes in each stratum were used as inputs in the ASCVD model, providing results similar to using binary variables representing that an individual has or does not have diabetes and further stratifying the sample. Diabetes prevalence data for those aged 45–64 years were applied to the 40–49 age group, data for those aged 55–64 years were applied to the 50–59 age group, and data for those aged 65+ years were applied to the 60–69 and 70–79 age groups.
ASCVD model coefficients	Sex: males, females Race: non-Hispanic White, non-Hispanic Black	Goff et al. (2014), Table A	For modeling purposes, the Hispanic subpopulation was assigned coefficients estimated for the non-Hispanic White subpopulation. The model applies to ages 40–89. ASCVD regressors include age, total cholesterol, HDLC, treated systolic blood pressure, untreated systolic blood pressure, smoking status, and diabetes status.
Annual all-cause death probability	Sex: males, females Age: integer ages 0 ... 100 Race/Ethnicity: all, non-Hispanic White, non-Hispanic Black, Hispanic	U.S. Life Tables, 2017 (Arias et al., 2019)	The quantity used in modeling is q_x (i.e., the probability of dying between ages x and $x + 1$). Life table data for the non-Hispanic other race category are not available; for subsequent modeling, all-race life tables are used for this category.
Annual non-CVD death probability for age 90+	Sex: males, females Age: integer ages 90 ... 100 Race/Ethnicity: all, non-Hispanic White, non-Hispanic Black, Hispanic	U.S. Life Tables, 2017 (Arias et al., 2019); U.S. Life Tables Eliminating Certain Causes of Death, 1999–2000 (Arias et al., 2013)	Annual non-CVD death probability is estimated by multiplying q_x from the 2017 U.S. life tables by the sex-specific ratio of non-CVD q_x to all-cause q_x from 1999–2000 U.S. life tables eliminating certain causes. Life table data for the non-Hispanic other race category are not available; for subsequent modeling, all-race life tables are used for this category. The 1999–2000 U.S. life tables eliminating certain causes are not race/ethnicity-specific; the U.S. general population ratios of non-CVD q_x to all-cause q_x were applied to all race/ethnicity categories. The 1999–2000 U.S. life tables eliminating certain causes are abridged and report 5-year rates. The corresponding 5-year ratios are applied to all individual years within the 5-year range.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
Annual non-CVD death probability for ages 40–89	Sex: males, females Age: integer ages 40 ... 89 Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic	U.S. Life Tables 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020)	Annual non-CVD death probability is estimated by multiplying q_x from 2017 U.S. life tables by the ratio of non-CVD q_x to all-cause q_x . The non-CVD q_x estimate was obtained for each integer age by sex combination as the difference between all-cause q_x from U.S. 2017 life tables and CVD q_x from CDC 1999–2019 cause-specific mortality rates. U.S. 2017 life table data for the non-Hispanic other race category are not available; life tables for the U.S. general population are used for this category.
Cardiovascular disease prevalence	Sex: males, females Age: age groups 18–44, 45–64, 65+ Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic Condition: MI, IS, other CHD, MI + IS + other CHD conditions combined	MEPS 2010–2017 (Agency for Healthcare Research and Quality, 2011, 2012a, 2012b, 2013a, 2013b, 2014a, 2014b, 2015a, 2015b, 2016a, 2016b, 2017b, 2017c, 2018, 2019a, 2019b, 2019c)	MEPS longitudinal files were used to obtain survey weights, design variables, and information on cardiovascular conditions (including age at diagnosis) that began prior to the start date for the survey panel. MEPS medical conditions files were used to obtain information on the newly diagnosed conditions of interest. Specifically, MI events were identified using ICD9=410 or MIDX=1, stroke events were identified using ICD9=433,434,435,436 or STRKDX=1, other CHD were identified using ICD9=413,414,427,428 or CHDDX=1, ANGIDX=1, OHRTDX=1. Cardiovascular disease prevalence was estimated based on persons whose condition started at an age prior to the age at which the MEPS round interview was conducted.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
Cardiovascular disease incidence in the non-CVD population	Sex: males, females Age: age groups 18–44, 45–64, 65+ Race/Ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic other, Hispanic Condition: MI, IS, other CHD	MEPS 2010–2017 (Agency for Healthcare Research and Quality, 2011, 2012a, 2012b, 2013a, 2013b, 2014a, 2014b, 2015a, 2015b, 2016a, 2016b, 2017b, 2017c, 2018, 2019a, 2019b, 2019c)	MEPS longitudinal files were used to obtain survey weights, design variables, and information on cardiovascular conditions (including age at diagnosis) that began prior to the start date for the survey panel. MEPS medical conditions files were used to obtain information on the newly diagnosed conditions of interest. Specifically, MI events were identified using ICD9=410 or MIDX=1, stroke events were identified using ICD9=433,434,435,436 or STRKDX=1, other CHD were identified using ICD9=413,414,427,428 or CHDDX=1, ANGIDX=1, OHRTDX=1. Cardiovascular disease incidence was estimated based on persons whose condition started at an age that was the same as the age at which the MEPS round interview was conducted.
In-hospital death probability for CVD events	Sex: males, females Age: age groups 18–44, 45–64, 65–84, 85+ Condition: MI, IS, other CHD	HCUP 2017 (Agency for Healthcare Research and Quality, 2017a)	Hospital death probabilities were estimated from condition-specific hospitalizations identified using the following ICD10 codes: ICD10=I21 for MI, ICD10=I63 for ischemic stroke, and ICD10=I20, I22–I25 for other CHD. HCUP reports death probabilities separately by sex or within age groups. EPA estimated age group- and sex-specific hospital death probabilities by assuming that male/female relative risk does not vary across age groups.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
1-year, 2-year, 3-year, 4-year, and 5-year all-cause mortality incidence in MI survivors ages 40–64	Sex: males, females Race: all Age: age groups 40–65 Condition: MI	Thom et al. (2001); MI incidence based on the MEPS 2010–2017 analysis, U.S. Life Tables, 2017 (Arias et al., 2019)	Thom et al. (2001) sex- and race-specific estimates for 1-year follow-up and 5-year follow-up all-cause mortality for ages 45–64 MI survivors are as reported in Roger et al. (2012) (the text of the original report is not accessible). Thom et al. (2001) generated separate estimates for non-Hispanic White and non-Hispanic Black persons. To derive sex-specific all-race/ethnicity estimates, EPA used MEPS-based race/ethnicity- and sex-specific MI incidence for ages 45–64 and assumed that non-Hispanic White mortality estimates apply to other race/ethnicity categories. To derive 2-year, 3-year, and 4-year all-cause post-MI mortality incidence, EPA further assumed that the annual probability of death between 1-year follow-up and 5-year follow-up was constant. Finally, EPA assumed that the resulting estimates apply to ages 40–44 MI survivors and age 65 MI survivors.
1-year, 2-year, 3-year, 4-year, 5-year, and 6-year all-cause mortality incidence in MI survivors and IS survivors age 65+	Sex: all Race: all Age: age group 65+ Condition: MI, IS	S. Li et al. (2019)	S. Li et al. (2019) estimates based on 2008 MI and 2008 IS Medicare cohorts (see Figure 1 of the paper) were used. Note that these estimates are neither race- nor sex-specific.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
1-year, 2-year, 3-year, 4-year, and 5-year CVD mortality incidence in MI survivors ages 40–65	Sex: males, females Race: non-Hispanic White, non-Hispanic Black, Age: age groups 40–65 Condition: MI	Thom et al. (2001); MI incidence based on the MEPS 2010–2017 analysis, U.S. Life Tables, 2017 (Arias et al., 2019); CVD death rates 1999–2019 (Centers for Disease Control and Prevention, 2020)	EPA used estimated annual age- and sex-specific non-CVD death probability (estimated as described above) to calculate the probability of non-CVD death within the next 1, 2, 3, 4, and 5 years. These probabilities were averaged over ages 45–64 using MI incidence-based weights estimated from MEPS 2010–2017 (estimated as described above). EPA then subtracted these estimates from 1-, 2-, 3-, 4-, and 5-year sex-specific all-cause mortality incidence in MI survivors aged 45–64 years (estimated as described above) to obtain 1-, 2-, 3-, 4-, and 5-year CVD mortality incidence. Based on this result, EPA estimated the sex-specific ratios of CVD mortality to all-cause mortality in MI survivors 1, 2, 3, 4, and 5 years after the initial event. These ratios were applied to non-Hispanic White and non-Hispanic Black all-cause post-MI mortality reported in Thom et al. (2001) to obtain post-acute CVD mortality estimates for these races. The other race/ethnicity categories used in modeling were assigned post-acute CVD mortality rates for non-Hispanic Whites. Finally, EPA assumed that the resulting estimates applied to ages 40–44 MI survivors and to age 65 MI survivors.

