International Journal of Epidemiology, 2019, 1004–1013
doi: 10.1093/ije/dyy264
Advance Access Publication Date: 7 December 2018
Original article

Miscellaneous

Use of a pooled cohort to impute cardiovascular disease risk factors across the adult life course

Adina Zeki Al Hazzouri,[1]* Eric Vittinghoff,[2] Yiyi Zhang,[3] Mark J Pletcher,[2] Andrew E Moran,[3] Kirsten Bibbins-Domingo,[2] Sherita H Golden[4,5] and Kristine Yaffe[2,6,7]

[1]Department of Public Health Sciences, University of Miami, Miami, FL, USA, [2]Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA, [3]Division of General Medicine, Columbia University Medical Center, New York, NY, USA, [4]Department of Medicine, Johns Hopkins University School of Medicine, Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA, [6]Department of Psychiatry and [7]Department of Neurology, University of California San Francisco, San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA

*Corresponding author. Division of Epidemiology, Department of Public Health Sciences, University of Miami, Miami, FL 33136, USA. E-mail: axz122@miami.edu

Editorial decision 19 October 2018; Accepted 10 November 2018

Abstract

Background: In designing prevention strategies, it may be useful to understand how early and midlife cardiovascular disease risk factor (CVDRF) exposures affect outcomes that primarily occur in mid to late life. Few single US cohorts have followed participants from early adulthood to late life.

Methods: We pooled four prospective cohorts that represent segments of the adult life course, and studied 15 001 White and Black adults aged 18 to 95 years at enrollment. We imputed early and midlife exposure to body mass index (BMI), glucose, lipids and blood pressure (BP). CVDRF trajectories were estimated using linear mixed models. Using the best linear unbiased predictions, we obtained person-specific estimates of CVDRF trajectories beginning at age 20 until each participant's end of follow-up. We then calculated for each CVDRF, summary measures of early and midlife exposure as time-weighted averages (TWAs).

Results: In the pooled cohort, 33.7% were Black and 54.8% were female. CVDRF summary measures worsened in midlife compared with early life and varied by sex and race. In particular, systolic and diastolic BP were consistently higher over the adult life course among men, and BMI was higher among Blacks, particularly Black women. Simulation studies suggested acceptable imputation accuracy, especially for the younger cohorts. Correlations of true and imputed CVDRF summary measures ranged from 0.53 to 0.99, and agreement ranged from 67% to 99%.

Conclusions: These results suggest that imputed CVDRFs may be accurate enough to be useful in assessing the effects of early and midlife exposures on later life outcomes.

Key words: Cardiovascular disease, cohort, imputation, life course

Key Messages

- Few single US cohorts have followed participants from early adulthood to late life. As such, there is a need to better understand whether using cardiovascular disease risk factor (CVDRF) patterns in younger cohorts to impute early and midlife CVDRF patterns for older cohorts yields imputation estimates that are accurate enough.
- We pooled data from four prospective cohorts with 15 001 White and Black adults aged 18 to 95 years at enrollment. We described our methods for imputing early and midlife CVDRFs, and reported simulations that assess the accuracy of those imputations.
- Our findings were in line with the well-established literature on sex and racial/ethnic differences in CVDRF levels. Our simulation studies suggested acceptable imputation accuracy, and agreement between true and imputed CVDRFs.
- The imputed CVDRFs may be accurate enough to be useful in assessing the effects of early and midlife exposures on late life outcomes, with implications for risk factor preventive intervention.

Introduction

In the USA, cardiovascular disease (CVD) affects one in three adults, remains the most common cause of morbidity and mortality[1,2] and is associated with nearly $450 billion per year in direct and indirect costs.[2] Moreover, CVD may in turn contribute to other costly adverse late life outcomes, including cognitive impairment and dementia.[3–10] The modifiable nature of CVD risk factors (CVDRFs), including lipids and blood pressure, makes them attractive targets for prevention.[11–13]

Emerging evidence suggests that adverse levels of CVDRFs develop early and tend to get worse over the life course.[2,14–17] For example, recent data from the National Health and Nutrition Examination Survey suggest that the prevalence of hypertension increases from approximately 24% in adulthood to over 50% in middle age and 70% in older age.[2] However, few epidemiological cohorts have overcome the logistic challenges of following participants from early adulthood to late life, with the result that few have been able to describe CVDRF trajectories over the life span, or estimate how early and midlife CVDRF exposures affect outcomes that primarily occur in mid to late life. Most cohorts have been restricted in age range, so that those with sufficient numbers of adverse outcomes that occur primarily in mid to late life lack information on early life CVDRF exposures. Longitudinal data linking CVDRFs in early or midlife with health outcomes occurring in mid life or later in life are important for evaluating the potential of early interventions that improve CVDRFs to prevent later adverse outcomes.

Another important issue is that most CVDRFs vary by sex or race/ethnicity.[17–19] For example, hypertension is more common and severe and develops at an earlier age among Blacks.[2,20,21] Obesity is also more common among women than men, and disproportionately affects minority populations.[17,22] However, few cohorts that span many decades recruited a racially and ethnically diverse sample.

