



# Systolic Blood Pressure and Cardiovascular Risk in Patients With Diabetes: A Prospective Cohort Study

Shishir Rao<sup>1</sup>, Yikuan Li, Milad Nazarzadeh<sup>2</sup>, Dexter Canoy<sup>3</sup>, Mohammad Mamouei<sup>4</sup>, Abdelaali Hassaine<sup>5</sup>, Gholamreza Salimi-Khorshidi, Kazem Rahimi<sup>6</sup>

**BACKGROUND:** Whether the association between systolic blood pressure (SBP) and risk of cardiovascular disease is monotonic or whether there is a nadir of optimal blood pressure remains controversial. We investigated the association between SBP and cardiovascular events in patients with diabetes across the full spectrum of SBP.

**METHODS:** A cohort of 49 000 individuals with diabetes aged 50 to 90 years between 1990 and 2005 was identified from linked electronic health records in the United Kingdom. Associations between SBP and cardiovascular outcomes (ischemic heart disease, heart failure, stroke, and cardiovascular death) were analyzed using a deep learning approach.

**RESULTS:** Over a median follow-up of 7.3 years, 16 378 cardiovascular events were observed. The relationship between SBP and cardiovascular events followed a monotonic pattern, with the group with the lowest baseline SBP of <120 mm Hg exhibiting the lowest risk of cardiovascular events. In comparison to the reference group with the lowest SBP (<120 mm Hg), the adjusted risk ratio for cardiovascular disease was 1.03 (95% CI, 0.97–1.10) for SBP between 120 and 129 mm Hg, 1.05 (0.99–1.11) for SBP between 130 and 139 mm Hg, 1.08 (1.01–1.15) for SBP between 140 and 149 mm Hg, 1.12 (1.03–1.20) for SBP between 150 and 159 mm Hg, and 1.19 (1.09–1.28) for SBP ≥160 mm Hg.

**CONCLUSIONS:** Using deep learning modeling, we found a monotonic relationship between SBP and risk of cardiovascular outcomes in patients with diabetes, without evidence of a J-shaped relationship. (*Hypertension*. 2023;80:598–607. DOI: 10.1161/HYPERTENSIONAHA.122.20489.) • **Supplemental Material**

**Key Words:** blood pressure ■ causality ■ cardiovascular diseases ■ diabetes mellitus ■ machine learning

**B**lood pressure (BP) reduction is a well-known primary and secondary preventive strategy for cardiovascular events in addition to diabetes.<sup>1</sup> Observational studies conducted in low-risk groups have suggested that the association between elevated BP and risk of major cardiovascular disease (CVD) is log-linear.<sup>2,3</sup> However, the association in patients with pre-existing cardiometabolic disorders is less well understood.

In people with diabetes, reports have been inconsistent in their conclusion about the nature of the association between systolic blood pressure (SBP) and risk of CVD. Although in a cohort of patients with diabetes free of known CVD, the risk of CVD increased continuously

with higher SBP, and in the same study when diabetic patients with or without prior CVD were considered, the relationship was J-shaped with the lowest observed risk of cardiovascular events among those with SBP between 130 and 139 mm Hg.<sup>4</sup> Another study similarly found a J-shaped pattern in patients with diabetes, with the nadir of risk between 135 and 139 mm Hg.<sup>5</sup> This apparent discontinuous relationship has found some support from conventional meta-analyses of randomized controlled trials. For instance, one study concluded that BP lowering increased the risk of cardiovascular death in trials of patients with diabetes when average SBP was <140 mm Hg at baseline.<sup>6</sup> However, previous

Correspondence to: Kazem Rahimi, Nuffield Department of Women's and Reproductive Health, University of Oxford, Level 3, Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU, UK. Email [kazem.rahimi@wrh.ox.ac.uk](mailto:kazem.rahimi@wrh.ox.ac.uk)

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/HYPERTENSIONAHA.122.20489>.

For Sources of Funding and Disclosures, see page 606.

© 2023 The Authors. *Hypertension* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited.

*Hypertension* is available at [www.ahajournals.org/journal/hyp](http://www.ahajournals.org/journal/hyp)

NOVELTY AND RELEVANCE

What Is New?

This study investigated the shape of the association between blood pressure and future risk of cardiovascular outcomes in a cohort of 49 000 patients with diabetes using a validated deep learning causal model (Targeted Bidirectional Electronic Health Records Transformer [T-BEHRT]).

What Is Relevant?

In contrast to conventional statistical approaches, the T-BEHRT model showed no evidence of a J-shaped pattern.

Patients in the lowest systolic blood pressure category of <120 mm Hg had the lowest risk of future cardiovascular disease.

Clinical/Pathophysiological Implications?

This study extends the lower, the better paradigm of hypertension to patients with diabetes and provides further reassurance about the role of intensive blood pressure lowering in this growing patient population.

Nonstandard Abbreviations and Acronyms

<b>BMI</b>	body mass index
<b>BP</b>	blood pressure
<b>CPRD</b>	Clinical Practice Research Datalink
<b>CVD</b>	cardiovascular disease
<b>EHR</b>	electronic health record
<b>IHD</b>	ischemic heart disease
<b>LR</b>	logistic regression
<b>RR</b>	risk ratio
<b>SBP</b>	systolic blood pressure
<b>T-BEHRT</b>	Targeted Bidirectional Electronic Health Records Transformer

In this study, we leveraged such novel deep learning methods to re-evaluate the relationship between SBP and cardiovascular events in a sample of 49 000 UK patients with diabetes.

METHODS

Data Availability

Requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the Clinical Practice Research Datalink (CPRD) organization (<https://cprd.com/data-access>).

Study Setting and Participants

We used prospectively collected EHR data from CPRD, which has been validated for epidemiological research.<sup>8,11,12</sup> The CPRD database contains retrospective anonymized patient data covering ≈7% of the UK population and is considered generally representative in terms of sex, age, and ethnicity. The approval for this work was given by the CPRD Independent Scientific Advisory Committee of UK. We used EHR data from 3 resources in which we identified a cohort of 49 000 individuals with prevalent diabetes: primary care, secondary care (Hospital Episode Statistics), and the Office of National Statistics (cause-specific mortality). We included people between 50 and 90 years of age with at least 1 BP measurement taken between the years 1990 and 2005. Consistent with previous studies on BP, the study entry (ie, baseline) was defined as the date of the first BP measurement in this time period.<sup>13–15</sup> We identified individuals as having diabetes at study entry using validated phenotyping methods; full list of codes for identifying diabetes are available in Table S1.<sup>11,16</sup> Consistent with previous epidemiological studies, patients with heart failure before study entry were excluded from the study.<sup>1</sup>

