

# UCLA

## UCLA Previously Published Works

### Title

Low HbA1c levels and all-cause or cardiovascular mortality among people without diabetes: the US National Health and Nutrition Examination Survey 1999-2015.

### Permalink

<https://escholarship.org/uc/item/9hc51535>

### Journal

International journal of epidemiology, 50(4)

### ISSN

0300-5771

### Authors

Inoue, Kosuke  
Nianogo, Roch  
Telesca, Donatello  
[et al.](#)

### Publication Date

2021-08-01

### DOI



10.1093/ije/dyaa263

Peer reviewed



Miscellaneous

# Low HbA1c levels and all-cause or cardiovascular mortality among people without diabetes: the US National Health and Nutrition Examination Survey 1999–2015

Kosuke Inoue,<sup>1\*</sup> Roch Nianogo,<sup>1</sup> Donatello Telesca,<sup>2</sup> Atsushi Goto ,<sup>3</sup>  
Vahe Khachadourian,<sup>4</sup> Yusuke Tsugawa,<sup>5,6</sup> Takehiro Sugiyama,<sup>7,8</sup>  
Elizabeth Rose Mayeda <sup>1</sup> and Beate Ritz<sup>1,9,10</sup>

<sup>1</sup>Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, CA, USA,

<sup>2</sup>Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA, USA,

<sup>3</sup>Department of Health Data Science, Graduate School of Data Science, Yokohama City University,

Yokohama, Japan, <sup>4</sup>Gerald & Patricia Turpanjian School of Public Health, American University of

Armenia, Yerevan, Armenia, <sup>5</sup>Division of General Internal Medicine and Health Services Research,

David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, <sup>6</sup>Department of Health Policy and

Management, UCLA Fielding School of Public Health, Los Angeles, CA, USA, <sup>7</sup>Diabetes and

Metabolism Information Center, Research Institute, National Center for Global Health and Medicine,

Tokyo, Japan, <sup>8</sup>Department of Health Services Research, Faculty of Medicine, University of Tsukuba,

Ibaraki, Japan, <sup>9</sup>Department of Environmental Health Sciences, UCLA Fielding School of Public Health,

Los Angeles, CA, USA and <sup>10</sup>Department of Neurology, UCLA David Geffen School of Medicine, Los

Angeles, CA, USA

\*Corresponding author. Department of Epidemiology, UCLA Fielding School of Public Health, 650 Charles E. Young Dr. South, Los Angeles, CA 90095, USA. E-mail: koinoue@ucla.edu

Editorial decision 19 November 2020; accepted 2 December 2020

## Abstract

**Objective:** It is unclear whether relatively low glycated haemoglobin (HbA1c) levels are beneficial or harmful for the long-term health outcomes among people without diabetes. We aimed to investigate the association between low HbA1c levels and mortality among the US general population.

**Methods:** This study includes a nationally representative sample of 39 453 US adults from the National Health and Nutrition Examination Surveys 1999–2014, linked to mortality data through 2015. We employed the parametric g-formula with pooled logistic regression models and the ensemble machine learning algorithms to estimate the time-varying risk of all-cause and cardiovascular mortality by HbA1c categories (low, 4.0 to <5.0%; mid-level, 5.0 to <5.7%; prediabetes, 5.7 to <6.5%; and diabetes,  $\geq$ 6.5% or taking antidiabetic medication), adjusting for 72 potential confounders including demographic characteristics, lifestyle, biomarkers, comorbidities and medications.

**Results:** Over a median follow-up of 7.5 years, 5118 (13%) all-cause deaths, and 1116 (3%) cardiovascular deaths were observed. Logistic regression models and machine learning algorithms showed nearly identical predictive performance of death and risk estimates. Compared with mid-level HbA1c, low HbA1c was associated with a 30% (95% CI, 16 to 48) and a 12% (95% CI, 3 to 22) increased risk of all-cause mortality at 5 years and 10 years of follow-up, respectively. We found no evidence that low HbA1c levels were associated with cardiovascular mortality risk. The diabetes group, but not the prediabetes group, also showed an increased risk of all-cause mortality.

**Conclusions:** Using the US national database and adjusting for an extensive set of potential confounders with flexible modelling, we found that adults with low HbA1c were at increased risk of all-cause mortality. Further evaluation and careful monitoring of low HbA1c levels need to be considered.

**Key words:** Low HbA1c, mortality, cardiovascular, parametric g-formula, machine learning, NHANES

#### Key Messages

- Using the large national database of US adults in 1999-2015, along with the parametric g-formula controlling for a high-dimensional set of confounders, we found that low HbA1c (4.0 to <5.0%) was associated with increased risk of all-cause mortality at 5 and 10 years of follow-up.
- The association was stronger among females than males, particularly at 5 years.
- We found no evidence that low HbA1c was associated with cardiovascular mortality risk.
- Our findings highlight that low HbA1c among people without diabetes may need to be carefully monitored.
- Future research is warranted to establish causality and identify underlying mechanisms that explain the relationship between low HbA1c and long-term health outcomes.

## Introduction

Glycated haemoglobin (HbA1c) is one of the major diagnostic biomarkers of diabetes, and it is well known that elevated HbA1c levels are associated with an increased risk of all-cause mortality as well as cardiovascular disease (CVD).<sup>1-3</sup> Additionally, previous studies have shown that relatively low HbA1c levels are associated with increased risk of all-cause mortality and CVD among people with diabetes, suggesting a potential health burden for intensive treatment of glucose levels.<sup>2-5</sup> However, the effect of relatively low HbA1c on long-term health outcomes among people without diabetes remains unclear, as results from previous studies have been inconsistent.<sup>3,6-12</sup> Relatively low HbA1c might be a proxy of malnutrition or an early stage of chronic disease.<sup>3</sup> Therefore, the observed increased risk of mortality may not reflect a causal effect of relatively low HbA1c, but instead be a reflection of the underlying poor health. Moreover, although the risk of mortality according to HbA1c levels may vary over time, previous studies have employed a Cox proportional hazard model to estimate hazard ratios, potentially violating proportional hazards assumption (i.e. relative hazards are

assumed to not vary over time). In this context, to investigate the impact of relatively low HbA1c on long-term health outcomes, analyses using flexible models that consider an array of confounders not previously accounted for and that account for time-varying risk in the estimation, are needed.