Table D-1: Summary of Inputs and Data Sources Used in the CVD Model

Data Element	Modeled Variability	Data Source	Notes
1-year, 2-year, 3-year, 4-year, 5-year, and 6-year CVD mortality incidence in MI survivors and IS survivors ages 65+	Sex: male, female Race: all Age: ages 66 ... 89 Condition: MI, IS	S. Li et al. (2019); U.S. Life Tables, 2017 (Arias et al., 2019); CVD death rates, 1999–2019 (Centers for Disease Control and Prevention, 2020); U.S. Life Tables Eliminating Certain Causes of Death, 1999–2000 (Arias et al., 2013)	EPA used estimated annual age- and sex-specific non-CVD death probability (estimated as described above) to calculate the probability of non-CVD death within the next 1, 2, 3, 4, 5, and 6 years. These results were averaged using S. Li et al. (2019) 2008 MI/IS cohort age and sex characteristics. In conjunction with all-cause post-MI/IS mortality estimates from S. Li et al. (2019), these estimates were used to estimate the ratio of CVD mortality to the general population non-CVD mortality 1, 2, 3, 4, 5, and 6 years after the initial MI/IS event. The sex- and age-specific probabilities of CVD death 1, 2, 3, 4, 5, and 6 years after the initial MI/IS event were estimated by applying these ratios to sex- and age-specific non-CVD mortality probabilities.

