

Protocol for IDA literature review 2018-06-20

Aim of review: To describe current practice of IDA reporting in observational studies in clinical journals

Objectives

- Identify the different elements of IDA reported in the study:
- Compare reporting of IDA in different types of journals
- Compare differences in reporting IDA in methods or results sections

From our framework article: The aim of IDA is to provide a data set and reliable findings on this data set which allows researchers to work with this data set in a responsible manner. The latter requires a full awareness of all data properties needed for a correct analysis and interpretation of the data thus minimizing the risk of producing numerical results or interpretations which are misleading or incorrect.

DATA COLLECTION FORM

1. Background/ general info of the study
 - a. What were the aims?
 - b. What study designs were found (prospective, retrospective, multicenter, cohort)?
 - c. What data sources were used (primary data collection, registries, ...)? Size of study? Location(s)?
 - d. What types of models are used for the primary outcomes (Cox regression, logistic regression, etc.)?
 - e. Was a statistician included as an author? or if not, acknowledged for their help? If a statistician was not involved, what type of training did the person appear to have that did the statistical analysis (if possible to assess)?
2. IDA processes (data screening). Score: 0=not mentioned, 1=mentioned, 2=sufficient detail, NA=not applicable
 - a. Was a statement about data cleaning included? (0-1 score)
 - b. Was a patient flow diagram (STROBE) included? Or is patient inclusion/exclusion adequately described in the text?
 - c. Description of non-outcome variables
 - i. Was information presented in the text?
 - ii. Was information presented in a table?
 - iii. Was information presented in a figure?
 - iv. Did this include item missingness?
 - v. Did this include unit missingness?
 - vi. Description of non-outcome variables for subgroups?
 - vii. Was there a description of methods?

- d. Were associations between non-outcome variables included?
 - i. Was information presented in the text?
 - ii. Was information presented in a table?
 - iii. Was information presented in a figure?
 - e. Description of outcome variables
 - i. Was information presented in the text?
 - ii. Was information presented in a table?
 - iii. Was information presented in a figure?
 - iv. Did this include item missingness?
 - v. Did this include unit missingness?
 - vi. Description of non-outcome variables for subgroups?
 - vii. Was there a description of methods?
 - f. If applicable. For repeatedly measured variables is the frequency of missingness described?
 - g. If applicable. Are data properties described for cluster variables (e.g. centers, year,...)?
 - h. Were there variable transformations ? (e.g. categorization, log-transform)
3. IDA process (updating analysis plan). Was there a change in the intended analyses?
 Score: 0=cannot be determined, 1=yes, but impact not stated, 2=yes, impact was justified, NA=not applicable
- a. Cannot be determined
 - b. Due to unexpected values (e.g leading to exclusion of variables of subjects)
 - c. Due to unexpected population heterogeneity (e.g. leading to subgroups, stratification)
 - d. Due to distribution of a variable (e.g. needing transformation)
 - e. Due to data properties (e.g do not fulfill requirements of model)
 - f. Due to missing data

METHODS

1. Sampling frame. Identify papers by a literature search as follows:
 - a. Journals: NEJM, JCO, The Lancet, JAMA, Circulation
 - b. Years: 2018
 - c. Search engine: Pubmed
 - d. Inclusion/exclusion terms see Pubmed search terms
 - e. Publication type: original research article (i.e. exclude letters, responses, editorials)
 - f. Sample size aim: 25
 - g. Exclude: no clear aim, abstracts, randomized controlled trials, letters, responses, editorials, meta-analysis, review, case-control, sample size less than 50, high dimensional data analysis (e.g genetics)

2. Data collection

- a. Search, browse, abstract for appropriateness (Keep track of number of papers excluded), and download papers (personnel: MH)
- b. Randomize order of papers in each journals and select the first 5 from each journal.
- c. If a paper does not fulfill criteria, the next paper from same journal is chosen. Keep track of number of papers excluded.
- d. A selection 25 of papers will be reviewed by 2 reviewers each to assess agreement.
- e. Collection form (personnel: LL and MH)
 - i. Background of study
 - ii. Text excerpts, Comments, references to tables or figures
 - iii. Scores are 0-1-2 according to questions listed under 2.

3. Summarizing data

- a. PRISMA flow chart
- b. Number of papers identified/selected/included by journal
- c. Location of reporting in a paper, e.g intro, methods, results, discussion, supplement (by category and by journal)
- d. Number of papers that report a category (by journal and by study design or model)
- e. IDA reporting in the methods section will be compared to IDA reporting in the results and discussion sections of the papers.

4. Participants:

- a. Marianne reviews all papers
- b. Carsten, Lara, Saskia, Werner review at least 5 papers
- c. Lara summarizes results
- d. Lara and Marianne write first draft of manuscript