To overcome these barriers to estimating the effects of early and midlife CVDRF exposures on later life outcomes, we pooled data from four large prospective cohorts, which together span the adult life course. Each includes White and Black participants. These cohorts include the Coronary Artery Risk Development in Young Adults (CARDIA) study of young to middle-aged adults, the Multi Ethnic Study of Atherosclerosis (MESA) of middle to older-aged adults, the Cardiovascular Health Study (CHS) and the Health, Aging and Body Composition (Health ABC) study of older adults. Based on the pooled data, we developed a method using early and midlife CVDRF patterns in the younger cohorts to impute earlier exposures in the older cohorts. In this paper, we describe our methods for imputing early and midlife CVDRFs, report simulations assessing the accuracy of those imputations, and compare imputed early and midlife CVDRF levels by sex and race.

Methods

Data sources and included cohorts

CARDIA[23] is an ongoing prospective cohort of 5115 adults recruited from four field centres at the University of

Alabama at Birmingham, the University of Minnesota, Northwestern University, and Kaiser Permanente Northern California. Participants were ages 18 to 30 years at baseline in 1985–86. By design, the sample was balanced within center by sex, age and education, and included 55% women and 52% Blacks. The baseline and eight follow-up examinations now span 30 years.

MESA[24] is an ongoing prospective study that includes 4515 White and Black adults recruited from six US communities: Baltimore, Chicago, Forsyth County in North Carolina, Los Angeles, New York City and St Paul. Participants were aged 45 to 84 years at baseline in 2000–01. The baseline sample included 53% women and 42% Blacks. The baseline and four follow-up examinations span 10 years.

CHS[25] is a recently completed prospective study of 5888 community-dwelling adults recruited from four US communities: Washington County in Maryland, Forsyth County in North Carolina, Sacramento County in California, and Allegheny County in Pennsylvania. Participants were 65 or older at baseline in 1990 and included 57% women and 12% Blacks. Participants were followed annually for up to 11 years.

Health ABC[26] is a recently completed prospective cohort study of 3075 community-dwelling adults aged 70–79 years at baseline in 1997. Participants were a random sample of Medicare-eligible older adults living in Memphis and Pittsburgh, including 52% women and 42% Blacks. Participants were followed annually or semi-annually for up to 11 years.

The inclusion of each of the four cohorts in our study was approved by local institutional review boards (IRBs) as well as the IRBs at the University of Miami and the University of California San Francisco, and the present analysis was approved by the Publications & Presentations committee of each study. All participants provided written informed consent.

Study sample

Our pooled cohort includes a total of 15 001 White and Black adult participants aged 18 to 95 years old at enrollment, with at least two repeated measurements of each of the CVD risk factors, including 4632 from CARDIA, 4238 from MESA, 3936 from CHS and 2195 from Health ABC. As shown in Figure 1, CARDIA participants contribute observations in early adulthood and middle age (18–60 years), MESA participants beginning in middle age (45+ years) and CHS and Health ABC beginning in older age (65+ years).

Measurement of CVDRFs

We used established measures from each study cohort to capture the CVDRFs that were measured at almost all study visits, using similar validated methods (see Supplementary Table 1, available as Supplementary data at *IJE* online), and that have been shown to influence large- and small-vessel atherosclerosis. As such, our primary CVDRFs include: body mass index (BMI, kg/m^2) calculated from measured height and weight; fasting glucose; fasting total cholesterol; fasting low-density lipoprotein (LDL) cholesterol; and systolic blood pressure (SBP) and

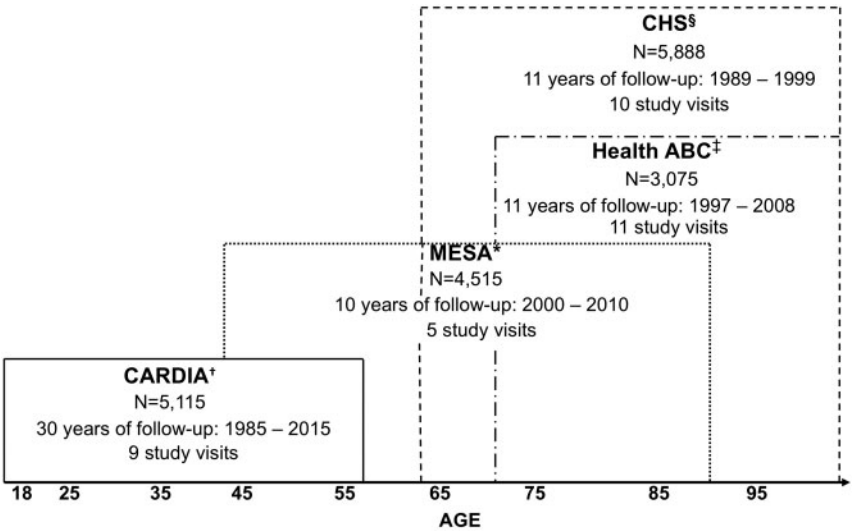

Figure 1. Description of the four study cohorts by age, sex and race.

diastolic blood pressure (DBP), each measured as the average of two seated measurements.[24,27–29] In addition, data were collected at most visits on additional risk factors that affect these CVDRFs, including current smoking status, age at onset of smoking, diabetes history, hypertension history and current use of diabetes, lipid-lowering and hypertensive medications.

Imputation procedures

The main aim of this study was to impute CVDRF levels earlier in the adult life course, beginning from age 20 years, for participants with follow-up data beginning in mid life or later in life. For example, for a MESA participant first seen at age 45 and last seen at age 55, the expected CVDRF levels were imputed each year from age 20 to 55. As shown in Supplementary Figure 2, available as Supplementary data at *IJE* online, the scope of the imputation is substantial for participants in the older cohorts.