Abbreviations: ASCVD – atherosclerotic cardiovascular disease, CHD – coronary heart disease, CVD – cardiovascular disease, HCUP – Healthcare Cost and Utilization Project, IS – ischemic stroke, MEPS – Medical Expenditure Panel Survey, MI – myocardial infarction, NCHS – National Center for Health Statistics, NHANES – National Health and Nutrition Examination Survey, NHIS – National Health Interview Survey, PFOA – perfluorooctanoic acid, PFOS – perfluorooctanesulfonic acid

Appendix E. CVD Model Uncertainty Analysis Approach

This appendix describes EPA's approach for future characterization of the sources of uncertainty embedded in the estimated impact of changes in serum perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) on the incidence of cardiovascular disease (CVD) events. Section E.1 summarizes the sources of uncertainty in CVD risk reduction modeling and identifies the sources that EPA selected for empirical evaluation. Section 0 describes the uncertainty modeling method for each source and an approach for prioritizing sources for characterization in the full-scale risk reduction analysis.

E.1 Characterization of Uncertainty Sources

Table E-1 describes sources of uncertainty for the following CVD risk reduction model components, including serum PFOA/PFOS, population size, and CVD risk. For each source of uncertainty, EPA notes whether this source is selected for the empirical uncertainty analysis, along with the reasoning for this decision.

Table E-1: Sources of Uncertainty in the CVD Risk Reduction Model			
Model Component	Source of Uncertainty	Will the Source Be Analyzed?*	Rationale for Including or Not Including
Serum PFOA/PFOS	Serum concentrations of PFOS and PFOA: The CVD risk reduction model requires serum concentrations of PFOS and PFOA as inputs. These inputs are derived from concentrations of PFOS and PFOA in drinking water using a PK model.	No	EPA is currently evaluating including the PK model in the uncertainty analysis. EPA may consider using an upper and lower bound for PK model parameters (e.g., half-life).
Population Size and Demographic Composition	PWS-specific served population size and demographic composition: The CVD risk reduction model is designed to be applied to the PWS-specific served population size, which is projected to 2023–2104 to evaluate the full extent of reduced CVD risk. The projections are developed for subpopulations defined by age, sex, and race/ethnicity.	No	PWS-specific population projections for demographic subgroups of interest are based on Woods & Poole Economics Inc. (2021) county-level projections. The analysis assumes that the growth and demographic composition of the PWS-specific population served is the same as that of the county in which it is located. Woods & Poole Economics Inc. (2021) does not provide uncertainty characterization for the projected population size and demographic composition.

Table E-1: Sources of Uncertainty in the CVD Risk Reduction Model

Model Component	Source of Uncertainty	Will the Source Be Analyzed?*	Rationale for Including or Not Including
CVD Risk	TC-serum PFOA slope factor; TC-serum PFOS slope factor: At this time, the slope factors that express the effects of PFOA and PFOS on serum TC markers are based on meta-analyses of five (PFOA) and six (PFOS) studies with high-quality data and clearly defined PFAS-TC level relationships. These studies provide a central estimate and 95% confidence interval values for the slope factors.	Yes	It is possible to characterize the statistical uncertainty of the slope factors. Values are obtained from meta-analysis and include 95% confidence intervals.
	ASCVD model to estimate the risk of first hard CVD event based on TC and other predictors: The ASCVD model coefficients are sex-specific and available for non-Hispanic White and non-Hispanic Black subpopulations (Goff et al., 2014).	Yes	It is possible to characterize the statistical uncertainty of the ASCVD model coefficients. Variance-covariance matrix estimates were obtained from the authors. However, it is not possible to characterize the extrapolation uncertainty that is associated with applying the coefficients estimated for the non-Hispanic Black subpopulation to other Hispanic and non-Hispanic other race subpopulations included in the CVD model.
	Non-TC ASCVD model predictors (HBP prevalence, controlled HBP prevalence, TC, HDLC, systolic blood pressure, smoking prevalence, diabetes prevalence): National-level age-, race/ethnicity- and sex-specific estimates of means for these variables come from NHIS and NHANES U.S. population surveys conducted by CDC.	Yes	It is possible to characterize the sampling uncertainty surrounding the NHIS and NHANES estimates.
	Standard life table: National-level race/ethnicity- and sex-specific life tables are estimates published by CDC.	No	CDC does not characterize the statistical uncertainty surrounding U.S. population life tables.
	Cause-eliminated life table: National-level race/ethnicity- and sex-specific cause-eliminated life tables are estimates published by CDC.	No	CDC does not characterize the statistical uncertainty surrounding U.S. population life tables.

Table E-1: Sources of Uncertainty in the CVD Risk Reduction Model

Model Component	Source of Uncertainty	Will the Source Be Analyzed?*	Rationale for Including or Not Including
	Annual CVD mortality: National-level age-, sex-, and race/ethnicity-specific CVD mortality rates published by CDC.	Yes	CDC reports the standard errors for the mortality rate estimates.
	Prevalence of CVD event history: EPA-estimated national-level age-, sex-, and race/ethnicity-specific prevalence of persons with a CVD event history in the United States using the MEPS data.	Yes	It is possible to characterize the sampling uncertainty surrounding these estimates.
	Distribution of ASCVD events by type: fatal and non-fatal MI, fatal and non-fatal IS, fatal CHD: EPA-estimated age-, sex-, and race/ethnicity-specific distribution of ASCVD events using the MEPS and HCUP data.	Yes	It is possible to partially characterize the sampling uncertainty surrounding these estimates. The sampling uncertainty characterization for the MEPS-based distribution of CVD incidence over MI, IS, and other CHD events is available. However, the HCUP hospital discharge data used to estimate MI, IS, and other CHD event mortality do not contain the uncertainty characterization.
	Post-acute CVD mortality rate: Age-, sex-, and event- (IS or MI) specific post-acute CVD event mortality rate estimates are based on a combination of two studies, CDC CVD mortality rates, and CDC life table data.	No	With the exception of CDC's CVD mortality rates, uncertainty information is not available for the sources contributing to the post-acute CVD mortality estimate.