The Transparency and Openness Promotion and the Reporting of Studies Conducted Using Observational Routinely-Collected Data reporting guidelines were followed for this cohort study. CPRD Independent Scientific Advisory Committee of UK (protocol: 16\_049R) gave the approval for this work.

observational studies, which implemented conventional statistical models are prone to issues of reverse causation and uncontrolled confounding,<sup>1,4</sup> and tabular meta-analyses of trials, which have provided the best level of evidence to date, are vulnerable to the ecological fallacy.<sup>7</sup>

As a consequence, the shape of the association between SBP and cardiovascular endpoints in patients with pre-existing diabetes remains contentious. Recently, better access to rich, longitudinal electronic health records (EHRs) in line with the development of deep learning models for causal inference have provided a new opportunity to overcome issues of uncontrolled confounding that low-dimensional expert-driven statistical approaches are prone to. Methods combining deep learning modeling for rich EHR data utilizing automatic feature extraction and confounding adjustment as well as semiparametric statistical modeling for mitigation of selection biases have shown promising results in several works utilizing semisynthetic and routine clinical EHR data.<sup>8–10</sup>

## Exposures

SBP was the exposure variable and was derived from CPRD measurement data. In CPRD, BP measurement is recorded by general practice staff during an in-person visit or consultation.<sup>12</sup> The European Society of Cardiology guidelines recommend 3 BP measurements measured 1 to 2 minutes apart with BP recording as the average of the last 2 BP readings.<sup>17</sup> In CPRD, the general practice staffs follow the same approach but a single BP measurement is recorded from each visit.<sup>12</sup> We excluded the SBP values <50 and >300 mm Hg as suggested by the previously published phenotyping methods to exclude outlier measurements.<sup>18</sup> All analyses were conducted with exposure status calculated as mean of SBP measurements in the first 12 months after study entry (ie, baseline exposure period). For example, for a hypothetical individual with 4 measurements in the first 12 months following study entry, the exposure would be the mean value of the 4 measurements. Patients who had a cardiovascular event or left the study during this baseline exposure period were removed from the analyses. Patients were categorized into 6 exposure categories of SBP: <120 mm Hg (reference), 120–129 mm Hg, 130–139 mm Hg, 140–149 mm Hg, 150–159 mm Hg, and ≥160 mm Hg.

## Outcomes

The primary outcome was fatal or nonfatal CVD, defined as a composite of ischemic heart disease (IHD), incident heart failure, stroke, and cardiovascular death. Secondary outcomes were components of the primary outcome: (1) IHD, (2) incident heart failure, and (3) stroke. All outcomes were identified by Read codes (primary care) and *International Classification of Diseases, Tenth Revision (ICD-10)* codes (secondary care and mortality data) as reported previously.<sup>11</sup> Follow-up period started 12 months after study entry; this feature of study design was implemented in order to avoid conducting inference within the exposure period (first 12 months after study entry). Thus, events that occurred between 12 and 120 months after study entry (eg, 108 months or 9 years of follow-up period) were captured for analysis (Figure S1).

## Statistical and Deep Learning Analyses

For the deep learning approach, we used Targeted Bidirectional Electronic Health Records Transformer (T-BEHRT) for risk ratio (RR) estimation of association between SBP and cardiovascular outcomes with SBP of <120 mm Hg considered as reference group.<sup>9</sup> For each of these comparisons, T-BEHRT was first trained to jointly predict exposure category (propensity score) and risk of outcome with 5-fold cross-validation implemented for training and testing.<sup>19</sup> The T-BEHRT model incorporated all recorded diagnoses and medications in medical history prior to study entry in addition to baseline smoking status (current, former, and never a smoker)—identified by last known status in the 12 months before baseline—and sex.

Initial estimates were computed on patients from test set of each of the 5 iterations. Second, equipped with these initial estimates, the T-BEHRT model updated risk estimates utilizing doubly-robust post hoc estimator, Cross-Validated Targeted Maximum Likelihood Estimation, to further mitigate biases.<sup>20</sup> RR and 95% CI were derived from this post hoc estimation procedure. More details and implementation of the T-BEHRT

approach can be found in the [Supplemental Methods](#) section (Figure S2; Table S2).

To compare against the T-BEHRT, logistic regression (LR) models were implemented to investigate the relationship between SBP and outcomes in individuals with diabetes. The exposure, SBP group was included as a categorical variable. The model was adjusted for age, sex, smoking status at baseline, body mass index (BMI), antihypertensive use at baseline, high-density lipoprotein, total cholesterol, triglycerides, atrial fibrillation, and chronic kidney disease. Baseline BMI, total cholesterol, high-density lipoprotein, and triglycerides were calculated as the average of measurements in the 12 months before baseline. Antihypertensives were identified by British National Formulary codes.<sup>21</sup> Smoking status (current, former, and never a smoker) was identified by last known status in the 12 months up to baseline. To comply with standard epidemiological research, we conducted imputation before inputting data to LR models. Multiple imputation using chained equations was used to impute missing variables BMI, high-density lipoprotein, total cholesterol, triglycerides, and smoking status; 25 imputations were conducted. For the LR, an estimate for the RR was obtained utilizing direct standardization.<sup>22</sup> We calculated RR as the average across the test sets of k-fold cross-validation (k=5) and calculated 95% CI over the 5 runs.<sup>23</sup> Additionally, the crude RR was also calculated as the average risk of the outcome in a particular exposure group divided by the average risk of the same outcome in the reference exposure group.

In addition, to check that findings of our binary regression models were not simply due to unaccounted informed censoring during follow-up, a Cox proportional hazards model was implemented with the same predictors as for the LR model. As is done traditionally for Cox proportional hazards modeling, censoring and time of cardiovascular event data were used for the modeling. The proportional hazards assumption was tested by plotting Schoenfeld residuals.

As a sensitivity analysis, to check that the impact of more complete adjustment of covariates, we extended our adjustment of the LR model with the following additional variables: insulin use at baseline, mean of measured values of hemoglobin A1c in the 12 months up to study entry, and duration of diabetes at baseline. As a measurement, hemoglobin A1c measurements were missing for 45.1% of patients in our cohort. Multiple imputation using chained equations was used to impute missing variables hemoglobin A1c, BMI, high-density lipoprotein, total cholesterol, triglycerides, and smoking status; 25 imputations were conducted.

Several additional sensitivity analyses were conducted using the T-BEHRT model. First, we conducted analyses stratified by sex (male and female) and age (≤75 and >75 years of age) analyses. Second, since antihypertensives can dilute associations, we repeated our main analysis with patients who had not taken antihypertensives after study entry.<sup>5</sup> Third, we conducted stratified analysis of the primary outcome additionally by baseline antihypertensive use. Fourth, to assess the possible impact of reverse causation, we excluded individuals who had cardiovascular events in the first 12 and 24 months of the follow-up period (24 and 36 months after study entry). Fifth, we repeated analysis without excluding patients who had an event in the first 12 months after study entry.