To impute early and midlife CVDRF levels, we required annual values of additional variables hypothesized to affect them, including smoking status, diabetes and hypertension and use of medications for diabetes, hypertension and hyperlipidaemia. Imputations were done in the following nested sequence, motivated by previous knowledge of the dominant causal pathways: (i) smoking status, based on race and sex; (ii) BMI, based on race, sex and smoking status; (iii) diabetes and hypertension status, based on race, sex and BMI; (iv) use of diabetes, lipid-lowering and antihypertensive medications, based on sex, race and diabetes or hypertension status; and (v) blood pressure, lipid levels and glucose level, based on race, sex, BMI, diabetes, smoking status and medication use. All imputation models included birth year and cohort. Details regarding the modelling procedure and imputation steps are provided in Supplementary Methods, available as Supplementary data at *IJE* online.

To initiate the imputations, age at onset of smoking, diabetes, hypertension and medication use were randomly drawn from conditional distributions estimated using log-normal survival models. Diabetes and hypertension were assumed to persist after onset. Change points for medication use and smoking status with discrepant values at visits more than one year apart were randomly imputed from a uniform distribution.

To impute early and midlife CVDRF levels, including BMI, glucose, lipids and blood pressure, trajectories were estimated using linear mixed models (LMMs). Based on the LMMs, we used best linear unbiased predictions (BLUPs)[30] to obtain person-specific estimates of CVDRFs trajectories, annually from age 20 years until the end of follow-up for each participant. The BLUPs are estimated for every age with complete covariate data, as provided by previously imputed annual values of smoking status, diabetes, hypertension and medication use. CVDRF measures were log-transformed for LMM analysis, and BLUPs were back-transformed to the measured scale. Finally, using those BLUP trajectories, we calculated period-specific time-weighted averages (TWAs) as summary measures of early (ages 20–39 years) and midlife (ages 40–59 years) CVDRF exposure.

Statistical analysis

We first described participant characteristics at baseline, across each of the four included cohorts. Then, using the pooled data, we showed the distribution of the estimated CVDRF TWAs (i.e. summary measures) in early (ages 20–39 years) and mid life (ages 40 to 59 years), across sex and race categories.

Our imputation procedures depend on the assumption that age trends in CVDRF levels do not vary by cohort. To examine this assumption, we plotted cohort-specific trajectories in combination with trajectories based on the pooled data, by sex and race. To further assess the accuracy of the imputation procedures and their utility in future analyses estimating early and midlife CVDRF effects on mid- and late-life outcomes, we conducted a simulation study in which we treated our imputed BLUP trajectories and TWAs as the true values, and then used these as the basis for simulating new observations, BLUPs and TWAs. Specifically, to obtain multiple simulated values of the CVDRFs, we added random normal errors, with standard deviation (SD) determined by the initial LMMs, to the imputed BLUPs at each observed age, and then obtained a new set of BLUPs and TWAs by refitting the LMMs to these imputed outcomes; 25 imputations were used. We then assessed agreement between the true and imputed TWAs using four summary statistics: (i) the bias of the imputed TWAs, scaled by the mean of the true values; (ii) the mean absolute deviation (MAD) of the imputed from the true values, scaled by the MAD of the true measure from its sample mean; (iii) correlation of the true and imputed continuous TWAs; and (iv) agreement of the true and imputed TWAs, after categorization at established clinical cut-points. We also present scatter plots of the true and imputed values. Finally, to assess the magnitude and direction of bias in regression coefficients for the TWAs resulting from imputation error, we generated binary outcomes under logistic models based on the true categorized TWAs, adjusting for age, then fitted correctly specified logistic models using the estimated TWAs in place of the true values, and reported the average bias of the coefficients. We first considered a joint model for the effects of TWAs for

Table 1. Description of participant characteristics at baseline, by cohort

Characteristic, mean (SD) or %	CARDIA 18-30 years $n = 4632$	MESA 45-84 years $n = 4238$	CHS 65-100 years $n = 3936$	Health ABC 70-80 years $n = 2195$
Age, years	24.9 (3.6)	62.2 (10.1)	72.2 (5.1)	73.4 (2.8)
Black	50.3%	41.0%	4.4%	37.3%
Men	44.9%	47.0%	42.5%	47.0%
BMI, kg/m^2	24.5 (5.0)	28.7 (5.5)	26.4 (4.4)	27.4 (4.7)
Fasting glucose, mg/dl	82.3 (14.3)	94.5 (26.0)	107 (29.1)	103 (32.2)
Total cholesterol, mg/dl	177 (33.4)	193 (35.6)	212 (38.4)	204 (38.2)
LDL cholesterol, mg/dl	110 (31.3)	117 (31.4)	130 (35.3)	123 (34.4)
HDL cholesterol, mg/dl	53.2 (13.1)	52.4 (15.5)	54.5 (15.6)	54.1 (16.8)
Systolic BP, mmHg	110 (10.9)	126 (21.0)	135 (20.9)	134 (20.3)
Diastolic BP, mmHg	68.7 (9.4)	71.9 (10.2)	70.1 (11.0)	71.2 (11.3)
Current smoker	29.6%	13.8%	10.6%	8.3%
Hypertension	2.8%	50.6%	48.4%	69.2%
Diabetes	0.54%	10.7%	12.9%	36.3%
Diabetes medication use	0.22%	8.5%	6.0%	10.5%
Hypertension medication use	1.0%	40.0%	42.9%	52.6%
Lipid medication use	0.35%	17.4%	5.1%	14.1%

ages 20–39 and 40–59, and then a reduced model for the effect of the TWAs for ages 20–59.

