

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

	Item No	Recommendation	Page No
<b>Title and abstract</b>	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	1-2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3-5
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed	3-4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
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	5
Bias	9	Describe any efforts to address potential sources of bias	4, 10
Study size	10	Explain how the study size was arrived at	N/A
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5
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) If applicable, explain how loss to follow-up was addressed (e) Describe any sensitivity analyses	5, fig 1
<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg 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	5, 6; fig 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	Table 1
Outcome data	15*	Report numbers of outcome events or summary measures over time	Appendix Table 1

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 (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	Table 2 and 3, pg 6-7
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	8
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	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10, 11
Generalisability	21	Discuss the generalisability (external validity) of the study results	Not explicitly discussed due to word limit
<b>Other information</b>			
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	Title page

\*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>.

The RECORD statement

	<b>Item No.</b>	<b>RECORD items</b>	<b>Location in manuscript where items are reported</b>
<b>Title and abstract</b>			
	1	<p>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.</p> <p>RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.</p> <p>RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.</p>	<p>1.1 Abstract (page 3). Since the number of databases is high, we did not have space to include them in the abstract</p> <p>1.2 Page 3, lines 17-19</p> <p>1.3 Page 3, line 17-19</p>
<b>Methods</b>			
Participants	6	<p>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.</p> <p>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.</p> <p>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</p>	<p>6.1 Appendix Tables 1-2</p> <p>6.2 Appendix Table 1-2, references 17-23</p>

		individuals with linked data at each stage	6.3 Figure 1. Because of the number of databases used, the complexity, and that data are linked at MCHP, we have not included all the details of each linked set
Variables	7	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.	7.1 Appendix Table 2. Since data linkages were done at MCHP, we do not have a complete list of codes but detailed definitions are provided.
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.	12.1 Page 4, line 42  12.2 This was not available but we have included multiple references on how robust the data is
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.	12.3 Page 5, line 45-54.
<b>Results</b>			
Participants	13	RECORD 13.1: Describe in detail the selection of the persons included in the study ( <i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in	13.1 Figure 1, pages 5-6, references 17-19

		the text and/or by means of the study flow diagram.	
<b>Discussion</b>			
Limitations	19	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.	19.1 Page 13, lines 17-31
<b>Other Information</b>			
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.	Because this is health data, it is not possible to access the raw data.

\*Reference: Benchimol EI, Smeeth L, Guttman 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.

\*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.