12. Statistical methods:

1. a). Describe all statistical methods, including those used to control for confounding.

Example

‘‘The adjusted relative risk was calculated using the MantelHaenszel technique, when evaluating if confounding by age or gender was present in the groups compared. The 95% confidence interval (CI) was computed around the adjusted relative risk, using the variance according to Greenland and Robins and Robins et al.’’

Explanation

In general, there is no one correct statistical analysis but, rather, several possibilities that may address the same question, but make different assumptions. Regardless, investigators should pre-determine analyses at least for the primary study objectives in a study protocol. Often additional analyses are needed, either instead of, or as well as, those originally envisaged, and these may sometimes be motivated by the data. When a study is reported, authors should tell readers whether particular analyses were suggested by data inspection. Even though the distinction between pre-specified and exploratory analyses may sometimes be blurred, authors should clarify reasons for particular analyses. If groups being compared are not similar with regard to some characteristics, adjustment should be made for possible confounding variables by stratification or by multivariable regression (see Box 5). Often, the study design determines which type of regression analysis is chosen. For instance, Cox proportional hazard regression is commonly used in cohort studies. whereas logistic regression is often the method of choice in case-control studies. Analysts should fully describe specific procedures for variable selection and not only present results from the final model. If model comparisons are made to narrow down a list of potential confounders for inclusion in a final model, this process should be described. It is helpful to tell readers if one or two covariates are responsible for a great deal of the apparent confounding in a data analysis. Other statistical analyses such as imputation procedures, data transformation, and calculations of attributable risks should also be described. Nonstandard or novel approaches should be referenced and the statistical software used reported. As a guiding principle, we advise statistical methods be described ‘‘with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results’’. In an empirical study, only 93 of 169 articles (55%) reporting adjustment for confounding clearly stated how continuous and multi-category variables were entered into the statistical model. Another study found that among 67 articles in which statistical analyses were adjusted for confounders, it was mostly unclear how confounders were chosen.

1. Describe any methods used to examine subgroups and interactions.

Example ‘‘Sex differences in susceptibility to the 3 lifestyle-related risk factors studied were explored by testing for biological interaction according to Rothman: a new composite variable with 4 categories (a b , ab þ, aþb , and a þb þ) was redefined for sex and a dichotomous exposure of interest where a and b denote absence of exposure. RR was calculated for each category after adjustment for age. An interaction effect is defined as departure from additivity of absolute effects, and excess RR caused by interaction (RERI) was calculated: RERI ¼ RRðaþbþÞ RRðabþÞ RRðaþbÞ 1 where RR(a þb þ) denotes RR among those exposed to both factors where RR(a b ) is used as reference category (RR ¼ 1.0). Ninety-five percent CIs were calculated as proposed by Hosmer and Lemeshow. RERI of 0 means no interaction’’.

Explanation

As discussed in detail under item 17, many debate the use and value of analyses restricted to subgroups of the study population. Subgroup analyses are nevertheless often done [4]. Readers need to know which subgroup analyses were planned in advance, and which arose while analysing the data. Also, it is important to explain what methods were used to examine whether effects or associations differed across groups (see item 17). Interaction relates to the situation when one factor modifies the effect of another (therefore also called ‘effect modification’). The joint action of two factors can be characterized in two ways: on an additive scale, in terms of risk differences; or on a multiplicative scale, in terms of relative risk (see Box 8). Many authors and readers may have their own preference about the way interactions should be analysed. Still, they may be interested to know to what extent the joint effect of exposures differs from the separate effects. There is consensus that the additive scale, which uses absolute risks, is more appropriate for public health and clinical decision making . Whatever view is taken, this should be clearly presented to the reader, as is done in the example above. A lay-out presenting separate effects of both exposures as well as their joint effect, each relative to no exposure, might be most informative. It is presented in the example for interaction under item 17, and the calculations on the different scales are explained in Box 8. 12

(c). Explain how missing data were addressed.