All data processing, data imputation, and analyses were undertaken in the programming language, Python. The

programming code for the T-BEHRT model is found as an open-source package on the Deep Medicine research group codebase (<https://github.com/deepmedicine>). Please see the Major Resources Table in the [Supplemental Material](#).

## RESULTS

### Population Statistics

A total of 49 000 patients with diabetes were included in this study ([Figure S3](#)). Descriptive statistics of the cohort according to the baseline SBP are shown in Table 1. Median age at baseline was 65 years of age. Forty-five percent of the individuals in the cohort were women and 39% current or former smokers. Patients in lower SBP exposure groups had higher prevalence of baseline IHD and lower proportion of antihypertensive use. BMI at baseline generally indicated an overweight cohort across all exposure groups. Prior to index date, patients had a median of 10 (interquartile range, 4–22) number of visits to general practice or hospital ([Table S3](#)). The median follow-up duration was 7.3 years. A total of 16 378 (33.4%) patients had cardiovascular events in the follow-up period, and 12 797 (26.1%) patients were censored (ie, patients who died due to causes other than CVD and who were otherwise lost to follow-up; Table 1;

[Figure S3](#)). Event rates are provided in the [Supplemental Material \(Table S4\)](#).

### Association of SBP on Cardiovascular Events

RR estimates from adjusted T-BEHRT model demonstrated a rise in the risk of cardiovascular events with a rise in SBP categories ([Figure 1](#)). Estimates investigating the association between SBP and secondary outcomes showed that the lowest risk of all secondary cardiovascular outcomes was observed at <120 mm Hg ([Figure 2](#)).

By contrast, the crude, LR, and the Cox proportional hazards models depicted a J-shaped pattern with all models capturing a nadir of risk at SBP between 130 and 139 mm Hg ([Figure 3](#)). These patterns were largely driven by the J-shaped associations of SBP with risk of incident heart failure and IHD ([Figure S4](#)). However, for stroke, the crude and adjusted LR models estimated that <120 mm Hg SBP demonstrated lowest risk similar to the T-BEHRT model. The adjusted Cox proportional hazards approach found that SBP between 130 and 139 mm Hg exhibited the lowest risk of all secondary outcome investigations. In sensitivity analysis of primary and secondary outcomes using an LR model with additional adjustment variables, the trends were almost identical to

**Table 1. Descriptive Statistics for Patients in Each Systolic Blood Exposure Group**

	Categories of baseline SBP					
	<120 mm Hg	120–129 mm Hg	130–139 mm Hg	140–149 mm Hg	150–159 mm Hg	≥160 mm Hg
No. (%)	2706 (5.5)	5881 (12.0)	10793 (22.0)	12203 (24.9)	8704 (17.8)	8713 (17.8)
Follow-up, y (IQR)	6.9 (2.5–9.0)	7.8 (3.1–9.0)	7.5 (3.2–9.0)	7.5 (3.2–9.0)	7.3 (3.0–9.0)	6.5 (2.8–9.0)
Censored (%)	821 (30.3)	1624 (27.6)	3007 (27.9)	3182 (26.1)	2132 (24.5)	2031 (23.3)
Age, y (IQR)	60.0 (52.0–70.0)	61.0 (53.0–70.0)	63.0 (55.0–72.0)	65.0 (56.0–73.0)	66.0 (58.0–74.0)	69.0 (61.0–76.0)
Women (%)	1045 (38.6)	2309 (39.3)	4466 (41.4)	5350 (43.8)	4048 (46.5)	4638 (53.2)
YOB (IQR)	1941 (1931–1948)	1940 (1931–1948)	1938 (1929–1947)	1936 (1928–1945)	1934 (1926–1942)	1931 (1924–1939)
BMI, kg/m <sup>2</sup> (IQR)*	29.3 (26.8–31.1)	29.6 (27.3–31.4)	29.5 (27.4–31.5)	29.5 (27.5–31.5)	29.4 (27.5–31.3)	29.1 (27.4–31.0)
HDL, mmol/L (IQR)*	1.3 (1.2–1.4)	1.2 (1.2–1.4)	1.3 (1.2–1.4)	1.3 (1.2–1.4)	1.3 (1.2–1.4)	1.3 (1.2–1.4)
TG, mmol/L (IQR)*	2.2 (1.6–2.6)	2.2 (1.7–2.6)	2.2 (1.7–2.6)	2.2 (1.8–2.6)	2.2 (1.8–2.6)	2.1 (1.8–2.5)
TC, mmol/L (IQR)*	5.1 (4.8–5.5)	5.1 (4.8–5.5)	5.1 (4.8–5.5)	5.2 (4.9–5.5)	5.2 (4.9–5.5)	5.3 (5.0–5.5)
Smoking status*:						
Current/former smoker (%)	1234 (45)	2604 (44)	4479 (41)	5005 (41)	3172 (36)	2928 (33)
Never smoker (%)	1472 (54)	3277 (55)	6314 (58)	7198 (58)	5532 (63)	5785 (66)
Disease at baseline:						
IHD (%)	409 (15.1)	763 (13.0)	1235 (11.4)	1219 (10.0)	870 (10.0)	823 (9.4)
Chronic kidney disease (%)	30 (1.1)	76 (1.3)	125 (1.2)	134 (1.1)	103 (1.2)	82 (0.9)
Stage 1 and 2 kidney disease (%)	37 (1.4)	67 (1.1)	123 (1.1)	143 (1.2)	79 (0.9)	103 (1.2)
Atrial fibrillation (%)	122 (4.5)	194 (3.3)	320 (3.0)	315 (2.6)	232 (2.7)	207 (2.4)
Medications prescribed at baseline:						
Antihypertensive use (%)	1036 (38.3)	2556 (43.5)	5459 (50.6)	6771 (55.5)	5045 (58.0)	5333 (61.2)

Values presented are median with IQR or %. BMI indicates body mass index; HDL, high-density lipoprotein; IHD, ischemic heart disease; IQR, interquartile range; SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; and YOB, year of birth.

\*The percentage of missing variables: BMI (55.1%), smoking status (29.2%), HDL (75.3%), TC (49.9%), and TG (68.7%).



those captured by the main adjusted LR approach (Table S5).

Shown in Figure 4, the sensitivity analyses investigating the association between SBP and the primary composite outcome using the T-BEHRT model preserved the trend found in the main analysis. In both sex and age-stratified analyses, the log-linear trend across the spectrum of SBP was generally preserved. In analysis of patients who had not taken antihypertensives during the exposure and follow-up periods, the RR estimates and corresponding 95% CI for each exposure group was slightly higher than their counterparts in the main analysis but overall mirrored the trend in the main analyses. Additionally, stratifying by antihypertensive usage at baseline, albeit the slight presence of a local minimum, the trend showed little material difference from the main result. Furthermore, excluding patients who had events in the first 12 and 24 months of follow-up also captured a similar trend as that of the main analysis. Lastly, incorporating those who had dropped out during the first 12 months following baseline (ie, the exposure period), the trend presented was similar to that of the main analysis.