One of the major impediments for effectively addressing the causal pathways from HbA1c to mortality is the complex multifactorial interactions between blood glucose levels and sociological, biological and clinical factors. Ample evidence exists that numerous factors are associated with both HbA1c and mortality, including demographics, socioeconomic status, diet, exercise, biomarkers, comorbidities and medication.<sup>13,14</sup> Due to such a high-dimensional data structure, it has often been challenging to integrate all this information and accurately establish a causal relationship between relatively low HbA1c and adverse health outcomes. Furthermore, given that interventions using a clinical trial approach to lower glucose levels among people without diabetes would not be ethical and feasible, causal analyses using observational data are needed on this topic.

In recent years, there have been substantial advances in the application of machine learning algorithms within the framework of causal inference including the g-formula,<sup>15</sup> propensity scores<sup>16</sup> and targeted maximum likelihood estimation.<sup>17</sup> For example, the g-formula framework allows the researcher to build an outcome prediction model based on observed quantities, and then predict potential outcomes under hypothetical exposure levels.<sup>18</sup> Given the rapidly expanding availability of data, flexible machine learning algorithms may offer advantages in applying this step of the g-formula to efficiently specify the prediction model, as they have the ability to discover whether it is important to include interactions and non-linear and higher-order effects which may not be easily covered by conventional regression models.<sup>18,19</sup>

In this study, using machine learning algorithms as well as conventional logistic regression within the parametric g-formula, we estimated the effect of relatively low HbA1c on all-cause and cardiovascular mortality among the US general adult population.

## Methods

The study protocols of NHANES were approved by the NCHS Institutional Review Board [<https://www.cdc.gov/nchs/nhanes/irba98.htm>]. The UCLA IRB review was not required for the present study (IRB#20-001055) because we only used publicly available data without identifiable information.

### Study design and patients

We used data from the U.S. National Health and Nutrition Examination Survey (NHANES) 1999–2014.<sup>20</sup> NHANES is a large-scale, multistage, nationally representative survey of the civilian non-institutionalized population in the USA, conducted by the National Center for Health Statistics (NCHS). Structured interview data and physical examination results, including urine and/or blood samples, are collected continuously and released in 2-year cycles.<sup>20</sup> All participants provided informed written consent at enrolment and completed a household interview followed by a physical examination in a mobile examination centre. The response rates of NHANES during the study period were 70–80%.<sup>21</sup> The study protocols of NHANES were approved by the NCHS Institutional Review Board.<sup>22</sup>

There were 39 520 participants aged  $\geq 20$  years at enrolment for whom HbA1c was available. We excluded participants with extremely low HbA1c levels ( $< 4.0\%$ ) that could be induced by severe liver disease or haemolysis ( $n = 18$ ).<sup>23</sup> We further excluded participants who lacked time-to-event data for death due to insufficient identifying

information when linking the mortality data ( $n = 49$ ). The final analytical cohort contained 39 453 participants.

## Measurement of variables

### Exposure and diagnosed diabetes ascertainment

During visits at the mobile examination centre, phlebotomists obtained blood samples from participants according to a standardized protocol after participants fasted at least 8 h and no more than 24 h. These samples were subsequently analysed to measure HbA1c, using high-performance liquid chromatography.<sup>24</sup> We stratified participants with HbA1c within the normal range into three groups by HbA1c levels as follows: low HbA1c, 4.0 to  $< 5.0\%$ ; mid-level HbA1c (referent group), 5.0 to  $< 5.7\%$ ; and prediabetes, 5.7 to  $< 6.5\%$ , as done in previous studies.<sup>6,25</sup> We also categorized participants with HbA1c  $\geq 6.5\%$  or taking antihyperglycaemic therapies into the ‘diabetes’ group. We included them in our analysis as a positive control group for whom an increased risk of mortality is expected compared with the mid-level HbA1c group.

### Outcomes ascertainment

We used the NCHS Public-Use Linked Mortality File through 31 December 2015, to ascertain death certificate information provided by the National Death Index (NDI)<sup>26</sup> through record matching by social security number, name, date of birth, race/ethnicity, sex, state of birth and state of residence. The primary outcome for the present study was all-cause mortality, and the secondary outcome was cardiovascular mortality. The cause of death was determined based on the International Classification of Diseases, Tenth Version (ICD-10). Cardiovascular disease was classified using ICD-10 codes I00–09, I11, I13, I20–51 and I60–69.<sup>27</sup>

### Covariates

Demographic variables included age, sex (male, female), race/ethnicity (non-Hispanic White, non-Hispanic Black, Mexican-American or others), citizenship status (US or other), educational status (less than high school, high school or General Education Degree, or more than high school), health insurance status (private, public, none), marital status (single, married) and the poverty-income ratio (the ratio of the family income to the poverty threshold; range, 0–5). Smoking status (never, current, former) and physical activity levels ( $\geq$  moderate or not) were self-reported. Diet information was obtained from 24-h dietary recall collected by trained interviewers using a computer-

based interactive platform ([Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online). As comorbidities, we selected anaemia, angina, arthritis, asthma, cancer, chronic heart failure, emphysema, heart attack, hypertension, liver failure and stroke (self-reported). Use of statins, antihypertensives and antidepressants was also self-reported. Biomarkers were measured according to NHANES laboratory procedure manuals ([Supplementary Table S1](#)).<sup>28</sup> All covariates were measured or reported at enrolment.