Example

‘‘Our missing data analysis procedures used missing at random (MAR) assumptions. We used the MICE (multivariate imputation by chained equations) method of multiple multivariate imputation in STATA. We independently analysed 10 copies of the data, each with missing values suitably imputed, in the multivariate logistic regression analyses. We averaged estimates of the variables to give a single mean estimate and adjusted standard errors according to Rubin’s rules’’.

Explanation

Missing data are common in observational research. Questionnaires posted to study participants are not always filled in completely, participants may not attend all follow-up visits and routine data sources and clinical databases are often incomplete. Despite its ubiquity and importance, few papers report in detail on the problem of missing data. Investigators may use any of several approaches to address missing data. We describe some strengths and limitations of various approaches in Box 6. We advise that authors report the number of missing values for each variable of interest (exposures, outcomes, confounders) and for each step in the analysis. Authors should give reasons for missing values if possible, and indicate how many individuals were excluded because of missing data when describing the flow of participants through the study (see also item 13). For analyses that account for missing data, authors should describe the nature of the analysis (e.g., multiple imputation) and the assumptions that were made (e.g., missing at random, see Box 6).

(d). Cohort study: If applicable, describe how loss to follow-up was addressed.

Example

‘‘In treatment programmes with active follow-up, those lost to follow-up and those followed-up at 1 year had similar baseline CD4 cell counts (median 115 cells per lL and 123 cells per lL), whereas patients lost to follow-up in programmes with no active follow-up procedures had considerably lower CD4 cell counts than those followed-up (median 64 cells per lL and 123 cells per lL). (...) Treatment programmes with passive follow-up were excluded from subsequent analyses’’ [116]. Explanation Cohort studies are analysed using life table methods or other approaches that are based on the person-time of follow-up and time to developing the disease of interest. Among individuals who remain free of the disease at the end of their observation period, the amount of follow-up time is assumed to be unrelated to the probability of developing the outcome. This will be the case if follow-up ends on a fixed date or at a particular age. Loss to follow-up occurs when participants withdraw from a study before that date. This may hamper the validity of a study if loss to follow-up occurs selectively in exposed individuals, or in persons at high risk of developing the disease (‘informative censoring’). In the example above, patients lost to follow-up in treatment programmes with no active follow-up had fewer CD4 helper cells than those remaining under observation and were therefore at higher risk of dying [116]. It is important to distinguish persons who reach the end of the study from those lost to follow-up. Unfortunately, statistical software usually does not distinguish between the two situations: in both cases follow-up time is automatically truncated (‘censored’) at the end of the observation period. Investigators therefore need to decide, ideally at the stage of planning the study, how they will deal with loss to follow-up

When few patients are lost, investigators may either exclude individuals with incomplete follow-up, or treat them as if they withdrew alive at either the date of loss to follow-up or the end of the study. We advise authors to report how many patients were lost to follow-up and what censoring strategies they used.

(d). Case-control study: If applicable, explain how matching of cases and controls was addressed. Example

‘‘We used McNemar’s test, paired t test, and conditional logistic regression analysis to compare dementia patients with their matched controls for cardiovascular risk factors, the occurrence of spontaneous cerebral emboli, carotid disease, and venous to arterial circulation shunt’’ .

Explanation

In individually matched case-control studies a crude analysis of the odds ratio, ignoring the matching, usually leads to an estimation that is biased towards unity (see Box 2). A matched analysis is therefore often necessary. This can intuitively be understood as a stratified analysis: each case is seen as one stratum with his or her set of matched controls. The analysis rests on considering whether the case is more often exposed than the controls, despite having made them alike regarding the matching variables. Investigators can do such a stratified analysis using the Mantel-Haenszel method on a ‘matched’ 2 by 2 table. In its simplest form the odds ratio becomes the ratio of pairs that are discordant for the exposure variable. If matching was done for variables like age and sex that are universal attributes, the analysis needs not retain the individual, person-to-person matching: a simple analysis in categories of age and sex is sufficient . For other matching variables, such as neighbourhood, sibship, or friendship, however, each matched set should be considered its own stratum. In individually matched studies, the most widely used method of analysis is conditional logistic regression, in which each case and their controls are considered together. The conditional method is necessary when the number of controls varies among cases, and when, in addition to the matching variables, other variables need to be adjusted for. To allow readers to judge whether the matched design was appropriately taken into account in the analysis, we recommend that authors describe in detail what statistical methods were used to analyse the data. If taking the matching into account does have little effect on the estimates, authors may choose to present an unmatched analysis.