DISCUSSION

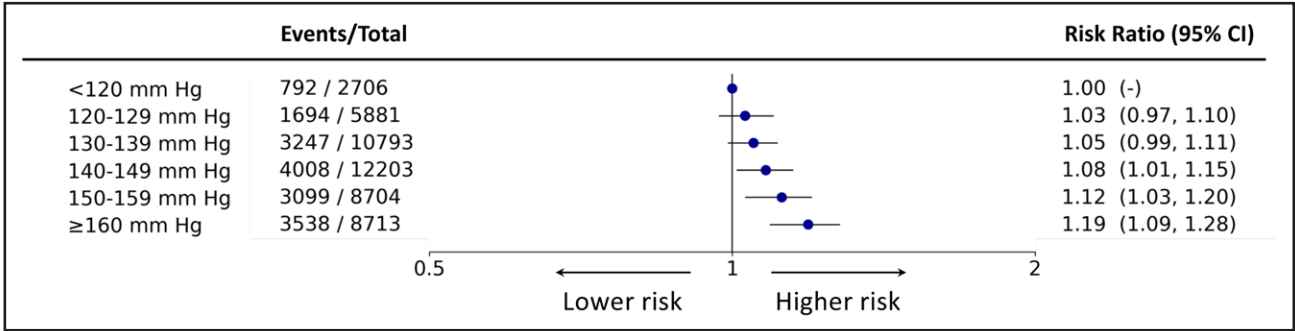
In this study, using deep learning modeling on a comprehensive dataset of UK EHR in a cohort of 49 000 individuals with diabetes, we found SBP to be monotonically associated to cardiovascular risk. Patients with SBP <120 mm Hg exhibited the lowest risk of cardiovascular outcomes. These results were compared with conventional approaches in epidemiology and were robust to several sensitivity analyses.

Findings from previous observational studies investigating the association of SBP and cardiovascular events in patients with diabetes utilizing conventional statistical modeling have shed light on a larger phenomenon: the shape of the association as a function of the comprehensiveness of confounding adjustment. Solely

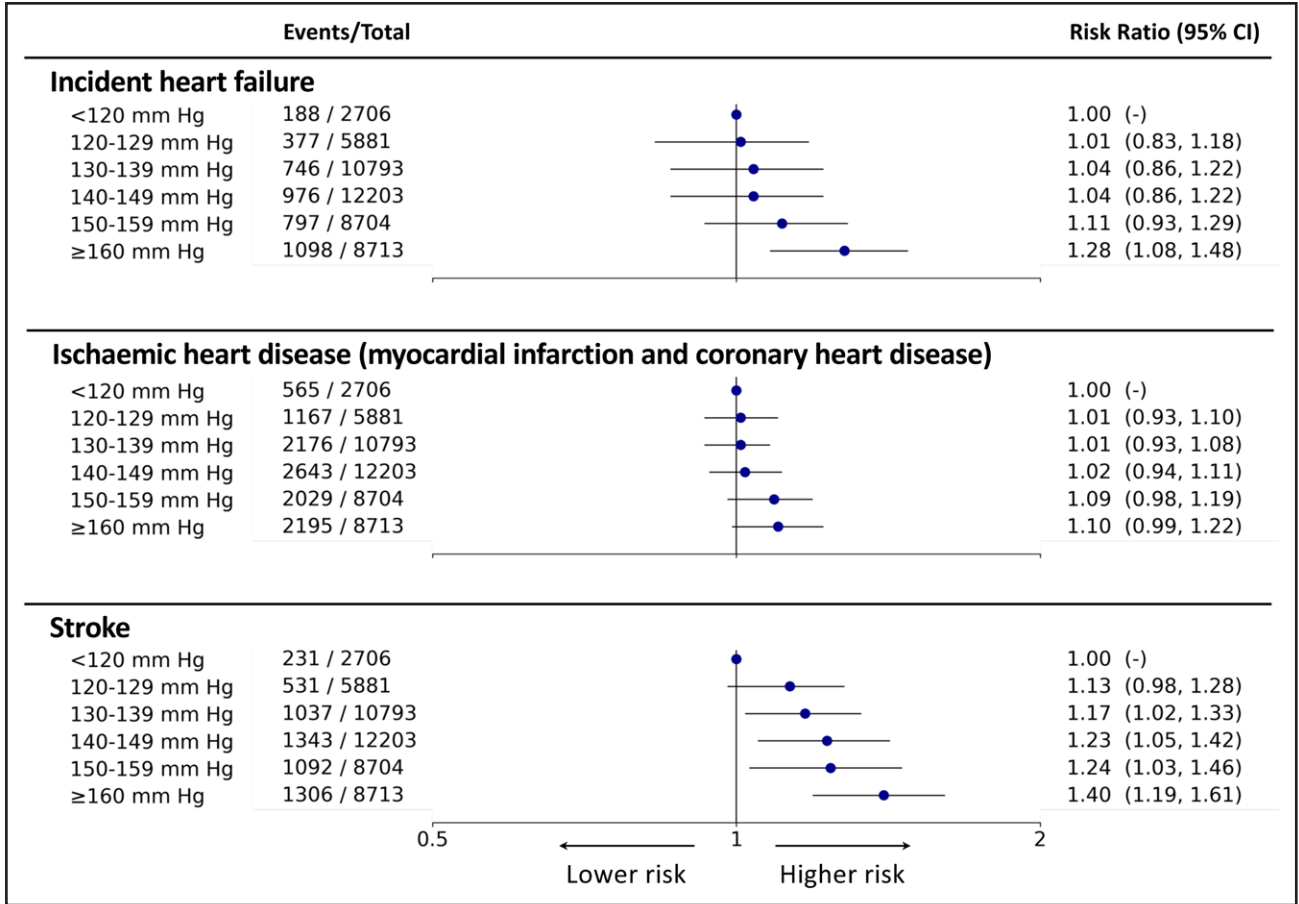
adjusting for confounders such as baseline age, sex, and demographic variables fell short of eliminating the J-shaped association in cohorts with diabetes free of CVDs as shown in past studies such as the Swedish National Diabetes Registry study with 54 000 patients and the ROSE study (Retrospective Epidemiological Study to Investigate Outcome and Mortality with Glucose Lowering Drug Treatment in Primary Care) with 34 000 patients.<sup>24,25</sup> However, extending the predictor set to include key cardiovascular risk factors more comprehensively modeled the association as compared to the former 2 studies and thus rejected the J-shaped association in a cohort of 187 000 patients.<sup>4</sup> However, in the same study, including patients with pre-existing cardiovascular conditions, the conventional modeling consistently captured the J-shaped relationship for all outcomes, with the exception of stroke.<sup>4</sup> Similarly, our analysis using LR modeling with adjustment for several commonly used confounders exhibited evidence of the J-shaped relationship between SBP and risk of composite CVD, incident heart failure, and IHD. More extensive adjustment or use of a survival model did not change the shape of the association.