## Statistical analyses

We employed the parametric g-formula algorithm, a generalization of the method known as standardization, to estimate the risk of death at 5 and 10 years for each HbA1c category. All models included continuous and quadratic terms for the follow-up year since NHANES enrolment, HbA1c category, an indicator variable for NHANES enrolment year, and all of the above-mentioned covariates. Missing data among covariates (28% of all participants had at least one missing value) were imputed with a random forest approach.<sup>29</sup>

In the parametric g-formula, we first fitted outcome prediction models using the exposure (HbA1c categories) and the above-mentioned 72 covariates, after arranging data into a person-time structure. To find the best predictive model for the outcome in this first step of the parametric g-formula, we developed a reference model and three machine learning models for each outcome using a training set (composed of a randomly selected 50% of the data). As the reference model, we fitted a pooled logistic regression model. In this model, we pooled observations from each follow-up year into a single dataset and employed logistic regression to predict the occurrence of each outcome. We also fitted tree-based machine learning algorithms (random forest<sup>30</sup> and gradient-boosted decision tree<sup>31</sup>) and SuperLearner.<sup>32</sup> To minimize the potential for overfitting, we performed 10-fold cross-validation for each model. After developing these prediction models, we computed the following prediction performance measures for each model in the test set (50% randomly selected samples): the area under the receiver operating characteristic curve (AUC) and confusion matrix results (i.e. sensitivity, specificity, positive predictive value and negative predictive value).

As the second step of the parametric g-formula, using the total sample, we employed the pooled logistic regression model and one of the machine learning algorithms with the best prediction performance and predicted the values for the potential outcomes under counterfactual

exposures. Then, we estimated the average marginal effect of the exposure on the outcome. We compared the estimated risk of death at 5 and 10 years had all eligible participants belonged to each of the HbA1c categories, using a risk ratio (RR) and a risk difference (RD) measure.<sup>33</sup> Robust 95% confidence intervals (CIs) were estimated by repeating these analyses on 200 bootstrapped samples. A more detailed discussion and coding for the parametric g-formula are presented elsewhere.<sup>34</sup> To evaluate mortality risks according to continuous HbA1c, we also employed restricted cubic spline models fitted with Cox proportional hazard regression with three knots (10th, 50th, and 90th percentile).

The stratum-specific analyses were conducted by age: younger (<65 years) versus older (≥65 years) and by sex: male versus female. *P*-value for heterogeneity was calculated using the method proposed by Altman and Bland.<sup>35</sup> We also performed the following sensitivity analyses to assess the robustness of our findings: (i) we performed complete case analysis with NHANES survey weights to account for unequal probabilities of selecting NHANES participants and non-response of those eligible and approached ( $n=28\,312$ )<sup>36</sup>; and (ii) we re-analysed data restricting participants to those with haemoglobin ≥13 g/dl in males and haemoglobin ≥12 g/dl in females who did not report anaemia, because HbA1c could be affected by anaemia ( $n=34\,740$ ).<sup>37,38</sup> All statistical analyses were conducted using R version 4.0.2.

## Results

The mean age of participants was 49.5 years (standard deviation, 18.3; median, 48; interquartile range, 34 to 64), and 48.1% were male. Demographic characteristics of participants across HbA1c groups are shown in [Table 1](#) and [Supplementary Table S1](#).

### Prediction of all-cause mortality and cardiovascular mortality

Overall, the median duration of follow-up was 7.5 years, and 5118 all-cause deaths and 1116 cardiovascular deaths were identified ([Supplementary Table S2](#), available as [Supplementary data](#) at *IJE* online). Kaplan-Meier survival curves by HbA1c levels are shown in [Supplementary Figure S1](#), available as [Supplementary data](#) at *IJE* online. All of the candidate algorithms, including the pooled logistic regression model, showed high prediction performance of all-cause mortality and cardiovascular mortality, with AUCs ranging from 0.86 to 0.90 ([Table 2](#)). Sensitivity, specificity and negative predictive value were also similar

**Table 1** Baseline characteristics according to glycated haemoglobin (HbA1c) levels in NHANES 1999–2014<sup>a,b</sup>

	HbA1c levels within the normal range			HbA1c $\geq$ 6.5% or antidiabetic medication
	Low (4.0% to <5.0%)	Mid-level (5.0% to <5.7%)	Prediabetes (5.7% to <6.5%)	
Total, <i>n</i>	4314	23 953	5921	5265
HbA1c %, median (IQR)	4.8 (4.7-4.9)	5.4 (5.2-5.5)	5.9 (5.8-6.1)	7.0 (6.5-8.2)
Age (years)	36.5 $\pm$ 15.0	46.8 $\pm$ 17.9	59.1 $\pm$ 15.6	61.7 $\pm$ 13.4
Sex (female), %	59.1	51.7	50.4	48.5
Race/ethnicity, %				
Non-Hispanic White	54.0	50.1	40.6	36.3
Non-Hispanic Black	17.4	17.2	27.0	26.8
Mexican-American	16.3	18.3	16.4	21.4
Others	12.3	14.4	16.0	15.5
Education status, %				
Less than 9th grade	7.3	10.7	16.4	22.0
9 <sup>th</sup> -11th grade	13.7	15.1	17.1	19.3
High school or GED	21.4	23.4	24.5	22.3
Higher than high school	57.6	50.8	42.0	36.4
Married, %	49.7	53.6	54.5	56.0
Smoking, %				
Never	59.4	54.2	49.8	48.8
Current	21.9	22.6	20.6	16.7
Former	18.7	23.2	29.6	34.5
Insurance status, %				
Private	33.1	34.0	38.5	32.9
Public	42.1	42.1	41.5	52.3
Uninsured	24.8	23.9	20.0	14.8
Poverty-income ratio	2.64 $\pm$ 1.66	2.62 $\pm$ 1.65	2.45 $\pm$ 1.57	2.23 $\pm$ 1.50
Physical activity levels ( $\geq$ moderate), %	68.7	65.2	56.8	47.6
Anaemia, %	4.7	3.3	3.5	6.3
Angina, %	0.7	2.1	3.9	7.8
Arthritis, %	12.2	22.5	36.4	44.3
Asthma, %	14.0	12.6	12.3	14.6
Cancer, %	4.2	8.1	12.2	12.9
Chronic heart failure, %	1.0	1.9	4.1	9.8
Emphysema, %	0.6	1.6	3.1	3.6
Heart attack, %	1.3	2.9	6.2	11.1
Hypertension, %	15.9	27.2	47.1	65.5
Liver failure, %	2.9	3.1	3.6	5.8
Stroke, %	1.7	2.6	5.2	8.7
Statin use, %	2.8	9.7	23.8	42.3
Antihypertensive use %	8.2	18.1	38.3	58.2
Antidepressant use, %	7.9	8.8	9.6	14.5
Systolic blood pressure (mmHg)	116.8 $\pm$ 16.5	123.3 $\pm$ 19.3	131.9 $\pm$ 20.8	134.2 $\pm$ 21.4
Diastolic blood pressure (mmHg)	67.8 $\pm$ 12.9	70.8 $\pm$ 12.9	71.7 $\pm$ 14.1	68.8 $\pm$ 15.6
Waist (cm)	91.6 $\pm$ 14.1	96.1 $\pm$ 14.9	103.4 $\pm$ 15.2	108.9 $\pm$ 15.6
BMI (kg/m <sup>2</sup> )	26.5 $\pm$ 5.5	28.0 $\pm$ 6.2	30.5 $\pm$ 6.9	32.4 $\pm$ 7.4

NHANES, National Health and Nutrition Examination Survey; IQR, interquartile range; GED, General Educational Development; BMI, body mass index.