(d). Cross-sectional study: If applicable, describe analytical methods taking account of sampling strategy. Example ‘‘The standard errors (SE) were calculated using the Taylor expansion method to estimate the sampling errors of estimators based on the complex sample design. (...) The overall design effect for diastolic blood pressure was found to be 1.9 for men and 1.8 for women and, for systolic blood pressure, it was 1.9 for men and 2.0 for women’’.

Explanation

Most cross-sectional studies use a pre-specified sampling strategy to select participants from a source population. Sampling may be more complex than taking a simple random sample, however. It may include several stages and clustering of participants (e.g., in districts or villages). Proportionate stratification may ensure that subgroups with a specific characteristic are correctly represented. Disproportionate stratification may be useful to over-sample a subgroup of particular interest. An estimate of association derived from a complex sample may be more or less precise than that derived from a simple random sample. Measures of precision such as standard error or confidence interval should be corrected using the design effect, a ratio measure that describes how much precision is gained or lost if a more complex sampling strategy is used instead of simple random sampling. Most complex sampling techniques lead to a decrease of precision, resulting in a design effect greater than 1. We advise that authors clearly state the method used to adjust for complex sampling strategies so that readers may understand how the chosen sampling method influenced the precision of the obtained estimates. For instance, with clustered sampling, the implicit trade-off between easier data collection and loss of precision is transparent if the design effect is reported. In the example, the calculated design effects of 1.9 for men indicates that the actual sample size would need to be 1.9 times greater than with simple random sampling for the resulting estimates to have equal precision.

(e)Describe any sensitivity analyses.

Example

‘‘Because we had a relatively higher proportion of ‘missing’ dead patients with insufficient data (38/148¼25.7%) as compared to live patients (15/437¼3.4%) (...), it is possible that this might have biased the results. We have, therefore, carried out a sensitivity analysis. We have assumed that the proportion of women using oral contraceptives in the study group applies to the whole (19.1% for dead, and 11.4% for live patients), and then applied two extreme scenarios: either all the exposed missing patients used second generation pills or they all used third-generation pills’’.

Explanation

Sensitivity analyses are useful to investigate whether or not the main results are consistent with those obtained with alternative analysis strategies or assumptions. Issues that may be examined include the criteria for inclusion in analyses, the definitions of exposures or outcomes, which confounding variables merit adjustment, the handling of missing data, possible selection bias or bias from inaccurate or inconsistent measurement of exposure, disease and other variables, and specific analysis choices, such as the treatment of quantitative variables (see item 11). Sophisticated methods are increasingly used to simultaneously model the influence of several biases or assumptions. In 1959 Cornfield et al. famously showed that a relative risk of 9 for cigarette smoking and lung cancer was extremely unlikely to be due to any conceivable confounder, since the confounder would need to be at least nine times as prevalent in smokers as in non-smokers. This analysis did not rule out the possibility that such a factor was present, but it did identify the prevalence such a factor would need to have. The same approach was recently used to identify plausible confounding factors that could explain the association between childhood leukaemia and living near electric power lines. More generally, sensitivity analyses can be used to identify the degree of confounding, selection bias, or information bias required to distort an association. One important, perhaps under recognised, use of sensitivity analysis is when a study shows little or no association between an exposure and an outcome and it is plausible that confounding or other biases toward the null are present.