

1 Time-to-Event Modeling of Hypertension Reveals the Nonexistence of True Controls

2

3 Daniel Shriner, Amy R. Bentley, Jie Zhou, Kenneth Ekoru, Ayo P. Doumatey, Guanjie Chen,

4 Adebowale Adeyemo, Charles N. Rotimi*

5

6 Center for Research on Genomics and Global Health, National Human Genome Research

7 Institute, Bethesda, Maryland 20892 USA

8

9 *Corresponding author: rotimic@mail.nih.gov

10 **Abstract**

11 Given a lifetime risk of ~90% by the ninth decade of life, it is unknown if there are true
12 controls for hypertension in epidemiological and genetic studies. Here, we use the Bayesian
13 framework to compare logistic and time-to-event approaches to modeling hypertension. Using a
14 proportional hazards model, we explored nonparametric and parametric models of the baseline
15 hazard function, accounting for interval censoring. In the Howard University Family Study
16 (HUFS), a population-based study of African Americans from Washington, D.C., the median age
17 at hypertension was 48 years, baseline hazard rates increased with age until 55 years, and the
18 probability of being free of hypertension at 85 years of age was 12.2%. In the nationwide
19 NHANES study, the median age at hypertension was 42 years in African Americans, in contrast
20 to 57 years in European Americans and 56 years in Mexican Americans. Baseline hazard rates
21 increased with age until 58 years in African Americans, comparable to 60 years in European
22 Americans and 58 years in Mexican Americans. The probability of being free of hypertension at
23 85 years of age was 8.4% in African Americans, in contrast to 21.4% in European Americans
24 and 20.6% in Mexican Americans. In all four groups, baseline hazard rates decreased but did not
25 reach zero, consistent with the nonexistence of controls. Model fits were comparable for a
26 proportional hazards model based on gamma-distributed hazard rates under a correlated prior
27 process and a logistic model adjusted for age and age². Last, using an agnostic model screening
28 approach of 38 potential covariates, we identified and replicated in all groups a model that
29 included chloride, low-density lipoprotein cholesterol, uric acid, and weight. With age and age²,
30 the variance explained by these covariates was 40.9% in African Americans, 34.8% in European
31 Americans, and 28.3% in Mexican Americans. Taken together, modeling of the baseline hazard
32 function of hypertension suggests that there are no true controls and that controls in logistic

33 regression are cases with a late age of onset. These findings shed considerable insights into the
34 design of genetic and epidemiological studies of hypertension with implications for ethnic health
35 disparities.

36 **Introduction**

37 Hypertension, or abnormally high blood pressure, is common in the US, affecting
38 approximately 45% of adults (Centers for Disease Control and Prevention, 2018). Hypertension
39 is a risk factor for heart disease and stroke and causes or contributes to nearly half a million
40 deaths a year (Centers for Disease Control and Prevention, 2019). Globally, an estimated 1.13
41 billion people have hypertension, and less than 20% of these people have their blood pressure
42 under control (World Health Organization, 2019).

43 Systolic blood pressure (SBP) has a general tendency to increase linearly with age, across
44 sexes and ethnic groups (Burt et al., 1995). Diastolic blood pressure (DBP) has a general
45 tendency to increase until the end of the fifth decade of life, after which DBP either stabilizes or
46 decreases, again across sexes and ethnic groups (Burt et al., 1995). In the Framingham Heart
47 Study, an individual who is normotensive at 55 to 65 years of age has an 80% to 90% residual
48 lifetime risk of developing hypertension, adjusted for competing causes of mortality (Vasan et
49 al., 2002). Compared to age-matched European Americans, hypertension in African Americans
50 develops at an earlier age and is more prevalent (Chobanian et al., 2003; Cooper, Liao, &
51 Rotimi, 1996; Mozaffarian et al., 2016).

52 A common approach in genetic epidemiology studies of hypertension involves coding the
53 outcome as a binary variable representing cases and controls and proceeds with logistic
54 regression. Given that the lifetime risk is so high, we first investigated whether a proportional
55 hazards model in time-to-event analysis yields a better fit than logistic regression. Second, as
56 time-to-event analysis assumes that the event will occur, *i.e.*, that every individual will become
57 hypertensive if they live long enough, we investigated a proportional hazards model including a
58 fraction of individuals that will never become hypertensive and hence are true epidemiological

59 controls. Third, using an agnostic model screening approach, we explored the issue of what
60 covariates to include and how much variance they explain. We performed these analyses in an
61 observational study of African Americans and then replicated and extended our findings in a
62 nationally representative study of African Americans, European Americans, and Mexican
63 Americans.

64 **Materials and Methods**

65 *Discovery study*

66 The Howard University Family Study (HUFS) is a population-based observational study of
67 African American families and unrelated individuals from Washington, D.C. (Adeyemo et al.,
68 2009). Ethics approval for the Howard University Family Study (HUFS) was obtained from the
69 Howard University Institutional Review Board and written informed consent was obtained from
70 each participant. All clinical investigation was conducted according to the principles expressed
71 in the Declaration of Helsinki. Families and individuals were not ascertained based on any
72 phenotype. Weight was measured on an electronic scale to the nearest 0.1 kg. Height was
73 measured on a stadiometer to the nearest 0.1 cm. Body mass index (BMI) was calculated as
74 weight divided by the square of height (kg/m^2). Waist circumference was measured to the nearest
75 0.1 cm at the narrowest part of the torso. Hip circumference was measured to the nearest 0.1 cm
76 at the widest part of the buttocks. The waist-hip ratio was calculated as waist circumference in
77 cm divided by hip circumference in cm. Fat mass and fat-free mass were estimated using
78 bioelectrical impedance analysis with a validated population-specific equation as previously
79 described (Luke et al., 1997). Percent fat mass was defined as fat mass divided by weight $\times 100$.
80 Blood pressure was measured while seated using an oscillometric device (Omron Healthcare,
81 Inc., Bannockburn, Illinois). Three readings were taken at 10-minute intervals. Reported readings
82 were the averages of the second and third readings. Hypertension was defined as SBP ≥ 140 mm
83 Hg, DBP ≥ 90 mm Hg, or treatment with anti-hypertensive medication. Blood was drawn after
84 an overnight fast of at least 8 hours and all collected samples were stored at -80°C pending
85 biochemical assay. Creatinine, total cholesterol, high-density lipoprotein cholesterol, low-density
86 lipoprotein cholesterol, triglycerides, fructosamine, glucose, alkaline phosphatase, alanine

87 aminotransferase, total bilirubin, sodium, potassium, chloride, calcium, uric acid, urea, C-
88 reactive protein, albumin, bicarbonate, and total protein were measured using COBAS
89 INTEGRA tests (Roche Diagnostics, Indianapolis, Indiana). Cortisol and insulin were measured
90 using Elecsys assays (Roche Diagnostics). Creatinine clearance was calculated using the
91 Cockcroft-Gault equation and the estimated glomerular filtration rate (eGFR) was calculated
92 using the 4-variable Modification of Diet in Renal Disease Study equation (National Kidney
93 Foundation, 2002; Levey et al., 1999). T2D case status was defined as fasting plasma glucose
94 level ≥ 126 mg/dL or treatment with anti-diabetic medication. T2D control status was defined as
95 fasting plasma glucose ≤ 100 mg/dL and no treatment with anti-diabetic medication. The last
96 completed grade of education and income were self-reported on a questionnaire. The proportions
97 of African and European ancestry were estimated as described previously (Shriner, Adeyemo, &
98 Rotimi, 2011). We extracted a subset of 1,014 unrelated individuals.

