14. Descriptive data:

14 (a). Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.

Explanation

Readers need descriptions of study participants and their exposures to judge the generalisability of the findings. Information about potential confounders, including whether and how they were measured, influences judgments about study validity. We advise authors to summarize continuous variables for each study group by giving the mean and standard deviation, or when the data have an asymmetrical distribution, as is often the case, the median and percentile range (e.g., 25th and 75th percentiles). Variables that make up a small number of ordered categories (such as stages of disease I to IV) should not be presented as continuous variables; it is preferable to give numbers and proportions for each category (see also Box 4). In studies that compare groups, the descriptive characteristics and numbers should be given by group, as in the example above. Inferential measures such as standard errors and confidence intervals should not be used to describe the variability of characteristics, and significance tests should be avoided in descriptive tables. Also, P values are not an appropriate criterion for selecting which confounders to adjust for in analysis; even small differences in a confounder that has a strong effect on the outcome can be important. In cohort studies, it may be useful to document how an exposure relates to other characteristics and potential confounders. Authors could present this information in a table with columns for participants in two or more exposure categories, which permits to judge the differences in confounders between these categories. In case-control studies potential confounders cannot be judged by comparing cases and controls. Control persons represent the source population and will usually be different from the cases in many respects. For example, in a study of oral contraceptives and myocardial infarction, a sample of young women with infarction more often had risk factors for that disease, such as high serum cholesterol, smoking and a positive family history, than the control group. This does not influence the assessment of the effect of oral contraceptives, as long as the prescription of oral contraceptives was not guided by the presence of these risk factors—e.g., because the risk factors were only established after the event (see also Box 5). In case-control studies the equivalent of comparing exposed and non-exposed for the presence of potential confounders (as is done in cohorts) can be achieved by exploring the source population of the cases: if the control group is large enough and represents the source population, exposed and unexposed controls can be compared for potential confounders.

(b). Indicate the number of participants with missing data for each variable of interest.

Example (a table refer to STROBE)

Explanation

As missing data may bias or affect generalisability of results, authors should tell readers amounts of missing data for exposures, potential confounders, and other important characteristics of patients (see also item 12c and Box 6). In a cohort study, authors should report the extent of loss to follow-up (with reasons), since incomplete follow-up may bias findings (see also items 12d and 13). We advise authors to use their tables and figures to enumerate amounts of missing data.

(c). Cohort study: Summarise follow-up time—e.g., average and total amount.

Example

‘‘During the 4366 person-years of follow-up (median 5.4, maximum 8.3 years), 265 subjects were diagnosed as having dementia, including 202 with Alzheimer’s disease’’.

Explanation

Readers need to know the duration and extent of follow-up for the available outcome data. Authors can present a summary of the average follow-up with either the mean or median follow-up time or both. The mean allows a reader to calculate the total number of person-years by multiplying it with the number of study participants. Authors also may present minimum and maximum times or percentiles of the distribution to show readers the spread of follow-up times. They may report total person-years of follow-up or some indication of the proportion of potential data that was captured [148]. All such information may be presented separately for participants in two or more exposure categories. Almost half of 132 articles in cancer journals (mostly cohort studies) did not give any summary of length of follow-up .