Results

Baseline characteristics of the four cohorts

Descriptive characteristics across the four cohorts are presented in Table 1. Mean baseline BMI was 24.5 (SD = 5.0) in the youngest cohort (CARDIA), 28.7 (SD = 5.5) in MESA, 26.4 (SD = 4.4) in CHS and 27.4 (SD = 4.7) in Health ABC. Prevalence of current smoking was highest (29.6%) in CARDIA compared with all other cohorts (13.8% in MESA, 10.6% in CHS and 8.3% in Health ABC). As expected, mean total and LDL cholesterol and systolic blood pressure were lowest in the youngest cohort. For example, mean baseline LDL cholesterol was 110 mg/dl (SD = 31.3) in CARDIA, 117 mg/dl (SD = 31.4) in MESA, 130 mg/dl (SD = 35.3) in CHS and 123 mg/dl (SD = 34.4) in Health ABC. Prevalence of hypertension, diabetes and medication use also increased with age in the respective cohorts.

Distribution of CVDRF TWAs (i.e. summary measures) in early (ages 20–39 years) and midlife (ages 40-59 years)

The results of the imputation of early and midlife CVDRF exposure from the pooled data are shown in Table 2, across sex and race categories. In early life (ages 20–39 years), average TWAs for BMI were highest among Black women (25.9 kg/m^2) and men (24.6 kg/m^2) compared with White women (21.9 kg/m^2) and men (23.2 kg/m^2). Average TWAs for fasting glucose were lowest among Black women (81.4 mg/dl) and highest among White men (87.7 mg/dl). Average TWAs for total cholesterol were highest among White women (190 mg/dl) and men (188 mg/dl). Average TWAs for SBP and DBP were highest among men (Black men: 118 mmHg and 72.8 mmHg, White men: 120 mmHg and 72.8 mmHg, respectively). TWAs were higher (i.e. worse) in midlife than early life, but race and sex differences were similar.

Consistency of CVDRF trajectories across cohorts

Figure 2 shows that in areas of overlap in midlife (ages 45–55 years) between CARDIA and MESA cohorts, as well as in areas of overlap in late life (ages 65–85 years) between MESA, CHS and Health ABC cohorts, the trajectories of the CVDRF variables are similar across individual cohorts and consistent with the combined trajectories from the pooled cohort. Although we did find some differences in level, especially in the older age range, these differences are accounted for by including cohort-specific intercepts in the imputation linear mixed models (LMMs). Life course trajectories also differed by sex and race. For example Blacks, especially Black women, had consistently higher BMI than their White counterparts over the adult life course. The trajectories of systolic and diastolic BP were also consistently higher over the life course among men, both White and Black, compared with women.

Table 2. Cardiovascular disease risk factor TWAs in early (ages 20–39 years) and midlife (ages 40–59 years), by sex and race

	Ages 20-39				Ages 40-59			
	Black women $n = 2887$	Black men $n = 2173$	White women $n = 5336$	White men $n = 4605$	Black women $n = 2887$	Black men $n = 2173$	White women $n = 5336$	White men $n = 4605$
BMI, kg/m^2	25.9 (23.3, 29.3)	24.6 (22.9, 26.9)	21.9 (20.5, 23.9)	23.2 (22.0, 25.0)	30.5 (26.3, 35.5)	27.8 (24.7, 31.3)	25.1 (22.5, 28.5)	26.1 (24.0, 28.8)
Fasting glucose, mg/dl	81.4 (77.8, 85.0)	85.3 (81.4, 88.7)	83.8 (78.3, 85.7)	87.7 (81.8, 89.9)	90.7 (86.0, 97.9)	93.7 (88.6, 101)	90.7 (85.9, 94.7)	95.6 (90.2, 100)
Total cholesterol, mg/dl	182 (168, 196)	180 (167, 194)	190 (177, 200)	188 (176, 199)	191 (171, 207)	185 (166, 203)	205 (190, 218)	201 (186, 214)
LDL cholesterol, mg/dl	112 (100, 125)	112 (99.2, 126)	114 (104, 123)	121 (111, 130)	114 (95.5, 129)	114 (95.7, 131)	118 (104, 132)	127 (113, 140)
Systolic BP, mmHg	112 (107, 117)	118 (113, 123)	111 (105, 116)	120 (114, 126)	123 (114, 132)	124 (116, 133)	118 (109, 126)	124 (116, 132)
Diastolic BP, mmHg	69.4 (66.5, 72.2)	72.8 (69.7, 75.5)	67.7 (64.8, 69.8)	72.8 (70.0, 75.0)	75.5 (70.7, 80.6)	78.1 (73.1, 82.8)	70.0 (65.7, 74.1)	75.6 (71.6, 79.6)

Data are presented as medians and interquartile ranges.