Conventional statistical models are usually implemented in observational studies with curated, low-risk cohorts relatively free of multimorbidity at baseline. By ensuring the cohort is healthy, statistical models may only require a handful of established confounding variables to sufficiently provide adjusted and trustworthy estimates. However, in cohorts like ours with high baseline BMI and a host of underlying conditions—including some cardiovascular in nature indicating prevalent multimorbidity—conventional approaches with expert selected confounders and regression modeling might be insufficient for capturing complex interactions and confounding.

Our work implementing deep causal modeling addresses these issues and fills some of the gaps in this research of SBP and cardiovascular outcomes. With T-BEHRT modeling, our research found no evidence of the J-shaped association between SBP and primary and



**Figure 1. Forest plot of risk ratio estimates of the Targeted Bidirectional Electronic Health Records Transformer (T-BEHRT) model with 95% CI for association of systolic blood pressure (SBP) and the primary composite outcome.** From the left, the 6 exposure groups are shown in first column. Number of events and total number of patients in each exposure group is shown in second column. The forest plot and corresponding risk ratio estimates are shown in the right-most column relative to reference class, <120 mm Hg. The forest plot is plotted in logarithmic scale. For all estimates for reference class, there is no CI.



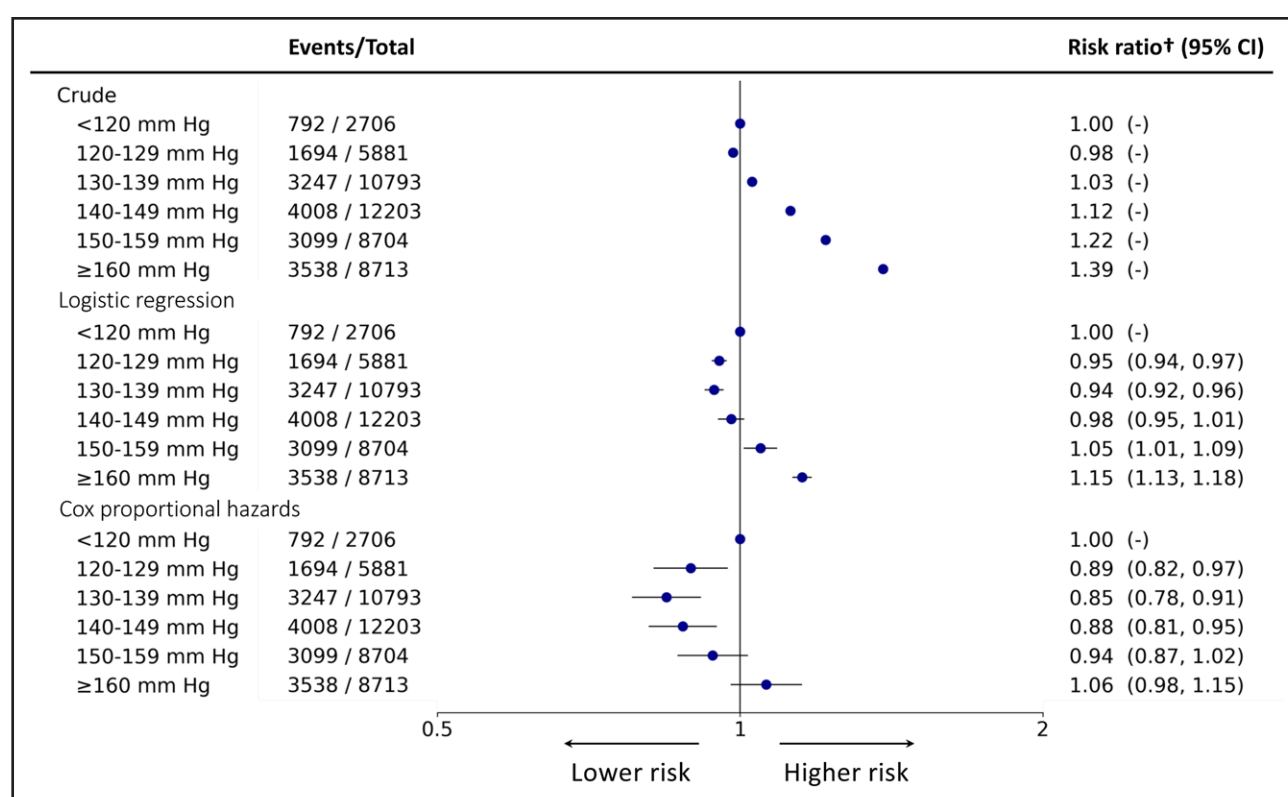
**Figure 2. Forest plot of risk ratio estimates of the Targeted Bidirectional Electronic Health Records Transformer (T-BEHRT) model with 95% CI for association of systolic blood pressure (SBP) and secondary outcomes.** From the left, the 6 exposure groups are shown in first column. Number of events and total number of patients in each exposure group is shown in second column. The forest plot and corresponding risk ratio estimates are shown in the right-most column relative to reference class, <120 mm Hg. The forest plot is plotted in logarithmic scale. For all estimates for reference class, there is no CI.

secondary outcomes. While a recent individual-participant data meta-analysis of randomized evidence has indirectly dismissed the existence of a J-shaped in patients with or without prior CVD,<sup>1</sup> evidence of heterogeneity of treatment effects in patients with diabetes has been controversial. In particular, a tabular meta-analysis of randomized trials in patients with diabetes has suggested that BP lowering might increase the risk of cardiovascular death when SBP is <140 mm Hg. While not directly analyzing BP lowering, our observational study, on the contrary, provides evidence that SBP maintains a log-linear relationship with cardiovascular outcomes across a wide range of baseline SBP categories in patients with diabetes.

We note in our cohort of high-risk patients, relatively modest associations were observed using the T-BEHRT approach. This, however, was also the case in our conventional modeling and comparable with previous research.<sup>4</sup> The weak associations might be due to 2 main reasons. First, as recently shown by trial evidence, the relative effect of BP lowering on cardiovascular outcomes was half as strong in people with diabetes as those without diabetes.<sup>26</sup> This might be due to the fact that part of the effect of BP

lowering on CVD outcomes is mediated through prevention of diabetes; a pathway that might not be relevant to people with pre-existing diabetes.<sup>27</sup> Second, our cohort included patients with several comorbidities and use of several medications. While all these attributes are typical in diabetes patients (and a suitable case for complex models such as T-BEHRT), they could lead to attenuation of true associations with CVD as compared to associations in lower risk cohorts.<sup>4,28</sup> This is supported by our sensitivity analyses, where exclusion of patients with use of antihypertensives at baseline or during follow-up led to stronger relationship between SBP and risk of CVD.