Abbreviations: ASCVD – atherosclerotic cardiovascular disease, CDC – Centers for Disease Control and Prevention, CHD – coronary heart disease, CVD – cardiovascular disease, HBP – high blood pressure, HCUP – Healthcare Cost and Utilization Project, HDLC – high-density lipoprotein cholesterol, IS – ischemic stroke, MEPS – Medical Expenditure Panel Survey, MI – myocardial infarction, NHANES – National Health and Nutrition Examination Survey, NHIS – National Health Interview Survey, PK – pharmacokinetic modeling, PWS – public water system, TC – total cholesterol

Note: * Selection of the source for further analysis was made upon consultation with EPA on March 31, 2021.

E.2 Evaluation of Uncertainty Source Contributions

Table E-2 provides details on uncertainty and variability characterization for each source of uncertainty selected for the evaluation. EPA will use sampling-based uncertainty characterization methods, such as Monte Carlo or Latin hypercube sampling, to integrate these sources of uncertainty and to express the uncertainty in the outputs of interest (U.S. EPA,

2011). The outputs of interest for this analysis are avoided cases of mortality and morbidity. EPA proposes to carry out uncertainty analysis for one or more public water systems (PWS) to ensure that robust conclusions about the most significant contributors to output uncertainty are drawn.

Table E-2: Source-Specific Uncertainty Characterization

Model Component	Source of Uncertainty	Uncertainty and Variability Characterization
CVD Impacts	TC-serum PFOA slope factor; TC-serum PFOS slope factor	Uncertainty: Uniform distribution within the estimated 95% confidence interval for slope factors. Details are provided in Appendix A. Variability: None
	ASCVD model to estimate the risk of first hard CVD event based on TC and other predictors	Uncertainty: Multivariate normal distribution for coefficients reported in the ASCVD model report (Goff et al., 2014) and variance-covariance matrices obtained from the authors. Variability: Sex, race/ethnicity
	Non-TC ASCVD model predictors: HBP prevalence, controlled HBP prevalence, TC, HDLC, systolic blood pressure	Uncertainty: Normal distribution using mean and standard error estimates for the parameters based on NHANES. Variability: Age, sex, race/ethnicity
	Non-TC ASCVD model predictors: smoking prevalence, diabetes prevalence	Uncertainty: Normal distribution using mean and standard error estimates for parameters based on NHIS. Variability: Age, sex
	Annual CVD mortality	Uncertainty: Normal distribution using mean and standard error estimates for parameters reported by CDC. Variability: Age, sex, race/ethnicity
	Prevalence of CVD event history	Uncertainty: Normal distribution using mean and standard error estimates for the parameters based on MEPS. Variability: Age, sex, race/ethnicity
	Distribution of ASCVD events by type: fatal and non-fatal MI, fatal and non-fatal IS, fatal CHD	Uncertainty: Normal distribution using mean and standard error estimates for the parameters based on MEPS and HCUP.* Variability: Age, sex, race/ethnicity

Abbreviations: ASCVD – atherosclerotic cardiovascular disease, CDC – Centers for Disease Control and Prevention, CHD – coronary heart disease, CVD – cardiovascular disease, HBP – high blood pressure, HCUP – Healthcare Cost and Utilization Project, HDLC – high-density lipoprotein cholesterol, IS – ischemic stroke, MEPS – Medical Expenditure Panel Survey, MI – myocardial infarction, NHANES – National Health and Nutrition Examination Survey, NHIS – National Health Interview Survey, TC – total cholesterol

Note: * Starting in April 2021, HCUP data provider, the Agency for Healthcare Research and Quality, no longer provides information on hospital mortality at the ICD10 level. For this information, EPA relies on 2017 HCUP data extracts obtained in March 2021.

To understand the drivers of uncertainty in the CVD risk reduction modeling output, EPA will compute standardized rank regression coefficients (SRRCs). SRRCs are coefficients from a multiple regression of the output value rank on the set of the input value ranks, which permits comparisons of input importance when inputs are combined in a nonlinear manner (Saltelli et al., 2009). Specifically, using Monte Carlo simulation results for a PWS EPA will estimate the following output-specific regression models:

$$Rank(O_{iabsr}) = \sum_{u \in U} \beta_u \cdot Rank(S_{uiabsr}) + \alpha_{ab} + \theta_s + \gamma_r + \varepsilon_{iabsr} \quad (\text{Eq. E-1})$$

where:

O_{iabsr} – output value in simulation draw i , for a cohort age a , birth year b , sex s , and race/ethnicity r . The outputs of interest are cases avoided over the analysis period;

S_{uiabsr} – input value for the uncertainty source u in simulation draw i , for a cohort age a , birth year b , sex s , and race/ethnicity r . The uncertain inputs are listed in Table E-2;

β_u – the marginal effect of a change in value rank for the uncertainty source u on the output value rank;

α_{ab} – fixed effect of cohort age a / birth year b combination;

θ_s – fixed effect of sex s ;

γ_r – fixed effect of race/ethnicity r ; and

ε_{iabsr} – residual error.