Accuracy of early and midlife CVDRF TWAs

Table 3 shows the results of our simulation study. The scaled bias of the imputed TWAs was generally less than 1% of the true mean. The MADs of the imputed-true differences were generally a small to moderate percentage of the MADs of the true measures. Likewise, correlation and agreement were strong to moderate. Accuracy and agreement were lower for the older cohorts, as expected, as well as for the lipid measures. Scatterplots for a 25% random sample of the true and imputed TWAs are presented in Supplementary Figure 2, available as Supplementary data at *IJE* online. These plots illustrate the correlation of the continuous TWAs as well as the agreement of the categorized values, as represented by the proportion of data points in the diagonal sectors of each plot. Finally, Table 4 shows the expected attenuation bias in logistic regression coefficients for the estimated TWAs for ages 20–39, 40–59 and 20–59.

Discussion

We describe the use of pooled data from four epidemiological cohorts, with ample representation of Black and White persons and age ranges covering the adult life span, to estimate summary measures of early and midlife cardiovascular risk factor exposures. By pooling these data, we used information on risk factor patterns in the younger cohorts to impute early and midlife CVDRF exposures for the older cohorts, where epidemiological outcomes of interest occur at higher rates. We use the pooled data to estimate differences in CVDRF levels by age, sex and race. We also provide evidence that the imputed young and midlife exposures may be accurate enough to be useful in assessing their effects on later life outcomes.

There is a well-established literature on sex and racial/ethnic differences in cardiovascular disease risk factor levels.[2,17] As expected, our findings from the pooled cohort data showed similar differences that persisted over the life course. We found TWA BMI to be higher among Blacks than Whites, across the adult life course. TWA systolic blood pressure was also found highest among men, especially Black men, which is supported by previous studies. Across sex and race categories, TWAs of all CVDRFs worsened from early life to mid life. In future analyses, we will examine whether those CVDRF differences in early and midlife affect later life health outcomes.

Our methods for leveraging pooled cohort data to impute early and midlife exposures for participants in older cohorts have several limitations. First, the 'true' TWAs for the older cohorts in our simulation study have restricted range, due to shrinkage toward the mean, possibly inflating our categorical agreement measure. Second, the TWAs for early and midlife CVDRF exposures are unquestionably subject to imputation error, which our simulation study showed was greater for participants in the older cohorts, CHS and Health ABC. The simulation study also showed that the imputation errors induce sometimes attenuation bias in estimates of the effects of the TWAs. Attenuation bias is of course conservative,[31] and in this context, the large size of the pooled cohort should help to overcome the resulting loss of power.