99

100 *Bayesian logistic regression and time-to-event analysis*

101 Let T represent the time of an event and $S(t) = \Pr(T > t)$ represent the survival function,
102 *i.e.*, the probability of being event-free as a function of observation time t . With cross-sectional
103 data, there is a single observation for each individual. If the individual has not yet experienced
104 the event, then the event time is right censored, because the event is presumed to occur some
105 unknown time after observation. If the individual has already experienced the event, then the
106 event time is left censored, because the event occurred at some unknown time prior to
107 observation. The combination of left and right censored data is known as interval censored data.
108 We performed interval censored proportional hazards analysis using the R package *icenReg*.

109 To perform Bayesian modeling, we used WinBUGS, version 1.4 with the R package
110 R2WinBUGS. We modeled logistic regression as:

$$111 \quad y_i \sim \text{Bernoulli}(\theta_i)$$

$$112 \quad \text{logit}(\theta_i) = \alpha$$

$$113 \quad \alpha \sim \text{Normal}(0, 10^6).$$

114 In this reduced model, the prior distribution for the intercept α follows a diffuse normal
115 distribution with mean 0 and variance 10^6 . We ran three chains of 10,000 iterations, with a burn-
116 in of 2,500 iterations and thinning of 10, yielding a posterior sample based on 2,250 iterations.
117 We assessed convergence using the potential scale reduction factor Rhat, which should equal 1 at
118 convergence for all monitored parameters. We assessed model fit using the deviance information
119 criterion (DIC). We then added age to the reduced model and ran three chains of 10,000
120 iterations, with a burn-in of 2,500 iterations and thinning of 10, yielding a posterior sample of
121 2,250 iterations. Next, we added age and age² to the reduced model and ran three chains of
122 100,000 iterations, with a burn-in of 10,000 and thinning of 50, yielding a posterior sample of
123 5,400 iterations. In all instances, effect sizes followed a diffuse normal prior distribution with
124 mean 0 and variance 10^6 .

125 We performed time-to-event analysis using a proportional hazards model (Congdon, 2003).
126 The hazard function $h(t)$ defines the instantaneous risk of the event at time t , conditional on
127 being event-free at that time:

$$128 \quad h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

129 Given the hazard function $h(t)$, the cumulative hazard function $H(t)$ and the survival function
130 $S(t)$ are defined as follows:

$$131 \quad H(t) = \int_0^t h(u) du$$

$$132 \quad S(t) = \Pr(T \geq t) = e^{-H(t)}.$$

133 In Cox's proportional hazards model (Cox, 1972), the hazard function $\lambda(t|z)$ is given by
134 $\lambda_0(t)e^{\beta z}$, in which $\lambda_0(t)$ is the baseline hazard function and z is a covariate with coefficient β .
135 In the absence of covariates, the hazard function is equivalent to the baseline hazard function,
136 $\lambda(t) = \lambda_0(t)$. We modeled the hazard rates as gamma-distributed with a correlated prior process
137 (Arjas & Gasbarra, 1994; Congdon, 2006). Specifically, the hazard rates were:

$$138 \quad \lambda_0(t=0) \sim \text{Gamma}(\alpha, \beta)$$

$$139 \quad \lambda_0(t+1) \sim \text{Gamma}\left(\alpha, \frac{\alpha}{\lambda_0(t)}\right)$$

$$140 \quad \alpha \sim \text{Uniform}(10, 100)$$

$$141 \quad \beta \sim \text{Uniform}(0.001, 0.01).$$

142 In this model, the expected value of $\lambda_0(t+1)$ equals $\lambda_0(t)$ and hence the hazard function is a
143 martingale. We divided time into intervals of one year and assumed that the baseline hazard was
144 constant within intervals.

145 A major assumption of time-to-event analysis is that the event will occur for every individual
146 at some point in time, although the individual may die before the event would have occurred.
147 This assumption can be relaxed by incorporating a permanent stayer or cured fraction, *i.e.*, a

148 fraction of individuals who will never experience the event. In our context, this fraction
149 represents individuals who never develop hypertension and hence are true epidemiological
150 controls. Let π represent the permanent stayer fraction. Then, the survival function for the entire
151 population $S_p(t)$ is the two-component mixture model $S_p(t) = \pi + (1 - \pi)S(t)$ (Gu, Sinha, &
152 Banerjee, 2011). We assigned to π the prior distribution $\text{Uniform}(0,1)$.

153

154 *Model selection*

155 We tested 38 covariates for inclusion in the model: sex; weight, height, hip circumference,
156 waist circumference, waist-hip ratio, body mass index; fat mass, fat-free mass, percent fat mass;
157 type 2 diabetes status, fasting glucose, fasting insulin, fructosamine; triglycerides, high-density
158 lipoprotein cholesterol, low-density lipoprotein cholesterol, total cholesterol; creatinine,
159 creatinine clearance, estimated glomerular filtration rate; alkaline phosphatase, alanine
160 aminotransferase, total bilirubin; last grade of education completed, income; percent African
161 ancestry; calcium, chloride, potassium, sodium; albumin, carbon dioxide, C-reactive protein,
162 total protein, uric acid, urea, and cortisol. We used a four-step forward-backward regression
163 procedure to perform model selection. First, we fit a single regression model for each covariate.
164 Second, we fit a multiple regression model with all significant predictors from step 1 and used
165 backward selection to remove nonsignificant predictors. Third, starting with the final model from
166 step 2, we reconsidered each nonsignificant covariate from step 1 using forward selection.
167 Fourth, we performed a final pruning step on the final model from step 3. For every test, we
168 declared a significance level of 0.05. Pseudo- r^2 values were estimated using the formula

169
$$r^2 = \frac{1 - \exp\left(\frac{D_1 - D_0}{n}\right)}{1 - \exp\left(\frac{-D_0}{n}\right)},$$
 in which D_1 is the deviance, D_0 is the null deviance, and n is the

170 sample size (Nagelkerke, 1991). Finally, we added age, age², and selected covariates to the
171 Bayesian logistic regression model described above and ran three chains of 1,000,000 iterations,
172 with a burn-in of 100,000 iterations and thinning of 1,000, yielding a posterior sample of 2,700
173 iterations. Effect sizes for all covariates followed a diffuse normal prior distribution with mean 0
174 and variance 10⁶.