These results have important implications for cardiovascular research. Our analysis provides some clarity concerning the relationship between SBP and cardiovascular events in patients with diabetes. Hence, while our investigation on its own is insufficient for recommending revisions of hypertension guidelines, our work rather functions as independent analyses complementing the findings of the individualized patient data meta-analysis of randomized evidences.<sup>26</sup> Together, they support the lower the better paradigm of SBP in patients with diabetes.



**Figure 3. Forest plot of relative risk estimates of various conventional statistical models with 95% CI for association of systolic blood pressure (SBP) and the primary outcome.**

From the left, the 6 exposure groups are shown in first column. Number of events and total number of patients in each exposure group is shown in second column. The forest plot and corresponding risk ratio (hazard ratio for Cox proportional hazards model) estimates are shown in the right-most column relative to reference class, <120 mm Hg. The forest plot is plotted in logarithmic scale. For all crude estimates and estimates for reference class, there is no CI. The † indicates hazard ratio for the Cox proportional hazards model.

## Strengths and Limitations

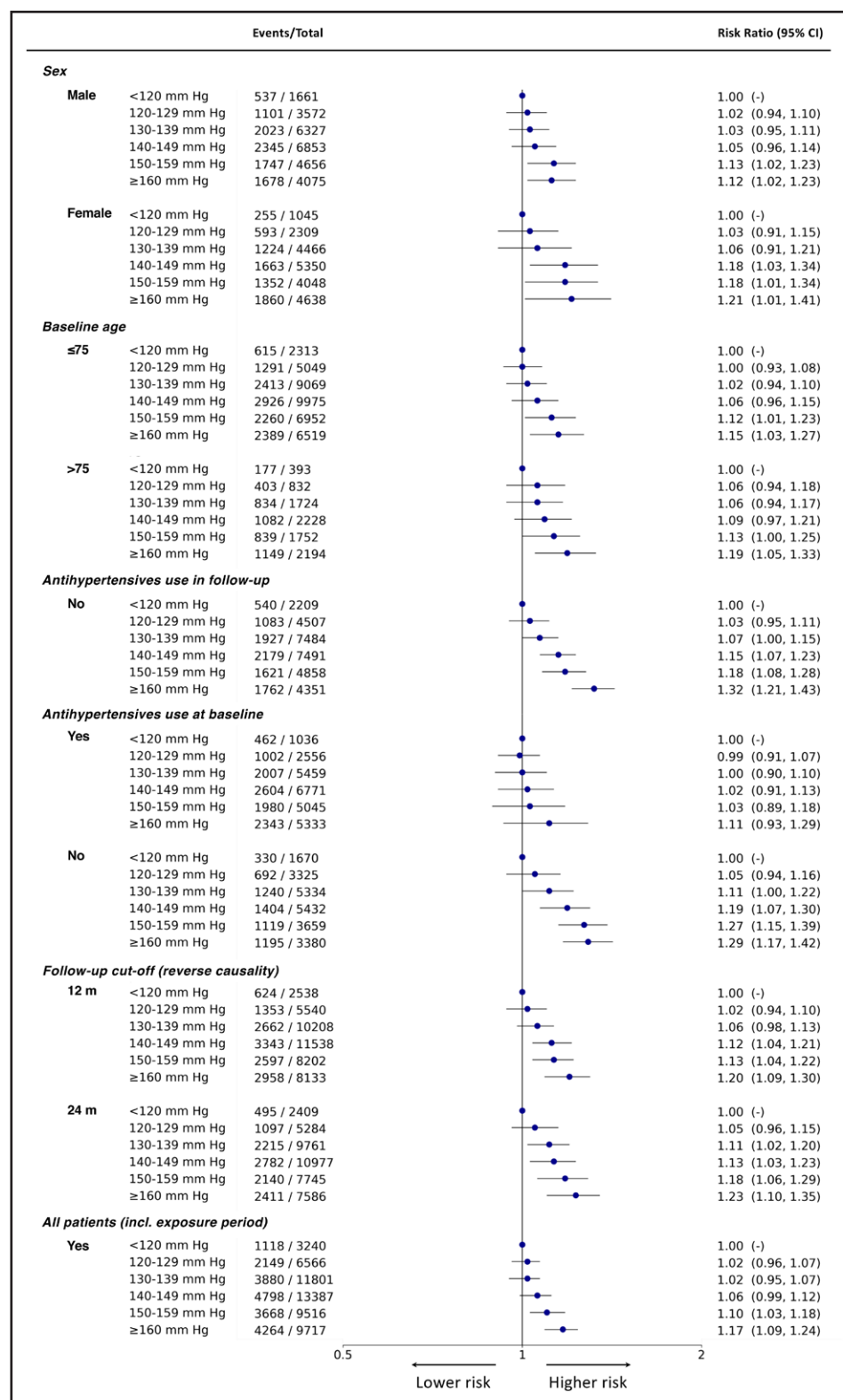
Our first strength is CPRD allows us access to a host of rich diagnosis, medication, and measurements variables. Also, we leveraged phenotyping methods validated for CPRD. Unlike previous studies of SBP and cardiovascular events that have excluded older individuals, patients between 50 and 90 years of age were included in this work. Exclusion based on baseline attributes was limited; only those with heart failure were excluded at baseline.

Second, deep learning feature extraction was used to automatically adjust for confounding variables and latent interactions in our input data. We also implemented conventional statistical approaches enabling direct comparison of T-BEHRT to established methods in the study of SBP and cardiovascular outcomes. We showed that by using superior adjustment methods, we can more effectively model the association in observational data, thereby rejecting the J-shaped argument.

This study also has several limitations. EHR data might have some level of measurement error or misclassification. Despite endorsement of the validity and reliability of physician diagnoses in the CPRD dataset, especially concerning studies of cardiovascular disorders,<sup>29</sup> case ascertainment of diabetes might have misclassifications

(eg, metabolic syndromes and related disorders as opposed to diabetes). Measurement errors are a natural issue with EHR data, especially that of BP, but we have attempted to mitigate these issues in the case of SBP data by taking an average of multiple measurements over the course of 12 months following baseline. However, we are unable to make any direct inference about the relationship in particular patient groups such as those with white coat hypertension.