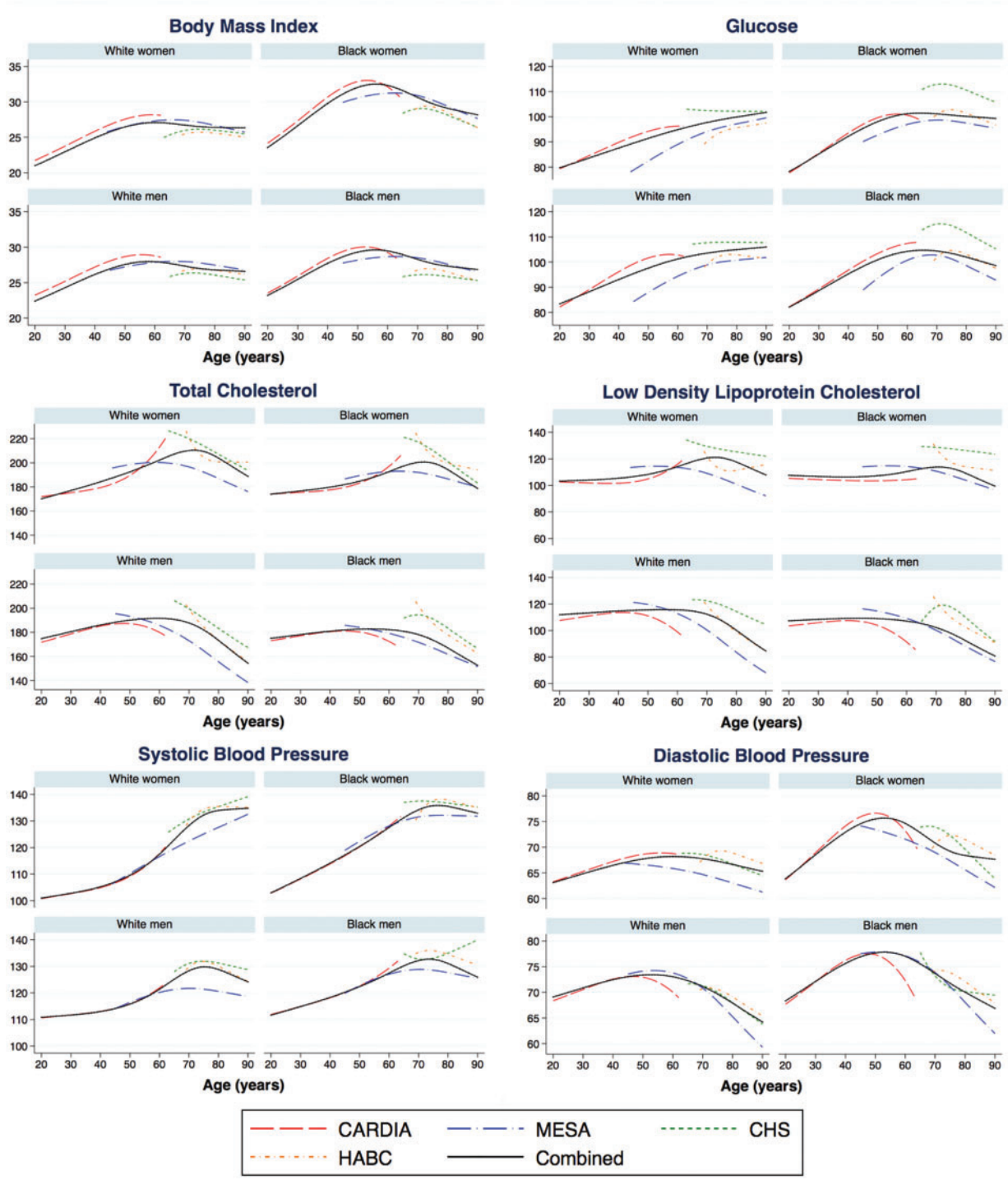

Figure 2. Life course trajectories of cardiovascular disease risk factors across categories of sex and race, results from the specific cohorts and pooled cohort.

Despite those limitations, a strength of our study is that we used innovative techniques to impute early and midlife CVDRFs and assessed the accuracy of those imputations. Our methods draw on an extensive literature using LMMs with splines for curve fitting,[32–34] extend recent work[35] using repeated risk factor measurements to predict later life events and allow us to make use of a priori knowledge about CVDRFs and the relationships among them—for example, that diabetes and hypertension almost never resolve once diagnosed, and that CVDRF trajectories are almost surely smooth. Simpler versions of the imputed early life exposures we propose here predicted mid- and late-life

Table 3. Measures of agreement between true and imputed TWAs, in early life (ages 20–39 years) and midlife (ages 40–59 years)

		Ages 20-39				Ages 40-59			
		CARDIA	MESA	CHS	Health ABC	CARDIA	MESA	CHS	Health ABC
BMI, kg/m^2	Bias (%)	0.06	0.37	−0.04	1.0	0.04	0.57	0.17	1.8
	MAD (%)	12	42	31	46	11	29	53	51
	Correlation	0.99	0.87	0.70	0.79	0.99	0.94	0.86	0.86
	Agreement (%)	94	82	91	80	94	83	75	72
Fasting glucose,mg/dl	Bias (%)	−0.03	−0.80	−0.96	−0.89	0.19	−0.09	−0.38	0.13
	MAD (%)	30	15	41	18	37	51	42	71
	Correlation	0.93	0.73	0.84	0.56	0.94	0.87	0.81	0.67
	Agreement (%)	99	99	99	99	89	93	83	84
Total cholesterol,mg/dl	Bias (%)	0.21	1.1	0.60	1.7	0.22	0.45	−0.48	0.89
	MAD (%)	27	59	96	86	35	50	42	62
	Correlation	0.95	0.70	0.44	0.51	0.93	0.87	0.81	0.66
	Agreement (%)	85	80	74	72	85	79	82	75
LDL cholesterol,mg/dl	Bias (%)	0.39	1.3	−3.4	−0.08	0.37	0.68	−3.8	−0.46
	MAD (%)	29	68	83	85	34	50	54	67
	Correlation	0.95	0.72	0.74	0.51	0.93	0.86	0.86	0.70
	Agreement (%)	84	71	79	67	83	74	75	68
Systolic BP, mmHg	Bias (%)	−0.02	0.19	0.28	0.69	0.08	0.66	0.80	1.4
	MAD (%)	21	26	36	47	37	48	42	49
	Correlation	0.93	0.92	0.92	0.86	0.92	0.86	0.82	0.72
	Agreement (%)	96	91	89	87	89	79	79	71
Diastolic BP, mmHg	Bias (%)	0.02	0.08	−0.17	−0.54	0.13	0.53	0.36	0.31
	MAD (%)	35	32	49	65	39	46	46	46
	Correlation	0.92	0.88	0.83	0.74	0.92	0.88	0.87	0.79
	Agreement (%)	97	99	99	98	90	87	89	77

Bias (%): mean difference between true and imputed TWA, as a percentage of the true mean TWA; MAD (%): mean absolute deviation of imputed from true TWA, as percentage of MAD of true values from mean; correlation: Pearson correlation of true and imputed TWAs; agreement: percentage agreement of categorized true and imputed TWAs. Clinical cut-points used in evaluating agreement: BMI (25, 30kg/m^2); glucose (100, 125 mg/dl); total cholesterol (160, 200 mg/dl); LDL cholesterol (100, 130 mg/dl); systolic BP (120, 140 mmHg); diastolic BP (80, 90 mmHg).

outcomes in previous work.[3,36,37] Our methods could also be used to derive other summary measures of early and midlife CVDRF exposures, including areas under the curve (AUCs), time spent over accepted thresholds, and average rates of change. In our study, we pooled multiple biracial cohorts with large sample sizes, repeated measures of CVDRFs, and long follow-up time. We also accounted for cohort and birth year effects. Risk factors for coronary heart disease, stroke and other manifestations of CVD may differ, though they overlap substantially. We focused on the primary underlying connection of atherosclerotic disease and risk factors that increase large- and small-vessel atherosclerosis, but our work can be extended to other risk factors and other study cohorts. In addition, in future analyses assessing the associations of early and midlife CVDRF exposures with mid- and late-life outcomes, potentially including death, we will develop inverse probability weights to address potential survival bias. MESA, CHS and Health ABC participants, who had to survive until cohort entry, may be somewhat atypical of the younger adult populations they are taken to represent. Finally, to account for estimation error in imputed CVDRF TWAs, and thus obtain valid standard errors in future analyses, we plan to use the methods developed for our simulation study to obtain multiple imputations of the BLUPs and TWAs; additional simulation studies will be used to determine the necessary number of imputations. We will then use established methods[38,39] for combining estimates based on multiply imputed data.