The SRRC coefficients for each uncertainty source are estimated using the equation below:

$$SRRC_u = \frac{\hat{\beta}_u}{s.e.(\beta_u)} \quad (\text{Eq. E-2})$$

where:

$SRRC_u$ – an estimate of the standardized rank regression coefficient for uncertainty source u ;

$\hat{\beta}_u$ – an estimate of the regression coefficient for uncertainty source u based on regression $Rank(O_{iabsr}) = \sum_{u \in U} \beta_u \cdot Rank(S_{uiabsr}) + \alpha_{ab} + \theta_s + \gamma_r + \varepsilon_{iabsr}$ (Eq. E-1); and

$s.e.(\beta_u)$ – an estimate of the standard error for the uncertainty source u regression coefficient in $Rank(O_{iabsr}) = \sum_{u \in U} \beta_u \cdot Rank(S_{uiabsr}) + \alpha_{ab} + \theta_s + \gamma_r + \varepsilon_{iabsr}$ (Eq. E-1).

Appendix F. Appendix References

- Agency for Healthcare Research and Quality. (2011). *MEPS HC-128: 2009 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-128
- Agency for Healthcare Research and Quality. (2012a). *MEPS HC-137: 2010 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-137
- Agency for Healthcare Research and Quality. (2012b). *MEPS HC-139: Panel 14 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-139
- Agency for Healthcare Research and Quality. (2013a). *MEPS HC-146: 2011 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-146
- Agency for Healthcare Research and Quality. (2013b). *MEPS HC-148: Panel 15 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-148
- Agency for Healthcare Research and Quality. (2014a). *MEPS HC-154: 2012 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-154
- Agency for Healthcare Research and Quality. (2014b). *MEPS HC-156: Panel 16 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-156
- Agency for Healthcare Research and Quality. (2015a). *MEPS HC-162: 2013 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-162
- Agency for Healthcare Research and Quality. (2015b). *MEPS HC-164: Panel 17 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-164
- Agency for Healthcare Research and Quality. (2016a). *MEPS HC-170: 2014 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-170
- Agency for Healthcare Research and Quality. (2016b). *MEPS HC-172: Panel 18 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from:
https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-172

- Agency for Healthcare Research and Quality. (2017a). *Healthcare Cost and Utilization Project*. Retrieved from: <https://hcupnet.ahrq.gov>
- Agency for Healthcare Research and Quality. (2017b). *MEPS HC-180: 2015 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-180
- Agency for Healthcare Research and Quality. (2017c). *MEPS HC-183: Panel 19 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-183
- Agency for Healthcare Research and Quality. (2018). *MEPS HC-193: Panel 20 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-193
- Agency for Healthcare Research and Quality. (2019a). *MEPS HC-190: 2016 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-190
- Agency for Healthcare Research and Quality. (2019b). *MEPS HC-199: 2017 Medical Conditions File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-199
- Agency for Healthcare Research and Quality. (2019c). *MEPS HC-202: Panel 21 Longitudinal Data File. Medical Expenditure Panel Survey*. Retrieved from: https://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-202
- Agency for Toxic Substances and Disease Registry. (2018). *Toxicological profile for Perfluoroalkyls*.
- Arias, E., Heron, M., & Tejada-Vera, B. (2013). United States life tables eliminating certain causes of death, 1999-2001.
- Arias, E., Heron, M., & Xu, J. (2019). *United States Life Tables, 2017*.
- Beatty, A. L., Ku, I. A., Bibbins-Domingo, K., Christenson, R. H., DeFillippi, C. R., Ganz, P., . . . Whooley, M. A. (2015). Traditional risk factors versus biomarkers for prediction of secondary events in patients with stable Coronary Heart Disease: from the Heart and Soul Study. *Journal of the American Heart Association*. doi:10.1161/JAHA.114.001646
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50(4), 1088-1101. doi:<https://doi.org/10.2307/2533446>
- Centers for Disease Control and Prevention. (2004). *2001-2002 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_B). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2001-2002/BPX_B.XPT
- Centers for Disease Control and Prevention. (2005). *2003-2004 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_C). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/BPX_C.XPT

- Centers for Disease Control and Prevention. (2007). *2005-2006 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_D). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/BPX_D.XPT
- Centers for Disease Control and Prevention. (2009a). *2001-2002 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_B). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2001-2002/DEMO_B.XPT
- Centers for Disease Control and Prevention. (2009b). *2003-2004 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_C). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DEMO_C.XPT
- Centers for Disease Control and Prevention. (2009c). *2005-2006 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_D). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/DEMO_D.XPT
- Centers for Disease Control and Prevention. (2009d). *2007-2008 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_E). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/DEMO_E.XPT
- Centers for Disease Control and Prevention. (2011). *2007-2008 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_E). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/BPX_E.XPT
- Centers for Disease Control and Prevention. (2013a). *2011-2012 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_G). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BMX_G.XPT
- Centers for Disease Control and Prevention. (2013b). *2011-2012 Data Documentation, Codebook, and Frequencies. Cholesterol - HDL (HDL_G). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/HDL_G.XPT
- Centers for Disease Control and Prevention. (2013c). *2011-2012 Data Documentation, Codebook, and Frequencies. Cholesterol - Total (TCHOL_G). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/TCHOL_G.XPT
- Centers for Disease Control and Prevention. (2015a). *2011-2012 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_G). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/DEMO_G.XPT
- Centers for Disease Control and Prevention. (2015b). *2013-2014 Data Documentation, Codebook, and Frequencies. Blood Pressure (BPX_H). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/BPX_H.XPT