T-BEHRT framework typically requires high-dimensional longitudinal data. The T-BEHRT estimator can function optimally with access to multiple EHR modalities (diagnoses, medications, etc) and associated temporal annotation (age and calendar year). Furthermore, as previously reported, when such rich data are provided, the unsupervised learning component (masked EHR modeling) of the T-BEHRT model works well in reducing bias in estimation.<sup>9</sup> When data are sparse or limited (finite sample), the doubly-robust estimation, with known benefits for finite-sample estimation, more accurately estimates RR than the variants without utilization of doubly-robust estimation.<sup>20</sup> Furthermore, when positivity (overlap) between exposure groups is limited, the T-BEHRT model fares better than other benchmark models.<sup>22</sup> Specifically, in our study of patients with



**Figure 4. Forest plot of risk ratio estimates of the Targeted Bidirectional Electronic Health Records Transformer (T-BEHRT) model with 95% CI for association of systolic blood pressure (SBP) and the primary outcome in sensitivity analyses.**

The particular sensitivity analysis is italicized on the left with strata indented. From the left, the 6 exposure groups are shown in the first column. Number of events and total number of patients in each exposure group is shown in the second column. The forest plot and corresponding risk ratio estimates are shown in the right-most column relative to reference class, <120 mm Hg. The forest plot is plotted in logarithmic scale. For all estimates for reference class, there is no CI.



cardiometabolic multimorbidity, where the traditional LR model insufficiently captures confounder relationships, deep learning modeling in tandem with semiparametric methods can be appropriately implemented for robust RR estimation. Further testing to assess model estimation accuracy is needed for settings in which the exposure is poorly defined (eg, generic painkiller) or when the outcome is categorical or continuous. Future confirmatory investigations on additional data sources would be invaluable for studying the association from different perspectives.

Our deep learning approach functions in the binary outcome framework as opposed to the time-to-event framework, which would be a more optimal framework for analyses where time-dependent bias is likely. Specifically, failure to carefully consider censoring in cohort can bias downstream estimation of the association. However, even though deep learning for survival analyses has been explored in the risk prediction setting, methodological issues in estimation of hazard ratios using nonlinear, deep learning models have limited pursuits of survival modeling for population-level association estimation. Hence, methodological studies would be necessary to develop and demonstrate the utility of deep learning-driven survival models for EHR, which could be applied to investigate the studied association and other challenging questions. Lastly, as is the case for all observational studies, our model is unable to overcome the challenge of fully taking account of unmeasured confounding and when this is of concern, alternative study designs such as randomized controlled trial might be required.

## Perspectives

Using large-scale, linked EHR, our investigation utilizing a deep learning model concluded that patients with diabetes in the lowest category of SBP of <120 mm Hg exhibited the lowest cardiovascular risk. This was in contrast to expert-dependent conventional approaches that have largely defended the paradoxical nonlinear J-shaped association between SBP and cardiovascular end points in those with diabetes. This difference is likely explained by the deep learning model's ability to more comprehensively capture known and latent confounders in routine EHR, thus reducing the potential of important confounders being missed out. In light of a recent meta-analysis of trials demonstrating that BP lowering medications reduces the risk of cardiovascular events in patients with diabetes regardless of baseline BP, our investigation provides complementary evidence against a causal link between low BP and risk of CVD in this population. Together these lines of evidence support the lower, the better paradigm of BP management. In terms of methodological insights, our work demonstrates that where complex confounding might be at play, conventional modeling is more likely to fall short of

delivering well-adjusted estimates of risk. In such scenarios, models such as T-BEHRT, which are capable of automatically capturing known and latent confounding in minimally processed, rich EHR provide a worthwhile alternative for minimizing biases. Future research applying and testing such data-driven deep learning methods in high-risk cohorts would be beneficial for disentangling the nuances of risk and protection in patients with pre-existing conditions.

## ARTICLE INFORMATION

Received October 18, 2022; accepted December 8, 2022.

### Affiliations

Deep Medicine, Oxford Martin School, University of Oxford, Oxford, United Kingdom (S.R., Y.L., M.N., M.M., G.S.-K., K.R.). Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, United Kingdom (S.R., Y.L., M.N., M.M., G.S.-K., K.R.). NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom (K.R.). Population Health Sciences Institute, University of Newcastle, Newcastle, United Kingdom (D.C.). Division of Informatics, Imaging and Data Sciences, University of Manchester, Manchester, United Kingdom (A.H.).

### Author Contributions

S. Rao and K. Rahimi conceived the study concept and designed the study. K. Rahimi and S. Rao acquired and curated the data. S. Rao conducted the data analysis and analysis visualizations. M. Nazarzadeh contributed to data analysis. S. Rao wrote the initial draft of the article. K. Rahimi is the guarantor for the study. All of the authors contributed to the revisions of the article and had full access to the data.

### Sources of Funding

This work was supported by grants from the British Heart Foundation ([BHF]; grant number, FS/PhD/21/29110 to Y. Li and K. Rahimi), (grant number, FS/19/36/34346 to M. Nazarzadeh and K. Rahimi), and (grant number, PG/18/65/33872 to K. Rahimi and D. Canoy); United Kingdom Research and Innovation (UKRI) Global Challenges Research Fund (GCRF; grant number, ES/P0110551/1 to K. Rahimi); Oxford National Institute of Health Research (NIHR) Biomedical Research Center (to K. Rahimi), Novo Nordisk grant (to M. Mamouei), and the Oxford Martin School (OMS), University of Oxford (to K. Rahimi). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. The views expressed are those of the authors and not necessarily those of the OMS, the BHF, the GCRF, the NIHR, or the Department of Health and Social Care.

### Disclosures

None.

### Supplemental Material

Supplemental Methods  
Tables S1–S5  
Figures S1–S4  
Major Resources Table

## REFERENCES

1. The Blood Pressure Lowering Treatment Trialists' Collaboration. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. *Lancet*. 2021;397:1625–1636. doi: 10.1016/S0140-6736(21)00590-0
2. Collaboration PS. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903–1913. doi: 10.1016/S0140-6736(02)11911-8
3. Lacey B, Lewington S, Clarke R, Kong XL, Chen Y, Guo Y, Yang L, Bennett D, Bragg F, Bian Z, et al. Age-specific association between blood pressure and vascular and non-vascular chronic diseases in 0.5 million adults in China: a prospective cohort study. *Lancet Glob Health*. 2018;6:e641–e649.