With CVD still the most common cause of morbidity and mortality in the USA, this study is relevant for population health, especially given the modifiable nature of CVDRFs such as lipids and blood pressure, making them attractive targets for prevention. We have described methods for pooling data from prospective cohorts that span the adult age range, in order to provide estimates of early and midlife exposure to CVDRFs that could be linked to health outcomes often occurring in midlife and later in life. We also provided evidence that the imputed exposures may be accurate enough to be useful in assessing the effects

Table 4. Percentage bias of regression coefficients based on estimated TWAs

		TWAs by age range		
		Joint model		Reduced model
	Level	20-39	40-59	20-59
BMI, kg/m^2	<25	ref	ref	ref
	25-30	−28.6	−13.7	−21.8
	>30	−10.3	−6.5	−13.9
Fasting glucose, mg/dl	<100	ref	ref	ref
	100-125	−27.1	−24.5	−16.4
	>125	−25.5	−16.7	−17.4
Total cholesterol, mg/dl	<160	ref	ref	ref
	160-200	−5.5	−8.6	−4.4
	>200	−20.8	−19.0	−23.8
LDL cholesterol, mg/dl	<100	ref	ref	ref
	100-130	−16.3	−5.5	−12.3
	>130	−23.2	−19.2	−26.2
Systolic BP, mmHg	<120	ref	ref	ref
	120-140	−10.3	−20.0	−27.3
	>140	−38.1	−38.1	−48.0
Diastolic BP, mmHg	<80	ref	ref	ref
	80-90	−11.0	−26.6	−43.3
	>90	−5.2	−36.5	−30.6

A joint model includes a TWA for early (ages 20–39 years) and a TWA for mid life (ages 40–59 years). A reduced model includes a TWA for ages 20–59 years.

of early and midlife exposures on late-life outcomes, with implications for risk factor preventive intervention.

Supplementary data

Supplementary data are available at *IJE* online.

Funding

This work was supported by grants from the National Institutes of Health, National Institute on Aging (1RF1AG054443 and K01AG047273) and National Heart, Lung, and Blood Institute (R01HL107475). The CARDIA study is supported by contracts HHSN268201300025C, HHSN268201300026C, HHSN268201300027C, HHSN268201300028C, HHSN268201300029C and HHSN268200900041C from the National Heart, Lung, and Blood Institute and the Intramural Research Program of the National Institute on Aging. The MESA study is supported by contracts HHSN2682015000003I, N01-HC-95159 through N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040 and UL1-TR-001079 from National Center for Research Resources. The CHS is supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083 and N01HC85086, and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided by R01AG023629 from the National Institute on Aging. A full list of principal CHS investigators and institutions can be found at [CHS-NHLBI.org]. The Health ABC study is supported by National Institute on Aging contracts N01-AG-6–2101, N01-AG-6–2103, N01-AG-6–2106, National Institute on Aging grant R01-AG028050 and National Institute of Nursing Research grant R01-NR012459. Health ABC was funded in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: None declared.

References

1. Wong MD, Shapiro MF, Boscardin WJ, Ettner SL. Contribution of major diseases to disparities in mortality. *N Engl J Med* 2002;**347**:1585–92.
2. Muntner P, Carey RM, Gidding S, Jones DW, Taler SJ, Wright JT Jr. Potential US population impact of the 2017 ACC/AHA high blood pressure guideline. *Circulation* 2018;**137**:109–18.
3. Yaffe K, Vittinghoff E, Pletcher MJ *et al*. Early adult to midlife cardiovascular risk factors and cognitive function. *Circulation* 2014;**129**:1560–67.
4. Zeki Al Hazzouri A, Elfassy T, Carnethon MR, Lloyd-Jones DM, Yaffe K. Heart rate variability and cognitive function in middle-age adults: the coronary artery risk development in young adults. *Am J Hypertens* 2018;**31**:27–34.
5. Iadecola C, Yaffe K, Biller J *et al*. Impact of hypertension on cognitive function: a scientific statement from the American Heart Association. *Hypertension* 2016;**68**:e67–94.
6. Yaffe K. Preclinical Alzheimer disease: Prevention Holy Grail or Pandora's Box?: Comment on "Heavy smoking in midlife and long-term risk of Alzheimer disease and vascular dementia". *Arch Intern Med* 2011;**171**:339–40.
7. Solomon A, Kivipelto M, Wolozin B, Zhou J, Whitmer RA. Midlife serum cholesterol and increased risk of Alzheimer's and vascular dementia three decades later. *Dement Geriatr Cogn Disord* 2009;**28**:75–80.
8. Fitzpatrick AL, Kuller LH, Lopez OL *et al*. Midlife and late-life obesity and the risk of dementia: cardiovascular health study. *Arch Neurol* 2009;**66**:336–42.
9. Whitmer RA, Gunderson EP, Barrett-Connor E, Quesenberry CP Jr, Yaffe K. Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study. *BMJ* 2005;**330**:1360.
10. Kalmijn S, van Boxtel MP, Verschuren MW, Jolles J, Launer LJ. Cigarette smoking and alcohol consumption in relation to cognitive performance in middle age. *Am J Epidemiol* 2002;**156**:936–44.
11. Heller DJ, Coxson PG, Penko J *et al*. Evaluating the impact and cost-effectiveness of statin use guidelines for primary prevention of coronary heart disease and stroke. *Circulation* 2017;**136**:1087–98.
12. Moise N, Huang C, Rodgers A *et al*. Comparative cost-effectiveness of conservative or intensive blood pressure treatment guidelines in adults aged 35-74 years: the cardiovascular disease policy model. *Hypertension* 2016;**68**:88–96.
13. Moran AE, Odden MC, Thanataveerat A *et al*. Cost-effectiveness of hypertension therapy according to 2014 guidelines. *N Engl J Med* 2015;**372**:447–55.