- Centers for Disease Control and Prevention. (2015c). *2013-2014 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_H). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/DEMO_H.XPT
- Centers for Disease Control and Prevention. (2016a). *2013-2014 Data Documentation, Codebook, and Frequencies. Cholesterol - HDL (HDL_H). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/HDL_H.XPT
- Centers for Disease Control and Prevention. (2016b). *2013-2014 Data Documentation, Codebook, and Frequencies. Cholesterol - Total (TCHOL_H). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/TCHOL_H.XPT
- Centers for Disease Control and Prevention. (2017a). *2015-2016 Data Documentation, Codebook, and Frequencies. Cholesterol - HDL (HDL_I). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/HDL_I.XPT
- Centers for Disease Control and Prevention. (2017b). *2015-2016 Data Documentation, Codebook, and Frequencies. Cholesterol - Total (TCHOL_I). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/TCHOL_I.XPT
- Centers for Disease Control and Prevention. (2017c). *2015-2016 Data Documentation, Codebook, and Frequencies. Demographic Variables & Sample Weights (DEMO_I). National Health and Nutrition Examination Survey*. Retrieved from: https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DEMO_I.XPT
- Centers for Disease Control and Prevention. (2018). *2017 Sample Adult File. National Health Interview Survey*. Retrieved from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2017/samadultcsv.zip
- Centers for Disease Control and Prevention. (2019). *2018 Sample Adult File. National Health Interview Survey*. Retrieved from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2018/samadultcsv.zip
- Centers for Disease Control and Prevention. (2020). *Underlying Cause of Death, 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program*. Retrieved from: <http://wonder.cdc.gov/ucd-icd10.html>
- Chang, H. Y., Nuyten, D. S. A., Sneddon, J. B., Hastie, T., Tibshirani, R., Sørli, T., . . . van de Vijver, M. J. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl. Acad. Sci. U. S. A.*, 102(10), 3738-3743. doi:10.1073/pnas.0409462102
- Château-Degat, M.-L., Pereg, D., Dallaire, R., Ayotte, P., Dery, S., & Dewailly, É. (2010). Effects of perfluorooctanesulfonate exposure on plasma lipid levels in the Inuit population of Nunavik (Northern Quebec). *Environmental research*, 110(7), 710-717.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101-129. doi:<https://doi.org/10.2307/3001666>

- Convertino, M., Church, T. R., Olsen, G. W., Liu, Y., Doyle, E., Elcombe, C. R., . . . Evans, T. R. J. (2018). Stochastic Pharmacokinetic-Pharmacodynamic Modeling for Assessing the Systemic Health Risk of Perfluorooctanoate (PFOA). *Toxicological Sciences*, 163(1), 293-306. doi:10.1093/toxsci/kfy035
- D'Agostino, R. B., Grundy, S., Sullivan, L. M., Wilson, P., & Group, C. H. D. R. P. (2001). Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA*, 286(2), 180-187. doi:10.1001/jama.286.2.180
- D'Agostino, R. B., Russell, M. W., Huse, D. M., Ellison, R. C., Silbershatz, H., Wilson, P. W., & Hartz, S. C. (2000). Primary and subsequent coronary risk appraisal: new results from the Framingham study. *Am. Heart J.*, 139(2 Pt 1), 272-281. doi:10.1067/mhj.2000.96469
- D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care. *Circulation*, 117(6), 743-753.
- D'Agostino, R. B., Wolf, P. A., Belanger, A. J., & Kannel, W. B. (1994). Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke*, 25(1), 40-43. doi:10.1161/01.str.25.1.40
- Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in medicine*, 21(11), 1575-1600.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3), 177-188.
- Dong, Z., Wang, H., Yu, Y. Y., Li, Y. B., Naidu, R., & Liu, Y. (2019). Using 2003–2014 US NHANES data to determine the associations between per-and polyfluoroalkyl substances and cholesterol: Trend and implications. *Ecotoxicology and environmental safety*, 173, 461-468.
- Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*, 95(449), 89-98. doi:10.1080/01621459.2000.10473905
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. doi:10.1111/j.0006-341X.2000.00455.x
- Dzierlenga, M. W., Crawford, L., & Longnecker, M. P. (2020). Birth weight and perfluorooctane sulfonic acid: a random-effects meta-regression analysis. *Environmental Epidemiology*, 4(3).
- Egger, M., Smith, G., Schneider, M., & Minder, C. (Eds.). (2008). *Systematic reviews in health care: Meta-analysis in context*. London, UK: BMJ Publishing Group.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629-634. doi:<https://doi.org/10.1136/bmj.315.7109.629>
- Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I., & Lau, J. (2000). Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in medicine*, 19(13), 1707-1728.