175

176 *Replication study*

177 The National Center for Health Statistics of the US Centers for Disease Control and
178 Prevention conducts the ongoing National Health and Nutrition Examination Survey
179 (NHANES). The survey comprises an in-home interview and a clinical examination by a mobile
180 examination center. We retrieved 16 years of examination data (from 1999 to 2014) from the
181 CDC portal (<http://wwwn.cdc.gov/Nchs/Nhanes>). We downloaded the variables BMXWT,
182 BPQ020, BPQ040A, BPXDI1, BPXDI2, BPXDI3, BPXDI4, BPXSY1, BPXSY2, BPXSY3,
183 BPXSY4, LBDLDL, LBXIN, LBXSCLSI, LBXSKSI, LBXSUA, RIDAGEEX, RIDRETH1,
184 SDVMVPSU, SDMVSTRA, and WTSAF2YR. SBP was defined as the average of BPXSY1,
185 BPXSY2, BPXSY3, and BPXSY4. DBP was defined as the average of BPXDI1, BPXDI2,
186 BPXDI3, and BPXDI4. Hypertension was defined as SBP ≥ 140 mm Hg, DBP ≥ 90 mm Hg,
187 treatment with anti-hypertensive medication, or having ever been diagnosed by a doctor. Strata,
188 clusters, and weights were designed to make statistical estimates representative of the non-
189 institutionalized, civilian US population. We included the strata variable SDMVSTRA as a
190 factor. Clusters were nested within strata but were ultimately omitted because the data were too

191 sparse after including only those individuals with fasting samples. To account for eight survey
192 cycles, we multiplied the weights WTSF2YR equally by 1/8. For each value of RIDRETH, we
193 rescaled the weights by dividing by the mean. For the survey cycle 2013–2014, we recalibrated
194 insulin to account for changes in the protocol: $\text{Insulin}_{2011-2012} = 10^{(0.9765 * \log_{10}(\text{Insulin}_{2013-2014} + 0.07832))}$
195 (https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/INS_H.htm). Across the eight survey cycles, we
196 retrieved data for a total of 23,628 participants, including 5,146 African Americans (“non-
197 Hispanic Blacks”), 10,023 European Americans (“non-Hispanic Whites”), and 5,059 Mexican
198 Americans.

199

200 *Code Availability*

201 WinBUGS code is available at <https://github.com/dshriner/Time-to-event>.

202 **Results**

203 *Time-to-event analysis*

204 Across African Americans, European Americans, and Mexican Americans, median systolic
205 blood pressure increased with age whereas median diastolic blood pressure increased and then
206 decreased (Figure 1). In time-to-event analysis, the probability of not having hypertension
207 decreased across the entire age range (Figure 2). However, in all three groups, there was an
208 inflection point in middle age after which the probability of having hypertension increased at a
209 slower rate (Figure 2). Despite this slowdown, the probability of not having hypertension was not
210 zero, even by the middle of the ninth decade (Figure 2). In HUFS, by 85 years of age, 12.2%
211 (95% credible interval (CI) [5.1%, 20.6%]) of African Americans remained free of hypertension
212 (Figure 2). In NHANES, by 85 years of age, 8.4% (95% CI [5.4%, 11.6%]) of African
213 Americans, 21.4% (95% CI [18.1%, 24.6%]) of European Americans, and 20.6% (95% CI
214 [13.6%, 27.3%]) of Mexican Americans remained free of hypertension (Figure 2). The median
215 age at which hypertension occurred was 48 (95% CI [45, 50]) years for African Americans in
216 HUFS and 42 (95% CI [40, 44]) years for African Americans, 57 (95% CI [55, 59]) years for
217 European Americans, and 56 (95% CI [54, 58]) years for Mexican Americans in NHANES
218 (Figure 2).

219 We next investigated seven distributions, six parametric and one nonparametric, for the
220 baseline hazard function in a proportional hazards model. Generalized gamma, loglogistic, and
221 exponential distributions all yielded poor fits to the HUFS data (Figure 3). Log-normal, gamma,
222 and Weibull distributions all yielded good fits to the HUFS data (Figure 3), with the log-normal
223 distribution yielding the highest likelihood of the six parametric distributions (Table 1). Overall,
224 a nonparametric hazard function yielded the highest likelihood (Table 1). However, we could not

225 perform model comparison between the log-normal distribution of the baseline hazard function
226 and the nonparametric hazard function in the frequentist framework because it is unclear how to
227 measure the dimensionality of the nonparametric hazard function.

228 Based on a Bayesian model in which the hazard rates were gamma distributed with a
229 correlated prior process, we estimated the underlying hazard rates. Across African Americans,
230 European Americans, and Mexican Americans, hazard rates increased and then decreased
231 (Figure 4). The initial hazard rate for African Americans (0.0123 (95% highest posterior density
232 interval [0.0091, 0.0161])) was larger than the initial hazard rate for both European Americans
233 (0.0028 (95% highest posterior density interval [0.0013, 0.0044])) and Mexican Americans
234 (0.0036 (95% highest posterior density interval [0.0019, 0.0052])). The maximum hazard rate for
235 African Americans (0.0630 (95% highest posterior density interval [0.0341, 0.0940])) was
236 trending larger than the maximum hazard rate for European Americans (0.0368 (95% highest
237 posterior density interval [0.0200, 0.0546])) and Mexican Americans (0.0417 (95% highest
238 posterior density interval [0.0219, 0.0635])). The age of maximum hazard was 55 (95% CI [43,
239 72]) years in HUFs and 58 (95% CI [48, 67]) years in African Americans, 60 (95% CI [52, 70])
240 years in European Americans, and 58 (95% CI [48, 69]) years in Mexican Americans in
241 NHANES (Figure 4).

242

243 *Proportional Hazards vs. Logistic Models*

244 Using the HUFs data, we found that logistic regression yielded a better fit than the
245 proportional hazards model based on the DIC, provided that both age and age² were included as
246 covariates in the logistic regression model (Table 2). Using logistic regression, the addition of
247 age to the reduced (intercept-only) model resulted in a substantially lower DIC (Table 2), with a

248 linear effect of age explaining 27.8% of the variance at the cost of one additional parameter. The
249 addition of age² further decreased the DIC (Table 2), explaining an additional 0.8% of the
250 variance at the cost of one additional parameter. With smoothing, the effective dimensionality of
251 the proportional hazards model was 3.3, comparable to the dimensionality of 3.0 for the logistic
252 model adjusted for age and age² (Table 2). We also found that inclusion of a permanent stayer
253 fraction increased the DIC of the proportional hazards model, indicating that inclusion of a
254 permanent stayer fraction was not supported (Table 2).

255

256 *Model selection*

257 We used forward-backward regression to perform model selection on a set of 38 potential
258 covariates. In the HUFSS data set, the final model included six of these covariates: chloride,
259 insulin, low-density lipoprotein cholesterol, potassium, uric acid, and weight. Compared to the
260 logistic model with age and age², these six covariates improved the fit ($p = 2.01 \times 10^{-15}$) and
261 explained an additional 8.1% of the variance, for a total of 36.7% of variance explained. Each of
262 these six covariates were replicated (p -values from 2.81×10^{-2} to 1.83×10^{-12}) and directionally
263 consistent in NHANES African Americans (Table 3). Furthermore, chloride, low-density
264 lipoprotein cholesterol, potassium, uric acid, and weight, but not insulin, were significant
265 covariates in NHANES European Americans, whereas chloride, insulin, low-density lipoprotein
266 cholesterol, uric acid, and weight, but not potassium, were significant covariates in NHANES
267 Mexican Americans (Table 4). In NHANES, the variance explained by age, age², chloride, low-
268 density lipoprotein cholesterol, uric acid, and weight was 40.9% in African Americans, 34.8% in
269 European Americans, and 28.3% in Mexican Americans.