<sup>a</sup>Data are presented as count (percentage) or mean  $\pm$  SD, otherwise specified.

<sup>b</sup>Other variables (dietary information and biomarkers) are shown in [Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online.

**Table 2** Predictive ability of pooled logistic regression model, tree-based algorithms and SuperLearner for all-cause and cardiovascular mortality

Models	AUC	Sensitivity	Specificity	PPV	NPV
Outcome: All-cause mortality					
Logistic regression model	0.87	0.80	0.78	0.05	>0.99
Random forest	0.86	0.91	0.54	0.03	>0.99
Gradient boosting	0.86	0.54	0.92	0.09	>0.99
SuperLearner	0.87	0.85	0.75	0.04	>0.99
Outcome: Cardiovascular mortality					
Logistic regression model	0.90	0.84	0.82	0.01	>0.99
Random forest	0.88	0.94	0.61	0.01	>0.99
Gradient boosting	0.89	0.42	0.96	0.03	>0.99
SuperLearner	0.90	0.90	0.74	0.01	>0.99

Each model included all 72 covariates listed in [Table 1](#) and in [Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online (e.g. demographic characteristics, diet, exercise, comorbidities, biomarkers, medications). Confusion matrix results (i.e. sensitivity, specificity, PPV and NPV) were at the cut-off value of the prevalence of each outcome. PPVs were generally low for all algorithms due to the small number of outcomes overall (i.e. all-cause and cardiovascular mortality).

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

for pooled logistic regression and SuperLearner, whereas random forest yielded a lower specificity and gradient boosting yielded a lower sensitivity.

### Estimated risk of all-cause mortality and cardiovascular mortality at 5 and 10 years

After adjusting for all potential confounders including demographic characteristics, diet, exercise, comorbidities, biomarkers and medications using a pooled logistic regression model within the parametric g-formula, compared with the mid-level HbA1c group, the low HbA1c group showed a 30% (95% CI, 16 to 48) and a 12% (95% CI, 3 to 22) increased risk of all-cause mortality at 5 years and 10 years of follow-up, respectively. On the absolute scale, the low HbA1c group showed a 1.83 (95% CI, 1.02 to 2.97) and a 1.66 (95% CI, 0.35 to 3.00) percentage points increase for all-cause mortality risk at 5 and 10 years of follow-up, respectively ([Figure 1](#), [Table 3](#)). We found no evidence of an association between HbA1c and cardiovascular mortality. The findings were qualitatively consistent when we used SuperLearner (which showed the highest predictive performance among the three machine learning algorithms) within the parametric g-formula ([Figure 1](#)).

We did not find evidence for an association in the prediabetes group with all-cause mortality at 5 and 10 years of follow-up ([Figure 1](#); [Supplementary Table S3](#), available as [Supplementary data](#) at *IJE* online). As expected, the diabetes group showed an increased risk of all-cause and cardiovascular mortality regardless of model specifications (i.e. either the pooled logistic regression model or SuperLearner) ([Figure 1](#); [Supplementary Table S4](#),

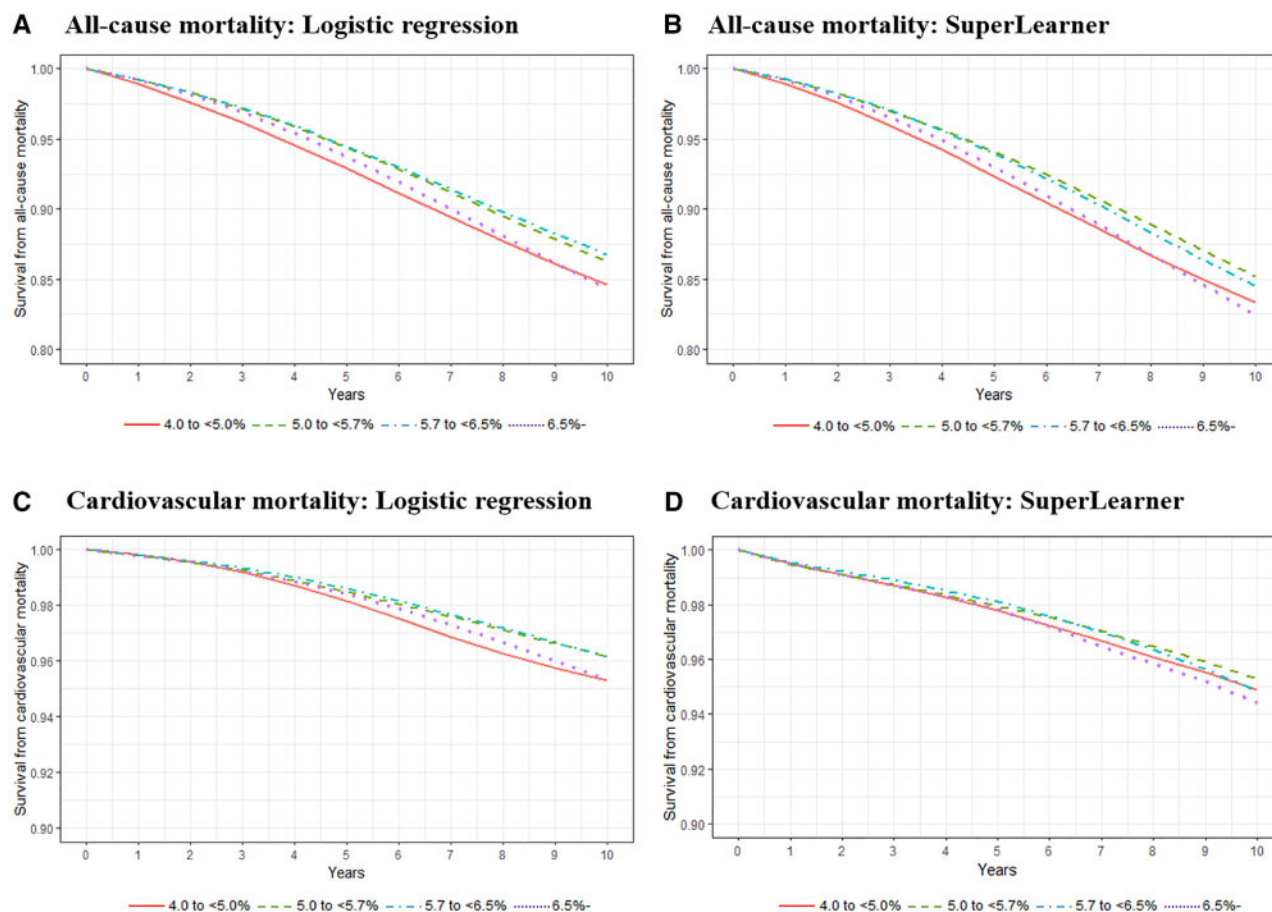
available as [Supplementary data](#) at *IJE* online). These associations were also found when restricted cubic spline curves were fitted with Cox proportional hazard regression ([Supplementary Figure S2](#), available as [Supplementary data](#) at *IJE* online).