- Eriksen, K. T., Raaschou-Nielsen, O., McLaughlin, J. K., Lipworth, L., Tjønneland, A., Overvad, K., & Sørensen, M. (2013). Association between plasma PFOA and PFOS levels and total cholesterol in a middle-aged Danish population. *PLoS One*, 8(2), e56969.
- Fan, Y., Li, X., Xu, Q., Zhang, Y., Yang, X., Han, X., . . . Lu, C. (2020). Serum albumin mediates the effect of multiple per-and polyfluoroalkyl substances on serum lipid levels. *Environmental Pollution*, 266, 115138.
- Fisher, M., Arbuckle, T. E., Wade, M., & Haines, D. A. (2013). Do perfluoroalkyl substances affect metabolic function and plasma lipids?—Analysis of the 2007–2009, Canadian Health Measures Survey (CHMS) Cycle 1. *Environmental research*, 121, 95-103.
- Fried, L. P., Borhani, N. O., Enright, P., Furberg, C. D., Gardin, J. M., Kronmal, R. A., . . . Newman, A. (1991). The cardiovascular health study: design and rationale. *Annals of epidemiology*, 1(3), 263-276.
- Friedman, G. D., Cutter, G. R., Donahue, R. P., Hughes, G. H., Hulley, S. B., Jacobs Jr, D. R., . . . Savage, P. J. (1988). CARDIA: study design, recruitment, and some characteristics of the examined subjects. *Journal of clinical epidemiology*, 41(11), 1105-1116.
- Fu, Y., Wang, T., Fu, Q., Wang, P., & Lu, Y. (2014). Associations between serum concentrations of perfluoroalkyl acids and serum lipid levels in a Chinese population. *Ecotoxicology and environmental safety*, 106, 246-252.
- Goff, D. C., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'Agostino, R. B., Gibbons, R., . . . Others. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Circulation*, 129(25 suppl 2), 49-73.
- He, X., Liu, Y., Xu, B., Gu, L., & Tang, W. (2018). PFOA is associated with diabetes and metabolic alteration in US men: National Health and Nutrition Examination Survey 2003–2012. *Science of the Total Environment*, 625, 566-574.
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*: John Wiley & Sons.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.
doi:<https://doi.org/10.1136/bmj.327.7414.557>
- Higgins, J. P. T., Chandler, T. J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2021). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*: Cochrane.
- Huang, M., Jiao, J., Zhuang, P., Chen, X., Wang, J., & Zhang, Y. (2018). Serum polyfluoroalkyl chemicals are associated with risk of cardiovascular diseases in national US population. *Environment international*, 119, 37-46.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Yasmeeen, F. (2021). Forecasting functions for time series and linear models. 2021. URL <http://pkg.robjhyndman.com/forecast>, 8.
- Hyndman, R., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, 27(1), 1-22.
- Jain, R. B., & Ducatman, A. (2019). Roles of gender and obesity in defining correlations between perfluoroalkyl substances and lipid/lipoproteins. *Science of the Total Environment*, 653, 74-81.

- Kannel, W. B., D'Agostino, R. B., Silbershatz, H., Belanger, A. J., Wilson, P. W., & Levy, D. (1999). Profile for estimating risk of heart failure. *Arch. Intern. Med.*, 159(11), 1197-1204. doi:10.1001/archinte.159.11.1197
- Li, S., Peng, Y., Wang, X., Qian, Y., Xiang, P., Wade, S. W., . . . Handelsman, Y. (2019). Cardiovascular events and death after myocardial infarction or ischemic stroke in an older Medicare population. *Clinical cardiology*, 42(3), 391-399.
- Li, Y., Barregard, L., Xu, Y., Scott, K., Pineda, D., Lindh, C. H., . . . Fletcher, T. (2020). Associations between perfluoroalkyl substances and serum lipids in a Swedish adult population with contaminated drinking water. *Environmental Health*, 19(1), 1-11.
- Li, Y., Fletcher, T., Mucs, D., Scott, K., Lindh, C. H., Tallving, P., & Jakobsson, K. (2018). Half-lives of PFOS, PFHxS and PFOA after end of exposure to contaminated drinking water. *Occupational and environmental medicine*, 75(1), 46-51.
- Lin, C. Y., Lee, H. L., Hwang, Y. T., & Su, T. C. (2020). The association between total serum isomers of per-and polyfluoroalkyl substances, lipid profiles, and the DNA oxidative/nitrative stress biomarkers in middle-aged Taiwanese adults. *Environmental research*, 182, 109064.
- Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, 74(3), 785-794.
- Lin, P.-I. D., Cardenas, A., Hauser, R., Gold, D. R., Kleinman, K. P., Hivert, M.-F., . . . Horton, E. S. (2019). Per-and polyfluoroalkyl substances and blood lipid levels in pre-diabetic adults—longitudinal analysis of the diabetes prevention program outcomes study. *Environment international*, 129, 343-353.
- Liu, H.-S., Wen, L.-L., Chu, P.-L., & Lin, C.-Y. (2018). Association among total serum isomers of perfluorinated chemicals, glucose homeostasis, lipid profiles, serum protein and metabolic syndrome in adults: NHANES, 2013–2014. *Environmental Pollution*, 232, 73-79.
- Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *The lancet*, 383(9921), 999-1008.
- Nelson, J. W., Hatch, E. E., & Webster, T. F. (2010). Exposure to polyfluoroalkyl chemicals and cholesterol, body weight, and insulin resistance in the general US population. *Environmental health perspectives*, 118(2), 197-202.
- O'Sullivan, A. K., Rubin, J., Nyambose, J., Kuznik, A., Cohen, D. J., & Thompson, D. (2011). Cost estimation of cardiovascular disease events in the US. *Pharmacoeconomics*, 29(8), 693-704.
- Palmer, T. M., & Sterne, J. A. C. (Eds.). (2016). *Meta-analysis in Stata: An updated collection from the Stata Journal* (2nd ed.). College Station, TX: Stata Press.
- PEACE Trial Investigators. (2004). Angiotensin-converting–enzyme inhibition in stable coronary artery disease. *New England Journal of Medicine*, 351(20), 2058-2068.
- Pencina, M. J., D'Agostino, R. B., Sr., Larson, M. G., Massaro, J. M., & Vasan, R. S. (2009). Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation*, 119(24), 3078-3084. doi:10.1161/CIRCULATIONAHA.108.816694