270 We further investigated two unexpected results of the model selection. One, sex was not
271 selected in the final model. Of the selected covariates, three showed sex dimorphisms in HUFS
272 by Welch's *t*-test, with uric acid and weight higher in males and low-density lipoprotein
273 cholesterol higher in females (Table 5). Second, across African Americans, European
274 Americans, and Mexican Americans, increasing low-density lipoprotein cholesterol was
275 associated with decreased risk of hypertension (Tables 3 and 4). As the direction of this effect
276 was unexpected, we reanalyzed the HUFS African Americans accounting for lipids medications.
277 Self-reported use of any lipid medication (coded as yes/no) was associated with increased risk of
278 hypertension ($p = 1.25 \times 10^{-3}$) but increasing low-density lipoprotein cholesterol remained
279 associated with decreased risk of hypertension ($p = 4.46 \times 10^{-2}$). Of the individuals in this sub-
280 analysis, 8.3% reported use of any lipid medication, all instances of which involved statins, so
281 drug class was not a confounder.

282 To investigate the possibility of time-dependent covariates, we added interaction terms to the
283 full logistic model (with age, age² and the selected covariates). For each of the selected
284 covariates, the Akaike information criterion was larger for the model containing a term that
285 interacted with age (Table 6). Thus, the evidence does not support time-dependence for any of
286 the selected covariates.

287

288 *2017 revised classification of hypertension*

289 We reanalyzed the HUFS data based on the 2017 reclassification of hypertension as SBP \geq
290 130 mm Hg or DBP \geq 80 mm Hg (Whelton et al., 2018). Under these more stringent thresholds,
291 the prevalence of hypertension increased from 48.3% to 66.1%, the median time to hypertension
292 decreased to 36 years (95% CI [33, 39]), the age at maximum hazard increased to 57 (95% CI

293 [33, 72]) years, and the lifetime risk at 85 years of age increased to 95.8% (95% CI [90.3%,
294 99.2%]). The 95% highest posterior density interval of the permanent stayer fraction included
295 zero.

296 **Discussion**

297 Time-to-event analysis of hypertension with a primary focus on modeling the baseline hazard
298 function recapitulated three known health disparities. One, the baseline hazard rate at the earliest
299 ages were higher in African Americans compared to European Americans and Mexican
300 Americans. Two, the median age when hypertension occurred was approximately a decade
301 earlier for African Americans compared to European Americans and Mexican Americans. Three,
302 by the middle of the ninth decade of life, the probability of African Americans remaining free of
303 hypertension was less than half the probability for European Americans and Mexican Americans.

304 The Bayesian model of the hazard rates revealed an inflection point, consistent with the log-
305 normal parametric distribution yielding the best fit among the parametric distributions we
306 explored. A decrease in hazard rates implies that the number of events decreases or that the
307 number of individuals at risk increases. We suggest that the simplest explanation for the
308 inflection point in hazard rates is the change in trajectory of DBP, with the decrease in DBP
309 leading to a decreased hazard of hypertension and consequently a reduced number of events. We
310 cannot rule out the possibility that there are individuals who are at less risk, although the finding
311 that the hazard rate does not reach zero (by 85 years of age) indicates the continued presence of
312 risk. It is also possible that mortality due to comorbidities begins to subside by around 60 years
313 of age, leading to reduced hazard rates of hypertension in older ages.

314 To enable model comparison between logistic regression and proportional hazards analysis,
315 we used the DIC, which is based on an estimate of dimensionality. The existence of the
316 inflection point in the baseline hazard function implies an approximately quadratic effect of age,
317 which in logistic regression is captured by an age^2 term with a negative regression coefficient.
318 Our results indicated that the proportional hazards model we used, which incorporated

319 smoothing to prevent over-parameterization, and logistic regression with age and age², provide
320 comparable fits.

321 In addition to non-zero hazard rates, evidence for the lack of existence of true controls comes
322 from the modeling of a permanent stayer fraction. We found that the permanent stayer fraction
323 essentially corresponded to the fraction of individuals who had not yet become hypertensive at
324 the last observed age. As the permanent stayer fraction did not capture any additional
325 information, models including the permanent stayer fraction had a worse fit compared to models
326 without that additional parameter. A major implication for proportional hazards analysis is that
327 the assumption that the event will occur for every individual at some point in time, unless the
328 individual dies first, is valid. A major implication for logistic regression is that controls should
329 be interpreted as individuals that have not yet become hypertensive, rather than as individuals
330 who will not become hypertensive.

331 Traditional risk factors for hypertension include excess body weight, excess dietary sodium
332 intake, reduced physical activity, deficiency of potassium, and excess alcohol intake (Chobanian
333 et al., 2003). Our results have three implications regarding these risk factors. One, we confirmed
334 that excess weight and low potassium, but not excess sodium, were risk factors in our multiple
335 regression model. It is possible that the inclusion of chloride and/or potassium could account for
336 the effects of sodium. Two, longitudinal analysis of 30 years of follow-up in the Framingham
337 Heart Study showed that the incidence rate of hypertension increased faster in females than in
338 males, with females having lower incidence under 50 years of age and higher incidence over 50
339 years of age (Dannenberg, Garrison, & Kannel, 1988). In contrast, we found that sex was not a
340 significant covariate, although low-density lipoprotein cholesterol, uric acid, and weight showed
341 sex dimorphisms. Three, we did not have data on birth weight, but our findings regarding hazard

342 rates at early ages are consistent with low birth weight being a risk factor for hypertension
343 (Lackland, Egan, & Ferguson, 2003) and African Americans having lower birth weight than
344 European Americans (David & Collins, 1997).

345 Low-density lipoprotein cholesterol has been reported to be positively associated with
346 hypertension, but this association generally does not remain significant after covariate
347 adjustment (Haffner, Ferrannini, Hazuda, & Stern, 1992; Laaksonen et al., 2008; Otsuka et al.,
348 2016; Sesso, Buring, Chown, Ridker, & Gaziano, 2005; Wildman et al., 2004). In contrast, we
349 found that low-density lipoprotein cholesterol was negatively associated with hypertension,
350 across African Americans, European Americans, and Mexican Americans. The explanation for
351 this discrepancy is unclear, but we present evidence against three possibilities. We obtained a
352 negative association in both single and multiple regression models, suggesting that the opposite
353 direction of effect was not due to other (known) covariates. We also found that the opposite
354 direction of effect was not due to a time-dependent effect. Furthermore, the opposite direction of
355 effect was not due to confounding by self-reported use of medication (*i.e.*, statins).

356 Uric acid is associated with hypertension, but whether this association is causal remains
357 unestablished. Mendelian randomization (MR) studies have provided conflicting evidence
358 regarding the causality of uric acid for hypertension, with evidence for no effect (Palmer et al.,
359 2013), protection (Sedaghat et al., 2014), and risk (Parsa et al., 2012). MR studies have reported
360 that uric acid is not causal for adiposity, chronic kidney disease, triglycerides, type 2 diabetes, or
361 obesity (Jordan et al., 2019; Kleber et al., 2015; Lyngdoh et al., 2012; Rasheed, Hughes, Flynn,
362 & Merriman, 2014). In contrast, higher adiposity, higher body mass index, lower high-density
363 lipoprotein cholesterol, and higher triglycerides are causally associated with increased uric acid
364 (Lyngdoh et al., 2012; Palmer et al., 2013; Rasheed et al., 2014; Yu, Chen, Huang, & Zeng,

365 2019). The association of uric acid with hypertension may reflect pleiotropy through linked
366 metabolic pathways, perhaps those involving lipid metabolism (Li et al., 2019).