### Stratum-specific analysis by age and sex

We found similar associations for low HbA1c and all-cause mortality in the younger and the older population on the relative risk scale ([Figure 2](#); [Supplementary Table S5](#), available as [Supplementary data](#) at *IJE* online). When we stratified by sex, we estimated increased risk of all-cause mortality for low HbA1c among females but not among males ([Figure 2](#); [Supplementary Table S5](#), available as [Supplementary data](#) at *IJE* online). We found no evidence for an association between low HbA1c and cardiovascular mortality in any subgroups stratified by age and sex ([Supplementary Table S6](#), available as [Supplementary data](#) at *IJE* online).

### Sensitivity analyses

The results for all-cause mortality did not substantially change when we performed complete-case analysis using NHANES survey weights ([Supplementary Table S7](#), available as [Supplementary data](#) at *IJE* online) and when we re-analysed the data after restricting to participants without anaemia, particularly at 5 years ([Supplementary Table S8](#), available as [Supplementary data](#) at *IJE* online).



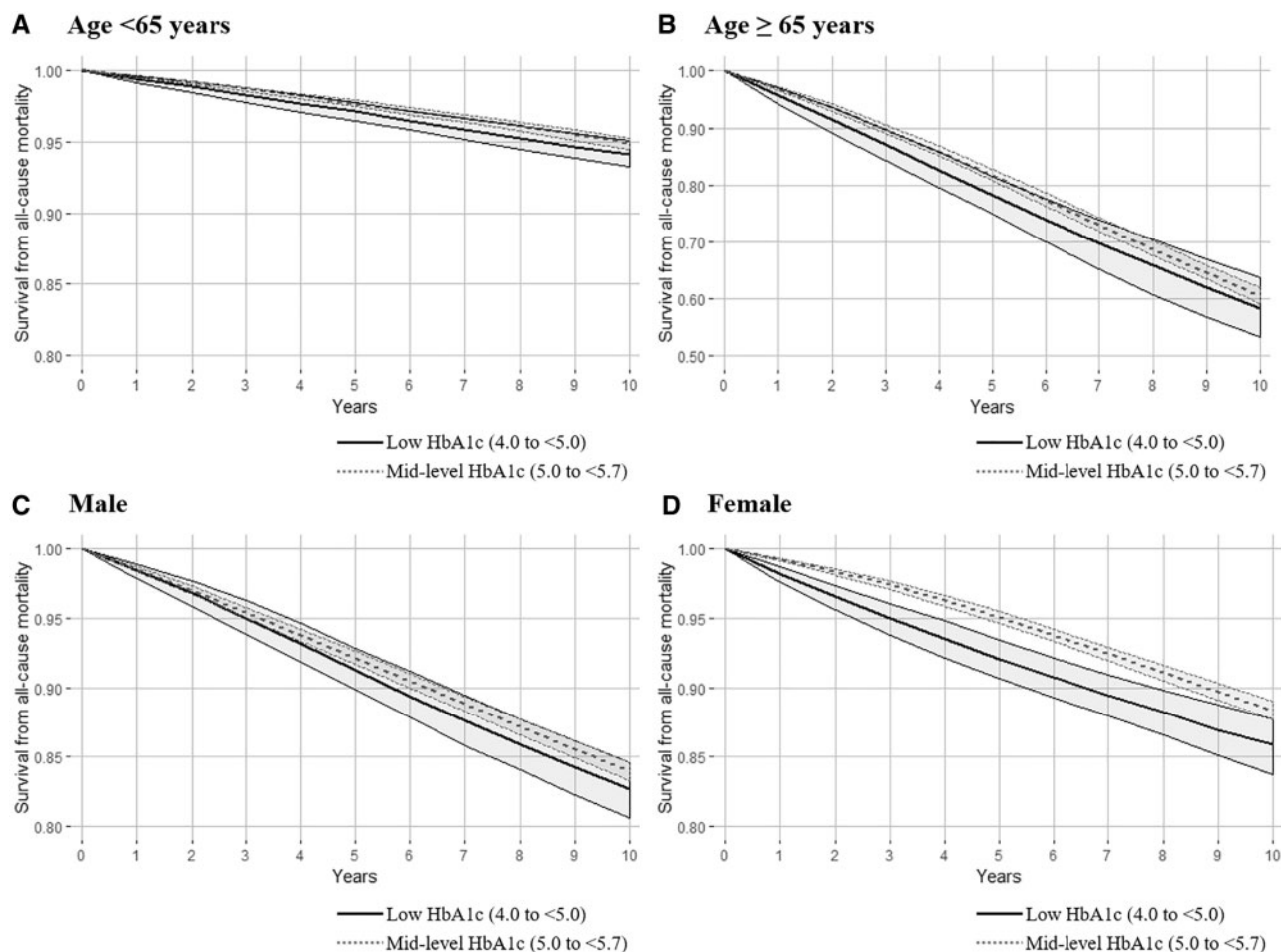
**Figure 1** Adjusted all-cause and cardiovascular mortality risk according to glycated haemoglobin (HbA1c) levels using parametric g-formula with pooled logistic regression models and SuperLearner. The ranges of the survival rate (Y-axis) presented in figures were 0.8 to 1.0 for all-cause mortality and 0.9 to 1.0 for cardiovascular mortality. Robust 95% confidence intervals (CIs) for each HbA1c category estimated by bootstrapping (in the pooled logistic regression model) are presented in [Table 3](#); and in [Supplementary Tables S3 and S4](#), available as [Supplementary data](#) at *IJE* online