- Available at: [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(18\)30217-1/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30217-1/fulltext). Accessed December 15, 2021.
4. Adamsson Eryd S, Gudbjörnsdóttir S, Manhem K, Rosengren A, Svensson AM, Miftaraj M, Franzén S, Björck S. Blood pressure and complications in individuals with type 2 diabetes and no previous cardiovascular disease: national population based cohort study. *BMJ*. 2016;354:4070. doi: 10.1136/bmj.4070.
  5. Sundström J, Arima H, Jackson R, Turnbull F, Rahimi K, Chalmers J, Woodward M, Neal B; Blood Pressure Lowering Treatment Trialists' Collaboration. Effects of blood pressure reduction in mild hypertension: a systematic review and meta-analysis. *Ann Intern Med*. 2015;162:184–191. doi: 10.7326/M14-0773
  6. Brunström M, Carlberg B. Effect of antihypertensive treatment at different blood pressure levels in patients with diabetes mellitus: systematic review and meta-analyses. *BMJ*. 2016;352:i717. doi: 10.1136/bmj.i717
  7. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res Synth Methods*. 2010;1:2–19. doi: 10.1002/jrsm.4
  8. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, Zhu Y, Rahimi K, Salimi-Khorshidi G. BEHRT: transformer for electronic health records. *Sci Rep*. 2020;10:7155. doi: 10.1038/s41598-020-62922-y
  9. Rao S, Mamouei M, Salimi-Khorshidi G, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Rahimi K. Targeted-BEHRT: deep learning for observational causal inference on longitudinal electronic health records [published online June 23, 2022]. *IEEE Trans Neural Netw Learn Syst*. doi: 10.1109/TNNLS.2022.3183864. <https://ieeexplore.ieee.org/document/9804397>
  10. Rao S, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Cleland JG, Lukasiewicz T, Salimi-Khorshidi G, Rahimi K. An explainable Transformer-based deep learning model for the prediction of incident heart failure. *IEEE J Biomed Health Inform*. 2022;26:3362–3372. doi: 10.1109/JBHI.2022.3148820
  11. Tran J, Norton R, Conrad N, Rahimi F, Canoy D, Nazarzadeh M, Rahimi K. Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: a population-based cohort study. *PLoS Med*. 2018;15:e1002513. doi: 10.1371/journal.pmed.1002513
  12. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, Smeeth L. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44:827–836. doi: 10.1093/ije/dyv098
  13. Emdin CA, Anderson SG, Woodward M, Rahimi K. Usual blood pressure and risk of new-onset diabetes evidence from 4.1 million adults and a meta-analysis of prospective studies. *J Am Coll Cardiol*. 2015;66:1552–1562. doi: 10.1016/j.jacc.2015.07.059
  14. Emdin CA, Anderson SG, Salimi-Khorshidi G, Woodward M, MacMahon S, Dwyer T, Rahimi K. Usual blood pressure, atrial fibrillation and vascular risk: evidence from 4.3 million adults. *Int J Epidemiol*. 2017;46:162–172. doi: 10.1093/ije/dyw053
  15. Emdin CA, Rothwell PM, Salimi-Khorshidi G, Kiran A, Conrad N, Callender T, Mehta Z, Pendlebury ST, Anderson SG, Mohseni H, et al. Blood pressure and risk of vascular dementia: evidence from a primary care registry and a cohort study of transient ischemic attack and stroke. *Stroke*. 2016;47:1429–1435. doi: 10.1161/strokeaha.116.012658
  16. Trowell WJ. "British national formulary". *Br Med J (Clin Res Ed)*. 1981;282:1078. doi: 10.1136/bmj.282.6269.1078
  17. Williams B, Mancia G, Spiering W, Rosei EA, Azizi M, Burnier M, Clement DL, Coca A, de Simone G, Dominiczak A, et al; ESC Scientific Document Group. 2018 ESC/ESH guidelines for the management of arterial hypertension. *Eur Heart J*. 2018;39:3021–3104. doi: 10.1093/eurheartj/ehy339
  18. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick N, Banerjee A, Dobson R, Fatemifar G, Kuan V, Lumbers T, Pasea L, et al. UK phenomics platform for developing and validating EHR phenotypes: CALIBER. *J Am Med Inform Assoc*. 2019;26:1545–1559. doi: 10.1093/jamia/ocz105
  19. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55. doi: 10.1093/biomet/70.1.41
  20. Levy J. An easy implementation of CV-TMLE. *arXiv*. Posted online November 13, 2022. <https://doi.org/10.48550/arXiv.1811.04573>
  21. Payne R, Denholm R. CPRD product code lists used to define long-term preventative, high-risk, and palliative medication. *University of Bristol*. Published online February 5, 2018. doi: 10.5523/bris.k38ghxkcub622603i5wq6bwag
  22. Hernán MA, Robins JM. (2020) *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. 2020.
  23. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. *Econometrics J*. 2018;21:C1–C68. doi: 10.1111/ectj.12097
  24. Cederholm J, Gudbjörnsdóttir S, Eliasson B, Zethelius B, Eeg-Olofsson K, Nilsson PM; NDR. Blood pressure and risk of cardiovascular diseases in type 2 diabetes: further findings from the Swedish National Diabetes Register (NDR-BP II). *J Hypertens*. 2012;30:2020–2030. doi: 10.1097/HJH.0b013e3283577bdf
  25. Sundström J, Sheikhi R, Östgren CJ, Svennblad B, Bodegard J, Nilsson PM, Johansson G. Blood pressure levels and risk of cardiovascular events and mortality in type-2 diabetes: cohort study of 34009 primary care patients. *J Hypertens*. 2013;31:1603–1610. doi: 10.1097/hjh.0b013e32836123aa
  26. Nazarzadeh M, Bidel Z, Canoy D, Copland E, Bennett DA, Dehghan A, Davey Smith G, Holman RR, Woodward M, Gupta A, et al; Blood Pressure Lowering Treatment Trialists' Collaboration. Blood pressure-lowering treatment for prevention of major cardiovascular diseases in people with and without type 2 diabetes: an individual participant-level data meta-analysis. *Lancet Diabetes Endocrinol*. 2022;10:645–654. doi: 10.1016/S2213-8587(22)00172-3
  27. Nazarzadeh M, Bidel Z, Canoy D, Copland E, Wamil M, Majert J, Smith Byrne K, Sundström J, Teo K, Davis BR, et al. Blood pressure lowering and risk of new-onset type 2 diabetes: an individual participant data meta-analysis. *Lancet*. 2021;398:1803–1810. doi: 10.1016/s0140-6736(21)01920-6
  28. Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D, Wamil M, Majert J, McManus R, Adler A, Agodoa L, et al. Age-stratified and blood-pressure-stratified effects of blood-pressure-lowering pharmacotherapy for the prevention of cardiovascular disease and death: an individual participant-level data meta-analysis. *Lancet*. 2021;398:1053–1064. doi: 10.1016/S0140-6736(21)01921-8
  29. Arana A, Margulis A, Varas-Lorenzo C, Bui CL, Gilman A, McQuay LJ, Reynolds M, Rebordosa C, Franks B, de Vogel S, et al. Validation of cardiovascular outcomes and risk factors in the Clinical Practice Research Datalink in the United Kingdom. *Pharmacoepidemiol Drug Saf*. 2021;30:237–247. doi: 10.1002/pds.5150