367 The African Americans in HUFS were all recruited and enrolled in Washington, D.C. The
368 fact that neither the last completed grade of education nor income were significant predictors
369 during model selection may reflect homogeneity of study participants within one city. In
370 contrast, the African Americans in NHANES were recruited nationwide. All covariates that were
371 significant during model selection in HUFS replicated in NHANES African Americans,
372 indicating that the findings in HUFS were generalizable to the national level.

373 In summary, by Bayesian modeling of the baseline hazard function in time-to-event analysis
374 of hypertension, we found that logistic regression, if adjusted for both linear and quadratic
375 effects of age, yielded a fit comparable to proportional hazards regression. We found no evidence
376 to support the existence of true controls, suggesting that if an individual lives long enough,
377 hypertension is inevitable. Finally, we found that the combination of chloride, low-density
378 lipoprotein cholesterol, uric acid, and weight, in addition to age and age², accounted for 40.9% of
379 the variance of hypertension in African Americans, 34.8% in European Americans, and 28.3% in
380 Mexican Americans, simultaneously consistent with common risk factors among the groups and
381 heterogeneity across the groups.

382 **Competing Interests**

383 The authors declare no competing interests.

384

385 **Acknowledgements**

386 The contents of this publication are solely the responsibility of the authors and do not
387 necessarily represent the official view of the National Institutes of Health. This research was
388 supported by the Intramural Research Program of the Center for Research on Genomics and
389 Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research
390 Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for
391 Information Technology, and the Office of the Director at the National Institutes of Health
392 (1ZIAHG200362).

393 **References**

- 394 Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., . . . Rotimi, C. (2009).
395 A genome-wide association study of hypertension and blood pressure in African
396 Americans. *PLoS Genet*, 5(7), e1000564. doi:10.1371/journal.pgen.1000564
- 397 Arjas, E., & Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored
398 survival data, using the Gibbs sampler. *Stat Sinica*, 4, 505-524.
- 399 Burt, V. L., Whelton, P., Roccella, E. J., Brown, C., Cutler, J. A., Higgins, M., . . . Labarthe, D.
400 (1995). Prevalence of hypertension in the US adult population. Results from the Third
401 National Health and Nutrition Examination Survey, 1988-1991. *Hypertension*, 25(3),
402 305-313. doi:10.1161/01.hyp.25.3.305
- 403 Centers for Disease Control and Prevention, National Center for Health Statistics. (2018).
404 Underlying Causes of Death, 1999-2017. CDC WONDER Online Database.
- 405 Centers for Disease Control and Prevention. (2019). *Hypertension Cascade: Hypertension*
406 *Prevalence, Treatment and Control Estimates Among US Adults Aged 18 Years and*
407 *Older Applying the Criteria From the American College of Cardiology and American*
408 *Heart Association's 2017 Hypertension Guideline-NHANES 2013-2016.*
- 409 Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., Jr., . . .
410 National High Blood Pressure Education Program Coordinating Committee. (2003).
411 Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and
412 Treatment of High Blood Pressure. *Hypertension*, 42(6), 1206-1252.
413 doi:10.1161/01.HYP.0000107251.49515.c2
- 414 Congdon, P. (2003). *Applied Bayesian Modelling*. England: John Wiley & Sons Ltd.

- 415 Congdon, P. (2006). *Bayesian Statistical Modelling* (Second ed.). England: John Wiley and Sons
416 Ltd.
- 417 Cooper, R. S., Liao, Y., & Rotimi, C. (1996). Is hypertension more severe among U.S. blacks, or
418 is severe hypertension more common? *Ann Epidemiol*, 6, 173-180. doi:10.1016/1047-
419 2797(96)00009-9
- 420 Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc. B*, 34(2), 187-220.
- 421 Dannenberg, A. L., Garrison, R. J., & Kannel, W. B. (1988). Incidence of hypertension in the
422 Framingham Study. *Am J Public Health*, 78(6), 676-679. doi:10.2105/ajph.78.6.676
- 423 David, R. J., & Collins, J. W., Jr. (1997). Differing birth weight among infants of U.S.-born
424 blacks, African-born blacks, and U.S.-born whites. *N Engl J Med*, 337(17), 1209-1214.
425 doi:10.1056/NEJM199710233371706
- 426 Gu, Y., Sinha, D., & Banerjee, S. (2011). Analysis of cure rate survival data under proportional
427 odds model. *Lifetime Data Anal*, 17(1), 123-134. doi:10.1007/s10985-010-9171-z
- 428 Haffner, S. M., Ferrannini, E., Hazuda, H. P., & Stern, M. P. (1992). Clustering of cardiovascular
429 risk factors in confirmed prehypertensive individuals. *Hypertension*, 20(1), 38-45.
430 doi:10.1161/01.hyp.20.1.38
- 431 Jordan, D. M., Choi, H. K., Verbanck, M., Topless, R., Won, H.-H., Nadkarni, G., . . . Do, R.
432 (2019). No causal effects of serum urate levels on the risk of chronic kidney disease: A
433 Mendelian randomization study. *PLoS Med*, 16(1), e1002725.
434 doi:10.1371/journal.pmed.1002725
- 435 Kleber, M. E., Delgado, G., Grammer, T. B., Silbernagel, G., Huang, J., Krämer, B. K., . . .
436 März, W. (2015). Uric Acid and Cardiovascular Events: A Mendelian Randomization
437 Study. *J Am Soc Nephrol*, 26(11), 2831-2838. doi:10.1681/ASN.2014070660

- 438 Laaksonen, D. E., Niskanen, L., Nyysönen, K., Lakka, T. A., Laukkanen, J. A., & Salonen, J. T.
439 (2008). Dyslipidaemia as a predictor of hypertension in middle-aged men. *Eur Heart J*,
440 29(20), 2561-2568. doi:10.1093/eurheartj/ehn061
- 441 Lackland, D. T., Egan, B. M., & Ferguson, P. L. (2003). Low birth weight as a risk factor for
442 hypertension. *J Clin Hypertens*, 5(2), 133-136. doi:10.1111/j.1524-6175.2003.01353.x
- 443 Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N., & Roth, D. (1999). A more
444 accurate method to estimate glomerular filtration rate from serum creatinine: a new
445 prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med*,
446 130(6), 461-470. doi:10.7326/0003-4819-130-6-199903160-00002
- 447 Li, X., Meng, X., He, Y., Spiliopoulou, A., Timofeeva, M., Wei, W.-Q., . . . Theodoratou, E.
448 (2019). Genetically determined serum urate levels and cardiovascular and other diseases
449 in UK Biobank cohort: A phenome-wide mendelian randomization study. *PLoS Med*,
450 16(10), e1002937. doi:10.1371/journal.pmed.1002937
- 451 Luke, A., Durazo-Arvizu, R., Rotimi, C., Prewitt, T. E., Forrester, T., Wilks, R., . . . Cooper, R.
452 S. (1997). Relation between body mass index and body fat in black population samples
453 from Nigeria, Jamaica, and the United States. *Am J Epidemiol*, 145(7), 620-628.
454 doi:10.1093/oxfordjournals.aje.a009159
- 455 Lyngdoh, T., Vuistiner, P., Marques-Vidal, P., Rousson, V., Waeber, G., Vollenweider, P., &
456 Bochud, M. (2012). Serum uric acid and adiposity: deciphering causality using a
457 bidirectional Mendelian randomization approach. *PLoS One*, 7(6), e39321.
458 doi:10.1371/journal.pone.0039321
- 459 Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., . . .
460 Turner, M. B. (2016). Heart Disease and Stroke Statistics-2016 Update: A Report From