**Table 3** Adjusted all-cause and cardiovascular mortality risk ratio and risk difference at 5 and 10 years among participants with low glycated haemoglobin (HbA1c; 4.0% to <5.0%) compared with those with mid-level HbA1c (5.0% to <5.7%) using parametric g-formula with pooled logistic regression models<sup>a</sup>

Outcomes	Follow-up periods	
	5 years	10 years
<b>All-cause mortality</b>		
Number of events/total number of participants among low HbA1c group	149/3453 (4.3%)	249/2205 (11.3%)
Number of events/total number of participants among mid-level HbA1c group	1125/18024 (6.2%)	2104/10906 (19.3%)
Adjusted risk ratio (95% CI)	1.30 (1.16 to 1.48)	1.12 (1.03 to 1.22)
Adjusted risk difference (95% CI)	+1.83% (1.02 to 2.97)	+1.66% (0.35 to 3.00)
<b>Cardiovascular mortality</b>		
Number of events/total number of participants among low HbA1c group	23/3327 (0.7%)	46/2002 (2.3%)
Number of events/total number of participants among mid-level HbA1c group	263/17162 (1.5%)	452/9254 (4.9%)
Adjusted risk ratio (95% CI)	1.17 (0.80 to 1.59)	1.21 (0.92 to 1.54)
Adjusted risk difference (95% CI)	+0.28% (-0.35 to 0.99)	+0.83% (-0.32 to 2.01)

<sup>a</sup>200 iterations were performed for bootstrapping to estimate 95% confidence interval.





**Figure 2** Adjusted all-cause mortality risk among participants low glycated haemoglobin (HbA1c; 4.0% to <5.0%) and mid-level HbA1c (5.0% to <5.7%) stratified by age and sex using parametric g-formula with pooled logistic regression models. The ranges of the survival rate (Y-axis) presented in figures were 0.8 to 1.0 except for the older population (range 0.5 to 1.0) who had a higher mortality rate than other groups. Robust 95% confidence intervals (CIs) were estimated by repeating these analyses on 200 bootstrapped samples

## Discussion

Using a nationally representative database of US adults and controlling for potential confounders (e.g. demographic characteristics, diet, exercise, comorbidities, biomarkers and medications) with several statistical algorithms, we found that individuals with low HbA1c (4.0% to <5.0%) were experiencing an increased risk of all-cause mortality at 5 and 10 years of follow-up compared with those with mid-level HbA1c (5.0% to <5.7%). This relationship was stronger for females than males. We found no evidence that low HbA1c was associated with cardiovascular mortality.

The question of whether relatively low HbA1c is beneficial or harmful for people without diabetes has been actively debated for a long time. Although some previous studies have reported associations between relatively low HbA1c and increased CVD and mortality,<sup>7–10</sup> its clinical and biological relevance has remained unclear. As HbA1c

is becoming more frequently (and routinely) measured based on clinical guidelines,<sup>39</sup> the chance that clinicians detect people with relatively low HbA1c might increase, and there is a need to answer this long-debated question about the potential burden of relatively low HbA1c levels on health. In this context, our findings provide new evidence, indicating that we may need to carefully monitor people with relatively low HbA1c.

Our findings were consistent with previous cohort studies investigating the association between low HbA1c and all-cause mortality among people without diabetes.<sup>7,9,10</sup> A recent study among US adults aged  $\geq 50$  years without diabetes, from the Health and Retirement Study, also reported a reverse J-shaped association between HbA1c and all-cause mortality, but not cardiovascular mortality.<sup>3</sup> Given the null association between relatively low HbA1c and all-cause mortality over 15 years of follow-up among Japanese adults,<sup>11</sup> the association might be heterogeneous across

race/ethnicity, which requires further investigation. A previous study in NHANES also did not find an association between relatively low HbA1c and all-cause mortality over a median follow-up of 9 years, but this analysis only included 7333 participants aged  $\geq 65$  years enrolled before 2004.<sup>6</sup> Furthermore, most of these studies might have suffered from limitations including a limited number of covariates adjusted for (i.e. unmeasured confounding bias), violations of proportional hazard assumptions, or a relatively short follow-up period. Our study, using high-dimensional data from a national survey with long follow-up time and flexible statistical modelling, overcomes some of these limitations and therefore helps to advance the current state of knowledge about the potential impact of low HbA1c on mortality.

Underlying biological mechanisms for the association between relatively low HbA1c and mortality have still not been established. Poor health status (e.g. malnutrition, unfavourable profiles of red blood cell-related factors, inflammation, decreased liver function or an early stage of chronic disease) has been proposed as an explanation for the association between relatively low HbA1c and mortality among people without diabetes.<sup>8,23,25</sup> Hypoglycaemia induces sympatho-adrenal activation, inflammation and endothelial dysfunction, all of which could lead to chronic and cardiometabolic diseases.<sup>40</sup> Given these proposed mechanisms and the fact that we did not find evidence for an association between low HbA1c and cardiovascular mortality (likely due to insufficient statistical power), further investigations with a larger sample size focusing on a high CVD risk population are warranted. Furthermore, we found a stronger association among females than males, particularly at 5-year follow-up. This was mainly due to the higher mortality risk for mid-level HbA1c among males than females, although both sexes showed similar mortality risks for low HbA1c. Given sex differences in glucose metabolism,<sup>41</sup> our findings also indicate the importance of evaluating HbA1c by sex.