- Pencina, M. J., D'Agostino, R. B., Sr., & Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat. Med.*, 30(1), 11-21. doi:10.1002/sim.4085
- Rappazzo, K. M., Coffman, E., & Hines, E. P. (2017). Exposure to perfluorinated alkyl substances and health outcomes in children: a systematic review of the epidemiologic literature. *International journal of environmental research and public health*, 14(7), 691.
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. *The handbook of research synthesis and meta-analysis*, 2, 295-316.
- Rodríguez-Barranco, M., Tobías, A., Redondo, D., Molina-Portillo, E., & Sánchez, M. J. (2017). Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC medical research methodology*, 17(1), 1-9.
- Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., . . . Ford, E. S. (2012). Heart disease and stroke statistics—2012 update: a report from the American Heart Association. *Circulation*, 125(1), e2.
- Rücker, G., Schwarzer, G., Carpenter, J., & Olkin, I. (2009). Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in medicine*, 28(5), 721-738.
- Saltelli, A., Chan, K., & Scott, E. (2009). Sensitivity Analysis. probability and statistics. In: Wiley, NY, USA.
- Seo, S.-H., Son, M.-H., Choi, S.-D., Lee, D.-H., & Chang, Y.-S. (2018). Influence of exposure to perfluoroalkyl substances (PFASs) on the Korean general population: 10-year trend and health effects. *Environment international*, 113, 149-161.
- Sidik, K., & Jonkman, J. N. (2005). A note on variance estimation in random effects meta-regression. *Journal of Biopharmaceutical Statistics*, 15(5), 823-838.
- StataCorp. (2019). *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LP.
- Steenland, K., Tinker, S., Frisbee, S., Ducatman, A., & Vaccarino, V. (2009). Association of perfluorooctanoic acid and perfluorooctane sulfonate with serum lipids among adults living near a chemical plant. *American journal of epidemiology*, 170(10), 1268-1278.
- Thom, T., Kannel, W., Silbershatz, H., & D'Agostino, R. (2001). Cardiovascular diseases in the United States and prevention approaches. *Hurst's the heart*, 1, 3-17.
- U.S. Census Bureau. (2017). *American Community Survey Single-Year Estimates. American Community Survey Demographic and Housing Estimates*. Retrieved from: <https://data.census.gov/cedsci/>
- U.S. Environmental Protection Agency. (2011). Exposure Factors Handbook 2011 Edition (Final Report). EPA/600/R-09/052F. Washington, DC: U.S. Environmental Protection Agency.
- U.S. Environmental Protection Agency. (2016a). *Health effects support document for Perfluorooctane Sulfonate (PFOS)*. (EPA 822-R-16-002). Retrieved from https://www.epa.gov/sites/production/files/2016-05/documents/pfos_hesd_final_508.pdf
- U.S. Environmental Protection Agency. (2016b). *Health effects support document for Perfluorooctanoic Acid (PFOA)*. (EPA 822-R-16-003). Retrieved from https://www.epa.gov/sites/production/files/2016-05/documents/pfoa_hesd_final-plain.pdf

- U.S. Environmental Protection Agency. (2020). *Safe Drinking Water Information System 2020 Quarter 4 Database*. Retrieved from <https://www.epa.gov/ground-water-and-drinking-water/safe-drinking-water-information-system-sdwis-federal-reporting>
- United States Environmental Protection Agency. (2021a). *Proposed Approaches for Deriving Maximum Contaminant Level Goals for PFOA in Drinking Water*. EPA 822-D-21-001.
- United States Environmental Protection Agency. (2021b). *Proposed Approaches for Deriving Maximum Contaminant Level Goals for PFOS in Drinking Water*. EPA 822-D-21-002.
- Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.*, 30(10), 1105-1117. doi:10.1002/sim.4154
- von Hippel, P. T. (2015). The heterogeneity statistic I² can be biased in small meta-analyses. *BMC medical research methodology*, 15(1), 1-8.
- Whooley, M. A., De Jonge, P., Vittinghoff, E., Otte, C., Moos, R., Carney, R. M., . . . Feldman, M. D. (2008). Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. *JAMA*, 300(20), 2379-2388.
- Williams, O. (1989). The Atherosclerosis Risk in Communities (ARIC) study-design and objectives. *American journal of epidemiology*, 129(4), 687-702.
- Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, 97(18), 1837-1847. doi:10.1161/01.cir.97.18.1837
- Woods & Poole Economics Inc. (2021). *Complete Demographic Database*. Retrieved from: <https://www.woodsandpoole.com/our-databases/united-states/all-geographies/>
- Yang, Q., Guo, X., Sun, P., Chen, Y., Zhang, W., & Gao, A. (2018). Association of serum levels of perfluoroalkyl substances (PFASs) with the metabolic syndrome (MetS) in Chinese male adults: A cross-sectional study. *Science of the Total Environment*, 621, 1542-1549. doi:10.1016/j.scitotenv.2017.10.074