- 461 the American Heart Association. *Circulation*, 133(4), e38-e360.
- 462 doi:10.1161/CIR.0000000000000350
- 463 Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination.
- 464 *Biometrika*, 78(3), 691-692.
- 465 National Kidney Foundation. (2002). K/DOQI clinical practice guidelines for chronic kidney
- 466 disease: evaluation, classification, and stratification. *Am J Kidney Dis*, 39(2 Suppl 1), S1-
- 467 s266.
- 468 Otsuka, T., Takada, H., Nishiyama, Y., Kodani, E., Saiki, Y., Kato, K., & Kawada, T. (2016).
- 469 Dyslipidemia and the Risk of Developing Hypertension in a Working-Age Male
- 470 Population. *J Am Heart Assoc*, 5(3), e003053. oi:10.1161/JAHA.115.003053
- 471 Palmer, T. M., Nordestgaard, B. G., Benn, M., Tybjaerg-Hansen, A., Davey Smith, G., Lawlor,
- 472 D. A., & Timpson, N. J. (2013). Association of plasma uric acid with ischaemic heart
- 473 disease and blood pressure: mendelian randomisation analysis of two large cohorts. *BMJ*,
- 474 347, f4262. doi:10.1136/bmj.f4262
- 475 Parsa, A., Brown, E., Weir, M. R., Fink, J. C., Shuldiner, A. R., Mitchell, B. D., & McArdle, P.
- 476 F. (2012). Genotype-based changes in serum uric acid affect blood pressure. *Kidney Int*,
- 477 81(5), 502-507. doi:10.1038/ki.2011.414
- 478 Rasheed, H., Hughes, K., Flynn, T. J., & Merriman, T. R. (2014). Mendelian randomization
- 479 provides no evidence for a causal role of serum urate in increasing serum triglyceride
- 480 levels. *Circ Cardiovasc Genet*, 7(6), 830-837. doi:10.1161/CIRCGENETICS.114.000556
- 481 Sedaghat, S., Pazoki, R., Uitterlinden, A. G., Hofman, A., Stricker, B. H. C., Ikram, M. A., . . .
- 482 Dehghan, A. (2014). Association of uric acid genetic risk score with blood pressure: the

- 483 Rotterdam study. *Hypertension*, 64(5), 1061-1066.
- 484 doi:10.1161/HYPERTENSIONAHA.114.03757
- 485 Sesso, H. D., Buring, J. E., Chown, M. J., Ridker, P. M., & Gaziano, J. M. (2005). A prospective
486 study of plasma lipid levels and hypertension in women. *Arch Intern Med*, 165(20), 2420-
487 2427. doi:10.1001/archinte.165.20.2420
- 488 Shriner, D., Adeyemo, A., & Rotimi, C. N. (2011). Joint ancestry and association testing in
489 admixed individuals. *PLoS Comput Biol*, 7(12), e1002325.
490 doi:10.1371/journal.pcbi.1002325
- 491 Vasan, R. S., Beiser, A., Seshadri, S., Larson, M. G., Kannel, W. B., D'Agostino, R. B., & Levy,
492 D. (2002). Residual lifetime risk for developing hypertension in middle-aged women and
493 men: The Framingham Heart Study. *JAMA*, 287(8), 1003-1010.
494 doi:10.1001/jama.287.8.1003
- 495 Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Jr., Collins, K. J., Dennison
496 Himmelfarb, C., . . . Wright, J. T., Jr. (2018). 2017
497 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the
498 Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A
499 Report of the American College of Cardiology/American Heart Association Task Force
500 on Clinical Practice Guidelines. *Hypertension*, 71(6), e13-e115.
501 doi:10.1161/HYP.0000000000000065
- 502 Wildman, R. P., Sutton-Tyrrell, K., Newman, A. B., Bostom, A., Brockwell, S., & Kuller, L. H.
503 (2004). Lipoprotein levels are associated with incident hypertension in older adults. *J Am
504 Geriatr Soc*, 52(6), 916-921. doi:10.1111/j.1532-5415.2004.52258.x

- 505 World Health Organization. (13 September, 2019). Hypertension. *Fact Sheets*. Retrieved from
506 <https://www.who.int/news-room/fact-sheets/detail/hypertension>
- 507 Yu, X., Chen, H., Huang, S., & Zeng, P. (2019). Evaluation of the causal effects of blood lipid
508 levels on gout with summary level GWAS data: two-sample Mendelian randomization
509 and mediation analysis. *medRxiv*, 19006296. doi:10.1101/19006296

510 **Figure Legends**

511 Figure 1. Blood pressure among normotensive and untreated individuals. The solid red horizontal
512 line represents the diagnostic threshold of 140 mm Hg and the solid blue horizontal line
513 represents the diagnostic threshold of 90 mm Hg.

514

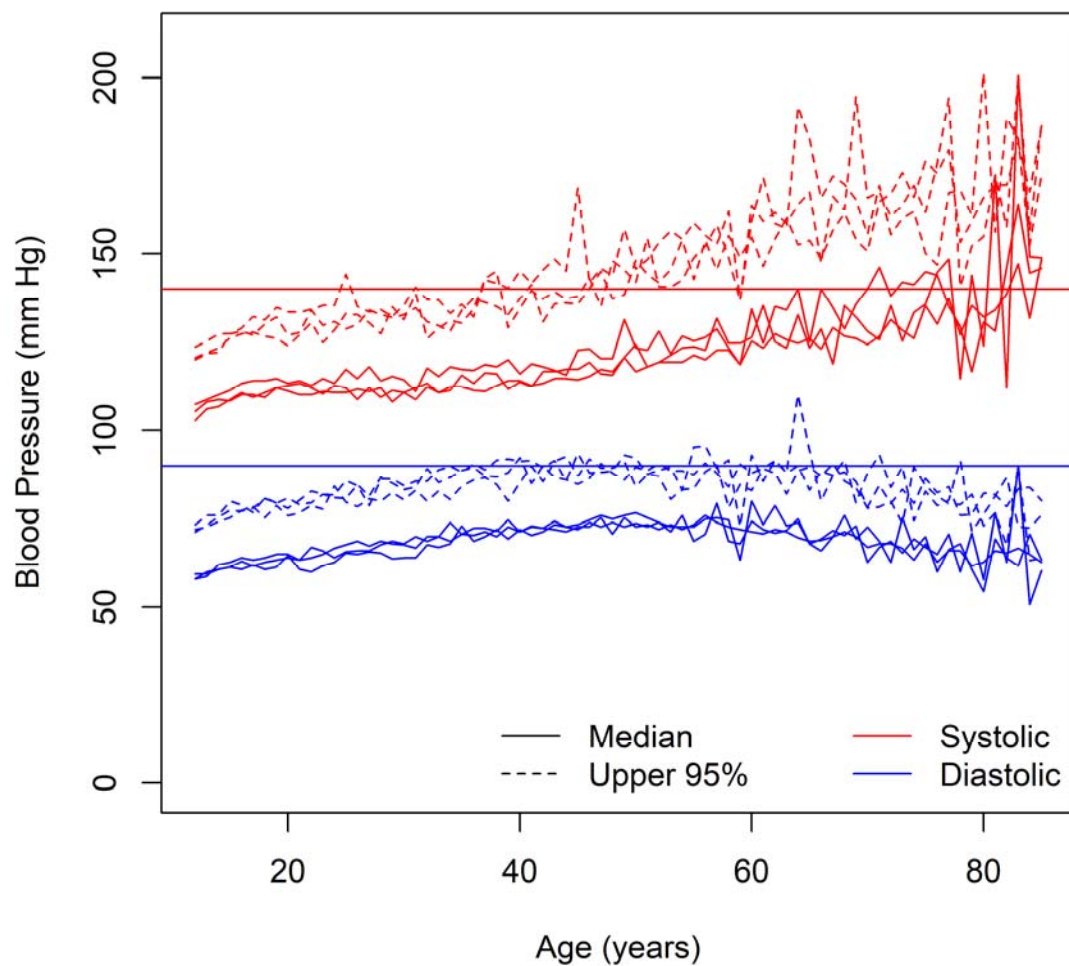
515 Figure 2. Time-to-event curves.