As expected based on ample previous evidence,<sup>42–44</sup> the diabetes group showed an increased risk of all-cause mortality. A cohort study of one million US adults reported diabetes to be associated with a higher risk of mortality for several diseases such as CVD, cancer, respiratory dysfunction, digestive diseases, genitourinary disorders and even accidents.<sup>42</sup> Although mortality and incidence of cardiovascular events among people with diabetes have decreased over the past two decades, mainly owing to remarkable advancements in the treatment of CVD and diabetes,<sup>43,44</sup> our findings highlight that there is still a need for further improvement of diabetes management to avoid complications.

The present study has three major strengths. First, our study used a large, nationally representative sample of the US general population with linkage to the most updated national mortality database. Second, we applied the parametric g-formula that does not require the proportional hazard assumption and allows us to estimate clinically meaningful absolute/relative risks.<sup>34</sup> This approach also enabled us to estimate risks at different time points (i.e. 5 and 10 years of follow-up). Last, we employed several ensemble machine learning algorithms to build the outcome prediction models in the first step of the parametric g-formula, including an ensemble method called SuperLearner that combines multiple machine learning algorithms with weights estimated to maximize performance.<sup>32</sup> Our results suggest that a conventional logistic regression modelling approach, which has much lower computational cost than machine learning algorithms, may well suffice to answer our research question, as risk factors are well known. But even in a scenario such as ours, comparing the findings from both logistic regression models and machine learning algorithms may provide a transparent approach to addressing the potential for bias due to model mis-specification.

Our study has limitations. First, although we included an extensive set of covariates, there is always a possibility for bias due to unmeasured confounding in observational study design. For example, low-density lipoprotein cholesterol levels were not available for many NHANES participants; therefore, we used statin use as a proxy for dyslipidaemia. The lack of detailed information on diabetes, such as family history and antibodies, may also limit the interpretation of the diabetes and mortality association, but does not affect the interpretation of our primary outcomes (i.e. low HbA1c and mortality among people without diabetes). The wide range of age in the present study may also raise a concern about residual confounding by age even after adjusting for age. However, as increasing age is negatively associated with low HbA1c and positively associated with mortality risks, such bias would be expected to cause an underestimation of the effect of low HbA1c on mortality. Given that clinical trials may not be feasible and ethical on this topic, future studies using other epidemiological approaches such as Mendelian randomization should be considered to validate our findings. Second, diet, lifestyle and comorbidities may have been mis-measured because these variables were self-reported. Third, as HbA1c was only measured at baseline, we had no information about how changes in HbA1c may or may not contribute to the increased risk of mortality. Last, covariate information was also only available at baseline. Thus, the exposure-confounders relationships were not well defined temporally, and we cannot rule out the possibility of reverse causation and over-adjustment. Further

longitudinal studies with measurements of HbA1c and other covariates at multiple time points are needed to overcome this limitation.

In conclusion, low HbA1c was associated with an increased risk of all-cause mortality at 5 and 10 years of follow-up among US adults. Our findings may indicate the importance of carefully monitoring individuals with relatively low HbA1c without diabetes, as well as individuals with diabetes in clinical practice. A better understanding of this relationship would enable health care professionals to design effective public health interventions to reduce the risk of long-term adverse health outcomes that may be related to relatively low HbA1c.

## Supplementary data

Supplementary data are available at *IJE* online.

## Funding

K.I. was supported by Toffler Award at the Department of Epidemiology, UCLA, the Burroughs Wellcome Fund Interschool Training Program in Chronic Diseases (BWF-CHIP), Honjo International Foundation Scholarship, and National Institutes of Health (NIH)/NIDDK grant F99DK126119. K.I. and Y.T. were supported by Amazon Web Services/ Computation Medicine Award, UCLA. E.R.M. was supported by NIH/NIA grant R00AG053410. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; or decision to submit the manuscript for publication. This article does not necessarily represent the views and policies of the NIH.

## Data availability

All data are publicly available at [<https://wwwn.cdc.gov/nchs/nhanes/Default.aspx>].

## Author contributions

K.I. developed the study concept, analysed and interpreted the data and wrote the manuscript. R.N., D.T., A.G. and T.S. developed the study concept, interpreted the data and reviewed/edited the manuscript. Y.T. and E.R.M. contributed to the discussion and reviewed/edited the manuscript. B.R.R. developed the study concept, supervised, and reviewed/edited the manuscript.

## Conflict of interest

None declared.

## References

- Forbes A, Murrells T, Mulnier H, Sinclair AJ. Mean HbA1c, HbA1c variability, and mortality in people with diabetes aged 70 years and older: a retrospective cohort study. *Lancet Diabetes Endocrinol* 2018;6:476–86.
- Currie CJ, Peters JR, Tynan A *et al.* Survival as a function of HbA(1c) in people with type 2 diabetes: a retrospective cohort study. *Lancet* 2010;375:481–89.
- Li F-R, Zhang X-R, Zhong W-F *et al.* Glycated Haemoglobin and all-cause and cause-specific mortality among adults with and without diabetes. *J Clin Endocrinol Metab* 2019;104:3345–54.
- Saydah S, Tao M, Imperatore G, Gregg E. GHb level and subsequent mortality among adults in the U.S. *Diabetes Care* 2009;32:1440–46.
- Goto A, Arah OA, Goto M, Terauchi Y, Noda M. Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis. *BMJ* 2013;347:f4533.
- Palta P, Huang ES, Kalyani RR, Golden SH, Yeh H-C. Haemoglobin A1c and mortality in older adults with and without diabetes: results from the National Health and Nutrition examination surveys (1988-2011. ). *Diabetes Care* 2017;40:453–60.
- Selvin E, Steffes MW, Zhu H *et al.* Glycated haemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med* 2010;362:800–11.
- Carson AP, Fox CS, McGuire DK *et al.* Low haemoglobin A1c and risk of all-cause mortality among US adults without diabetes. *Circ Cardiovasc Qual Outcomes* 2010;3:661–67.
- Paprott R, Schaffrath Rosario A, Busch MA *et al.* Association between haemoglobin A1c and all-cause mortality: results of the mortality follow-up of the German National Health Interview and Examination Survey 1998. *Diabetes Care* 2015;38:249–56.
- Brewer N, Wright CS, Travier N *et al.* A New Zealand linkage study examining the associations between A1C concentration and mortality. *Diabetes Care* 2008;31:1144–49.
- Sakurai M, Saitoh S, Miura K *et al.*; for the NIPPON DATA90 Research Group. HbA1c and the risks for all-cause and cardiovascular mortality in the general Japanese population: NIPPON DATA90. *Diabetes Care* 2013;36:3759–65.
- Schöttker B, Rathmann W, Herder C *et al.*; on behalf of the CHANCES group. HbA1c levels in non-diabetic older adults. No J-shaped associations with primary cardiovascular events, cardiovascular and all-cause mortality after adjustment for confounders in a meta-analysis of individual participant data from six cohort studies. *BMC Med* 2016;14:26.
- US Department of Health and Human Services, Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2020*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.
- NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* 2016;387:1513–30.
- Austin PC. Using ensemble-based methods for directly estimating causal effects: an investigation of tree-based G-computation. *Multivariate Behav Res* 2012;47:115–35.
- Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Stat Med* 2010;29:337–46.

17. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol* 2017;**185**:65–73.
18. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 2011;**173**:731–38.
19. Kuhn M, Johnson K. *Applied Predictive Modeling*. Vol 26. 2013. <http://appliedpredictivemodeling.com> 11 January 2020, date last accessed.
20. National Center for Health Statistics. *Centers for Disease Control and Prevention National Health and Nutrition Examination Survey*. 2019. [www.cdc.gov/nchs/nhanes.htm](http://www.cdc.gov/nchs/nhanes.htm) (11 February 2019, date last accessed).
21. National Center for Health Statistics. Centers for Disease Control and Prevention. *NHANES Response Rates and Population Totals*. <https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx> (18 February 2019, date last accessed).
22. National Center for Health Statistics. Centers for Disease Control and Prevention NCHS Research Ethics Review Board (ERB) Approval. 2017. [www.cdc.gov/nchs/nhanes/irba98.htm](http://www.cdc.gov/nchs/nhanes/irba98.htm) (18 February 2019, date last accessed).
23. Christman AL, Lazo M, Clark JM, Selvin E. Low glycosylated haemoglobin and liver disease in the U.S. population. *Diabetes Care* 2011;**34**:2548–50.
24. Selvin E, Parrinello CM, Sacks DB, Coresh J. Trends in prevalence and control of diabetes in the United States, 1988–1994 and 1999–2010. *Ann Intern Med* 2014;**160**:517–25.
25. Aggarwal V, Schneider ALC, Selvin E. Low haemoglobin A1c in nondiabetic adults: an elevated risk state? *Diabetes Care* 2012;**35**:2055–60.
26. Fillenbaum GG, Burchett BM, Blazer DG. Identifying a national death index match. *Am J Epidemiol* 2009;**170**:515–18.
27. Inoue K, Ritz B, Brent GA, Ebrahimi R, Rhee CM, Leung AM. Association of subclinical hypothyroidism and cardiovascular disease with mortality. *JAMA Netw Open* 2020;**3**:e1920745.
28. Centers for Disease Control and Prevention. *NHANES MEC Laboratory Procedures Manual*. [https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/manuals/2013\\_MEC\\_Laboratory\\_Procedures\\_Manual.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/manuals/2013_MEC_Laboratory_Procedures_Manual.pdf) (19 October 2019, date last accessed).
29. Stekhoven DJ, Bühlmann P. MissForest - non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;**28**:112–18.
30. RDRR. *Ranger: A Fast Implementation of Random Forests Version 0.11.2 from CRAN*. <https://rdr.io/cran/ranger/> (19 October 2019, date last accessed).
31. RDRR. *xgboost: Extreme Gradient Boosting Version 0.90.0.2 from CRAN*. <https://rdr.io/cran/xgboost/> (19 October 2019, date last accessed).
32. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;**6**:Article25.
33. Dickerman BA, Giovannucci E, Perner CH, Mucci LA, Hernán MA. Guideline-based physical activity and survival among US men with nonmetastatic prostate cancer. *Am J Epidemiol* 2019;**188**:579–86.
34. Keil AP, Edwards JK, Richardson DR, Naimi AI, Cole SR. The parametric G-formula for time-to-event data: towards intuition with a worked example. *Epidemiology* 2014;**25**:889–97.
35. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;**326**:219.
36. NHANES. *Continuous NHANES Web Tutorial—Specifying Weighting Parameters*. <https://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/intro.htm> (18 February 2019, date last accessed).
37. World Health Organization. *Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity*. 2011. <https://www.who.int/vmnis/indicators/haemoglobin.pdf> (14 March 2020, date last accessed).
38. Kim C, Bullard KM, Herman WH, Beckles GL. Association between iron deficiency and A1C levels among adults without diabetes in the National Health and Nutrition Examination Survey, 1999–2006. *Diabetes Care* 2010;**33**:780–85.
39. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020. *Diabetes Care* 2020;**43**:S14–S31
40. Davis IC, Ahmadizadeh I, Randell J, Younk L, Davis SN. Understanding the impact of hypoglycemia on the cardiovascular system. *Expert Rev Endocrinol Metab* 2017;**12**:21–33.
41. Varlamov O, Bethea CL, Roberts CT. Sex-specific differences in lipid and glucose metabolism. *Front Endocrinol (Lausanne)* 2014;**5**:241.
42. Campbell PT, Newton CC, Patel AV, Jacobs EJ, Gapstur SM. Diabetes and cause-specific mortality in a prospective cohort of one million U.S. adults. *Diabetes Care* 2012;**35**:1835–44.
43. Rawshani A, Rawshani A, Franzén S *et al*. Mortality and cardiovascular disease in Type 1 and Type 2 diabetes. *N Engl J Med* 2017;**376**:1407–18.
44. Gregg EW, Cheng YJ, Srinivasan M *et al*. Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet* 2018;**391**:2430–40.