Deep learning facilitates rapid cohort identification using human and veterinary clinical narratives

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Author's assessment
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	See Methods in the Abstract.
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	See Methods and Findings in the Abstract.
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	See Introduction, paragraph 1
Objectives	3	State specific objectives, including any prespecified hypotheses	See "Learning on human and veterinary medical records" in the Introduction section
Methods			
Study design	4	Present key elements of study design early in the paper	See Study Design in the Methods section.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	See Clinical Setting in the Methods section
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	See Patients in the Methods section paragraph 1.
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	Not applicable.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	See categories in Table 2 of the Methods Section; and Table 1 of Supplementary Material 2.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	See Clinical Setting and Patients in the Methods section.
Bias	9	Describe any efforts to address potential sources of bias	See Figure 1 in Study Design in the Methods section.
Study size	10	Explain how the study size was arrived at	See Patients in the Methods section.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If	See Deep learning models and Natural Language Processing in the Methods section.

		applicable, describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	See Study Design and Statistical analysis in the Methods section.
		(*b*) Describe any methods used to examine subgroups and interactions	Not applicable.
		(*c*) Explain how missing data were addressed	See Patients in the Methods section.
		(*d*) If applicable, explain how loss to follow-up was addressed	Not applicable since it is a retrospective study.
		(*e*) Describe any sensitivity analyses	Not applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	See Table 2 in the Methods section.
		(b) Give reasons for non-participation at each stage	Not applicable since this is a chart review retrospective study, from de-identified data previously collected as part of care delivery.
		(c) Consider use of a flow diagram	Table 2 shows sufficient information.
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	See Table 2.
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable.
		(c) Summarise follow-up time (eg, average and total amount)	Not applicable.
Outcome data	15*	Report numbers of outcome events or summary measures over time	Not applicable.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	See Table 3 in the Results section.
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable.
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	Not applicable.

Discussion			
Key results	18	Summarise key results with reference to study objectives	See Paragraph 1 in the Discussion section.
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	See Paragraph 4 in the Discussion section.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	See Paragraphs 2-3 in the Discussion section.
Generalisability	21	Discuss the generalisability (external validity) of the study results	See Public Health Implications in the Discussion section
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	See Funding in the Declarations section

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.