516

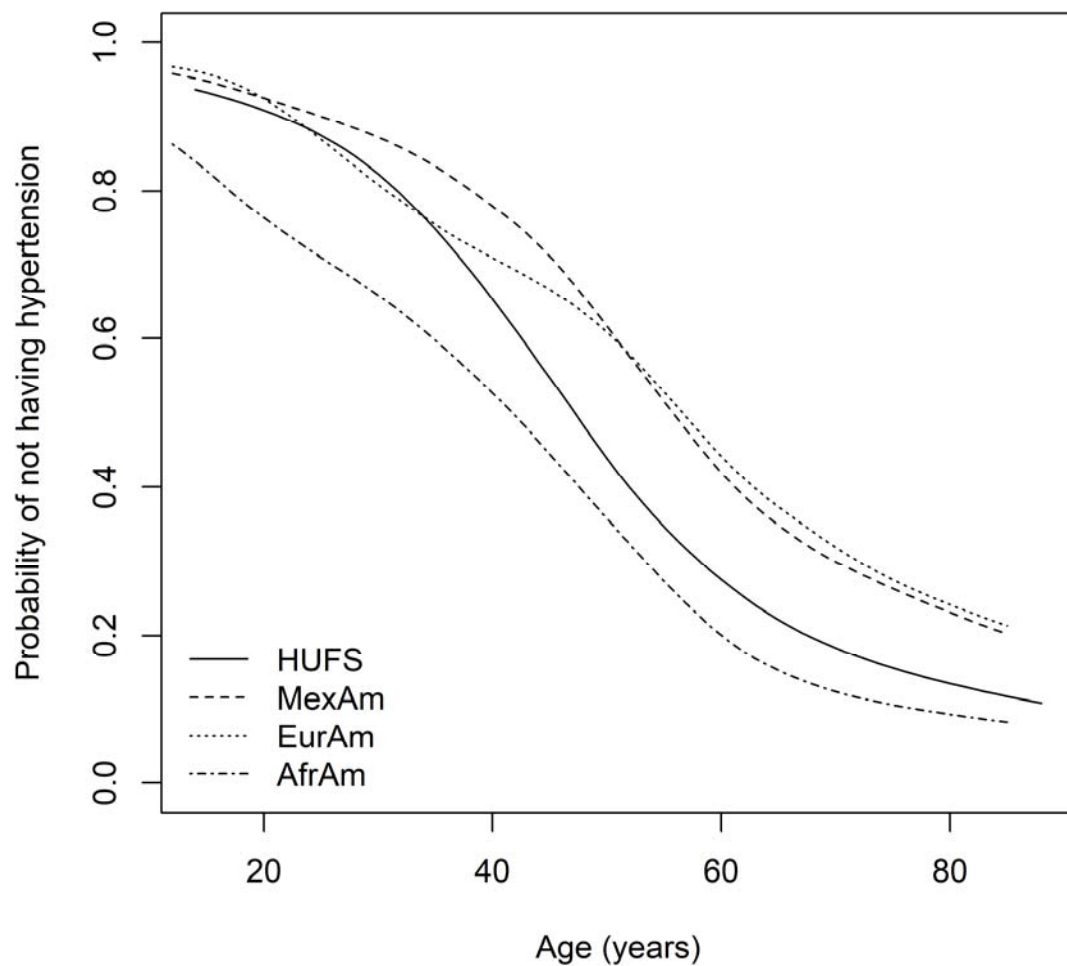
517 Figure 3. Parametric hazard functions.

518

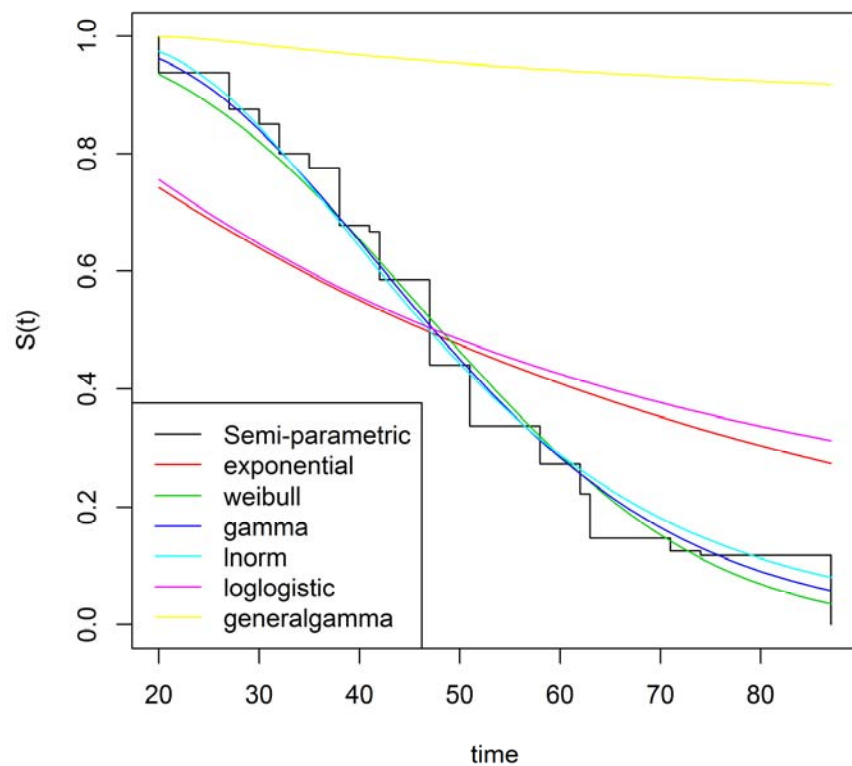
519 Figure 4. Nonparametric baseline hazard functions.