14. Freedman DS, Goodman A, Contreras OA, DasMahapatra P, Srinivasan SR, Berenson GS. Secular trends in BMI and blood pressure among children and adolescents: the Bogalusa Heart Study. *Pediatrics* 2012;**130**:e159–66.
15. Chen W, Srinivasan SR, Li S, Xu J, Berenson GS. Clustering of long-term trends in metabolic syndrome variables from childhood to adulthood in Blacks and Whites: the Bogalusa Heart Study. *Am J Epidemiol* 2007;**166**:527–33.
16. Wills AK, Lawlor DA, Matthews FE *et al*. Life course trajectories of systolic blood pressure using longitudinal data from eight UK cohorts. *PLoS Med* 2011;**8**:e1000440.
17. Benjamin EJ, Blaha MJ, Chiuve SE *et al*. Heart disease and stroke statistics - 2017 update: a report from the American Heart Association. *Circulation* 2017;**135**:e146–603.
18. Boykin S, Diez-Roux AV, Carnethon M, Shrager S, Ni H, Whitt-Glover M. Racial/ethnic heterogeneity in the socioeconomic patterning of CVD risk factors: in the United States: the multi-ethnic study of atherosclerosis. *J Health Care Poor Underserved* 2011;**22**:111–27.
19. Hickson DA, Diez Roux AV, Gebreab SY *et al*. Social patterning of cumulative biological risk by education and income among African Americans. *Am J Public Health* 2012;**102**:1362–69.
20. Dyer AR, Liu K, Walsh M, Kiefe C, Jacobs DR Jr, Bild DE. Ten-year incidence of elevated blood pressure and its predictors: the CARDIA study. Coronary Artery Risk Development in (Young) Adults. *J Hum Hypertens* 1999;**13**:13–21.
21. Liu K, Ruth K, Flack JM *et al*. Blood pressure in young Blacks and Whites: relevance of obesity and lifestyle factors in determining differences: the CARDIA study. *Circulation* 1996;**93**:60–66.
22. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United States, 2011-2014. *NCHS Data Brief* 2015;**219**:1–8.
23. Friedman GD, Cutter GR, Donahue RP *et al*. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* 1988;**41**:1105–16.
24. Bild DE, Bluemke DA, Burke GL *et al*. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 2002;**156**:871–81.
25. Fried LP, Borhani NO, Enright P *et al*. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;**1**:263–76.
26. Goodpaster BH, Carlson CL, Visser M *et al*. Attenuation of skeletal muscle and strength in the elderly: the Health ABC Study. *J Appl Physiol* 2001;**90**:2157–65.
27. Matthews KA, Katholi CR, McCreath H *et al*. Blood pressure reactivity to psychological stress predicts hypertension in the CARDIA study. *Circulation* 2004;**110**:74–78.
28. Rosano C, Abebe KZ, Aizenstein HJ *et al*. Longitudinal systolic blood pressure characteristics and integrity of white matter tracts in a cohort of very old black and white adults. *Am J Hypertens* 2015;**28**:326–34.
29. Tell GS, Rutan GH, Kronmal RA *et al*. Correlates of blood pressure in community-dwelling older adults. The Cardiovascular Health Study. Cardiovascular Health Study (CHS) Collaborative Research Group. *Hypertension* 1994;**23**:59–67.
30. Robinson GK. That BLUP is a good thing: the estimation of random effects. *Stat Sci* 1991;**6**:15–51.
31. Frost C, Thompson S. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *J R Stat Soc A* 2000;**163**:173–89.
32. Brumback BA, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves. *J Am Stat Assoc* 1998; **93**:961–76.
33. Rice JA, Wu CO. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 2001;**57**: 253–59.
34. Ye W, Lin X, Taylor JMG. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics* 2008;**64**: 1238–46.
35. Berry JD, Dyer A, Cai X *et al*. Lifetime risks for cardiovascular disease. *N Engl J Med* 2012;**366**:321–29.
36. Pletcher MJ, Bibbins-Domingo K, Liu K *et al*. Nonoptimal lipids commonly present in young adults and coronary calcium later in life: the CARDIA (Coronary Artery Risk Development in Young Adults) study. *Ann Intern Med* 2010;**153**:137–46.
37. Pletcher MJ, Vittinghoff E, Thanataveerat A, Bibbins-Domingo K, Moran AE. Young adult exposure to cardiovascular risk factors and risk of events later in life: the Framingham Offspring Study. *PLoS One* 2016;**11**:e0154288.
38. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;**8**:3–15.
39. Lopiano KK, Young LJ, Gotway CA. A comparison of errors in variables methods in regression models with spatially misaligned data. *Stat Methods Med Res* 2011;**20**:29–47.