The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Main data type is described in title, more details in Abstract. Location of study population and timeframe for data collection is in Abstract.
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Info is dispersed across all 5 paragraphs of Introduction.
Objectives	3	State specific objectives, including any prespecified hypotheses			2nd + 3rd paragraph of Introduction
Methods					
Study Design	4	Present key elements of study design early in the paper			First paragraph of Methods section describes derivation and and validation of Cox proportional-hazards models

Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Subsection "Patient selection and data retrieval"
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Subsection "Patient selection and data retrieval" We wrote custom code to preprocess data. The code was not particularly noteworthy. The code was tested. We include a flow diagram in Supplementary Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Subsection " Artificial intelligence-enabled adjudication of cardiovascular outcomes", Table 1, Supplemental Table 2, 3
Data sources/ measurement	8	For each variable of interest, give sources of data and details			Subsection " Patient selection and data retrieval

		of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			" and " Artificial intelligence-enabled adjudication of cardiovascular outcomes"
Bias	9	Describe any efforts to address potential sources of bias			Models are trained on dataset from one hospital, then validated on dataset from a different hospital
Study size	10	Explain how the study size was arrived at			Collection of hematology data started in 2017, in connection with other projects. This start date is the most important factor limiting the size of our study.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Subsections "Hematologic Predictors" and "Non-Hematologic Predictors". And Supplemental Table 2, 3
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical			a) Info is dispersed throughout this section. See, in particular, subsections "Derivation and validation of the models", 'Evaluation of model performance", "Other atatistical analyses" b) No subgroup analysis, but we did separate survival models for men and for women. c) Missing predictors were rare. See flow diagram in Supp. Fig. 1. Supp. Table 1B shows performance of classifiers that adjudicate

		methods taking account of sampling strategy (e) Describe any sensitivity analyses			CVD outcomes, though imperfect sensitivity is not handled as a missing-data problem.
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Authors had access to the primary Enterprise Data Warehouse. An accounting of data-cleaning methods is contained in the flow diagram s in Supplemental Figure 1
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	See the flow diagrams in Supp. Fig. 1, and Methods subsection "Patient selection and data retrieval"
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders			Subsection "Baseline characteristics" Table 2, and Supp. Fig. 2

		(b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			Subsection "Rates of cardiovascular disease and death". Table 3.
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			a) Subsections "Discrimination and calibration" "Calibration curves", "Parameter estimates'; Fig. 1 and 2; Table 4, 5, 6; Supp. Table 4. b) No such categorical variables. c) Not relevant
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Subsection "Related Analyses" Supp. Table 8
Discussion					
Key results	18	Summarise key results with reference to study objectives			First two paragraphs

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	6th paragraph
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			7th paragraph
Generalisability	21	Discuss the generalisability (external validity) of the study results			6th paragraph
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Funding was provided by One Brave Idea, a research grant awarded to Brigham and Women's Hospital.
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Code used for data processing and analysis will be available upon request.

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.