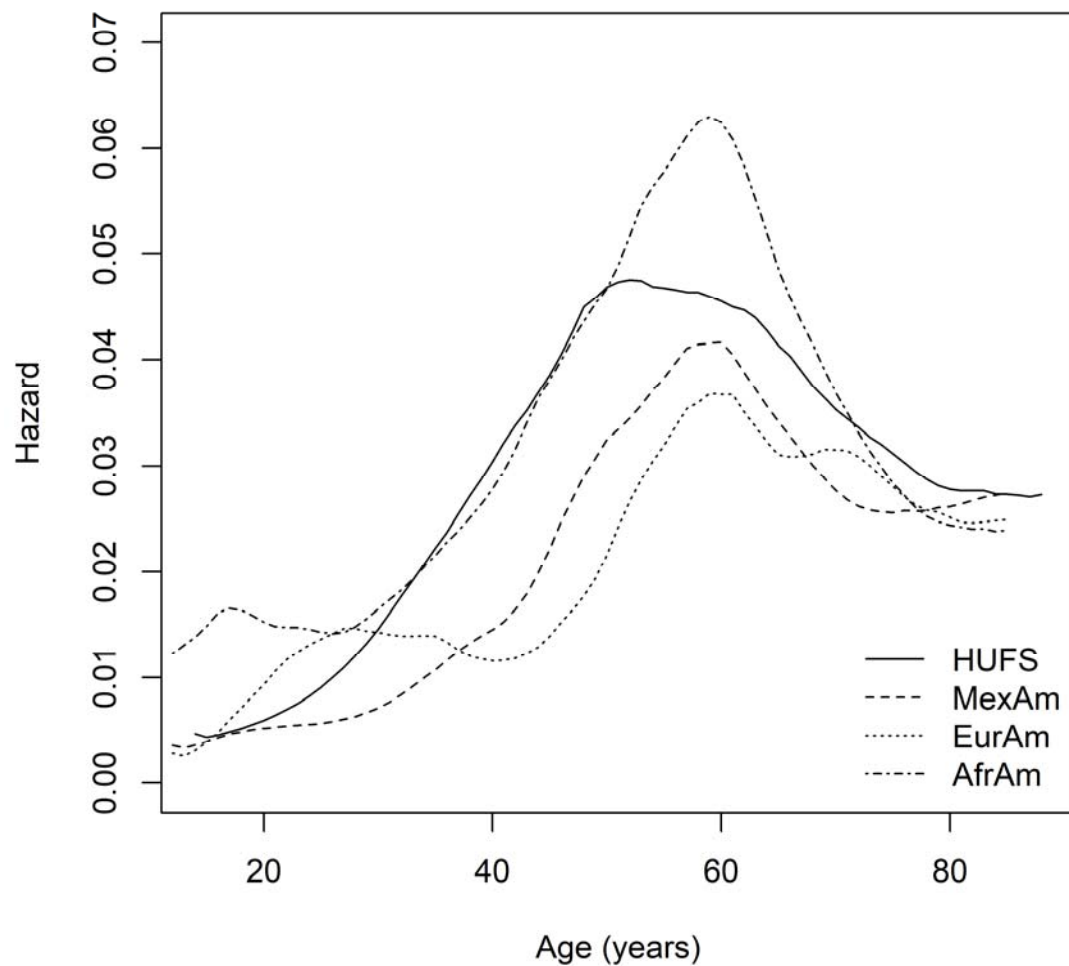
520



521



522



523

524 Table 1. Likelihoods for several parametric and nonparametric baseline hazard functions.

Distribution	<u>-ln(Likelihood)</u>
nonparametric	553.690
log-normal	580.417
gamma	580.589
Weibull	582.688
exponential	624.069
loglogistic	625.762
generalgamma	638.588

525

526 Table 2. Effective dimensionality of logistic and proportional hazards models.

Model	Parameters	DIC	pD
logistic	intercept	1406.6	1.0
logistic	intercept, age	1171.2	2.0
logistic	intercept, age, age ²	1165.3	3.0
proportional hazards	$\alpha, \beta, \lambda_0(t)$	1166.5	3.3
proportional hazards	$\alpha, \beta, \lambda_0(t), \pi$	1167.3	4.0

527

528 Table 3. Replication of covariates from model selection in African Americans.

Variable	HUFS (<i>n</i> =1,014)			NHANES (<i>n</i> =5,135)		
	Estimate	SE	<i>P</i> -value	Estimate	SE	<i>P</i> -value
Chloride (mmol/L)	-0.0253	0.0122	3.81E-02	-0.0241	0.0110	2.81E-02
Insulin (mIU/L)	0.0112	0.0054	3.61E-02	0.0115	0.0044	9.04E-03
LDL Cholesterol (mg/dL)	-0.0031	0.0013	1.75E-02	-0.0018	0.0008	3.04E-02
Potassium (mmol/L)	-0.2553	0.1016	1.20E-02	-0.5695	0.0930	9.30E-10
Uric Acid (mg/dL)	0.1713	0.0351	1.04E-06	0.1283	0.0227	1.66E-08
Weight (lbs.)	0.0033	0.0011	2.36E-03	0.0106	0.0015	1.83E-12

529

530 Table 4. Covariates from model selection in European Americans and Mexican Americans.

Variable	European Americans (<i>n</i> =10,023)			Mexican Americans (<i>n</i> =5,040)		
	Estimate	SE	<i>P</i> -value	Estimate	SE	<i>P</i> -value
Chloride (mmol/L)	-0.0656	0.0077	1.99E-17	-0.0534	0.0153	4.85E-04
Insulin (mIU/L)	0.0022	0.0051	6.76E-01	0.0111	0.0039	5.05E-03
LDL Cholesterol (mg/dL)	-0.0023	0.0006	1.38E-04	-0.0049	0.0013	1.88E-04
Potassium (mmol/L)	-0.3524	0.0683	2.49E-07	-0.2331	0.1298	7.25E-02
Uric acid (mg/dL)	0.1391	0.0165	3.16E-17	0.0833	0.0315	8.28E-03
Weight (lbs.)	0.0150	0.0016	5.61E-21	0.0164	0.0027	2.72E-09

531

532 Table 5. Sexual dimorphism among covariates in HUFS.

Variable	Male (<i>n</i> =414)	Female (<i>n</i> =600)	<i>P</i> -value
Chloride (mmol/L)	101.4 (4.4)	101.1 (5.3)	0.482
Insulin (mIU/L)	10.9 (15.3)	12.0 (11.9)	0.228
LDL Cholesterol (mg/dL)	111.3 (38.0)	117.9 (38.7)	7.28E-03
Potassium (mmol/L)	4.5 (0.7)	4.4 (1.3)	0.052
Uric Acid (mg/dL)	6.3 (1.5)	5.0 (1.5)	6.36E-35
Weight (lbs.)	194.4 (52.2)	186.1 (52.5)	0.014

533

534 Table 6. Time-dependence for selected covariates in HUF5.

Interaction term	AIC
none	1096.053
age × chloride	1096.695
age × insulin	1097.190
age × low-density lipoprotein cholesterol	1097.748
age × potassium	1097.956
age × uric acid	1097.919
age × weight	1097.013

535