Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration

Jan P. Vandenbroucke[1], Erik von Elm[2,3], Douglas G. Altman[4], Peter C. Gøtzsche[5], Cynthia D. Mulrow[6], Stuart J. Pocock[7], Charles Poole[8], James J. Schlesselman[9], Matthias Egger[2,10*] for the STROBE Initiative

1 Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, **2** Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland, **3** Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany, **4** Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, United Kingdom, **5** Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark, **6** University of Texas Health Science Center, San Antonio, United States of America, **7** Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom, **8** Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, United States of America, **9** Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, and University of Pittsburgh Cancer Institute, Pittsburgh, United States of America, **10** Department of Social Medicine, University of Bristol, Bristol, United Kingdom

Funding: The initial STROBE workshop was funded by the European Science Foundation (ESF). Additional funding was received from the Medical Research Council Health Services Research Collaboration and the National Health Services Research & Development Methodology Programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Citation: Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLoS Med 4(10): e297. doi:10.1371/journal.pmed.0040297

Received: July 20, 2007
Accepted: August 30, 2007
Published: October 16, 2007

Abbreviations: CI, confidence interval; RERI, Relative Excess Risk from Interaction; RR, relative risk; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology

* To whom correspondence should be addressed. E-mail: strobe@ispm.unibe.ch

ABSTRACT

Much medical research is observational. The reporting of observational studies is often of insufficient quality. Poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalisability of its results. Taking into account empirical evidence and theoretical considerations, a group of methodologists, researchers, and editors developed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations to improve the quality of reporting of observational studies. The STROBE Statement consists of a checklist of 22 items, which relate to the title, abstract, introduction, methods, results and discussion sections of articles. Eighteen items are common to cohort studies, case-control studies and cross-sectional studies and four are specific to each of the three study designs. The STROBE Statement provides guidance to authors about how to improve the reporting of observational studies and facilitates critical appraisal and interpretation of studies by reviewers, journal editors and readers. This explanatory and elaboration document is intended to enhance the use, understanding, and dissemination of the STROBE Statement. The meaning and rationale for each checklist item are presented. For each item, one or several published examples and, where possible, references to relevant empirical studies and methodological literature are provided. Examples of useful flow diagrams are also included. The STROBE Statement, this document, and the associated Web site (http://www.strobe-statement.org/) should be helpful resources to improve reporting of observational research.

Introduction

Rational health care practices require knowledge about the aetiology and pathogenesis, diagnosis, prognosis and treatment of diseases. Randomised trials provide valuable evidence about treatments and other interventions. However, much of clinical or public health knowledge comes from observational research [1]. About nine of ten research papers published in clinical speciality journals describe observational research [2,3].

The STROBE Statement

Reporting of observational research is often not detailed and clear enough to assess the strengths and weaknesses of the investigation [4,5]. To improve the reporting of observational research, we developed a checklist of items that should be addressed: the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (Table 1). Items relate to title, abstract, introduction, methods, results and discussion sections of articles. The STROBE Statement has recently been published in several journals [6]. Our aim is to ensure clear presentation of what was planned, done, and found in an observational study. We stress that the recommendations are not prescriptions for setting up or conducting studies, nor do they dictate methodology or mandate a uniform presentation.

STROBE provides general reporting recommendations for descriptive observational studies and studies that investigate associations between exposures and health outcomes. STROBE addresses the three main types of observational studies: cohort, case-control and cross-sectional studies. Authors use diverse terminology to describe these study designs. For instance, 'follow-up study' and 'longitudinal study' are used as synonyms for 'cohort study', and 'prevalence study' as synonymous with 'cross-sectional study'. We chose the present terminology because it is in common use. Unfortunately, terminology is often used incorrectly [7] or imprecisely [8]. In Box 1 we describe the hallmarks of the three study designs.

The Scope of Observational Research

Observational studies serve a wide range of purposes: from reporting a first hint of a potential cause of a disease, to verifying the magnitude of previously reported associations. Ideas for studies may arise from clinical observations or from biologic insight. Ideas may also arise from informal looks at data that lead to further explorations. Like a clinician who has seen thousands of patients, and notes one that strikes her attention, the researcher may note something special in the data. Adjusting for multiple looks at the data may not be possible or desirable [9], but further studies to confirm or refute initial observations are often needed [10]. Existing data may be used to examine new ideas about potential causal factors, and may be sufficient for rejection or confirmation. In other instances, studies follow that are specifically designed to overcome potential problems with previous reports. The latter studies will gather new data and will be planned for that purpose, in contrast to analyses of existing data. This leads to diverse viewpoints, e.g., on the merits of looking at subgroups or the importance of a predetermined sample size. STROBE tries to accommodate these diverse uses of observational research - from discovery to refutation or confirmation. Where necessary we will indicate in what circumstances specific recommendations apply.

How to Use This Paper

This paper is linked to the shorter STROBE paper that introduced the items of the checklist in several journals [6], and forms an integral part of the STROBE Statement. Our intention is to explain how to report research well, not how research should be done. We offer a detailed explanation for each checklist item. Each explanation is preceded by an example of what we consider transparent reporting. This does not mean that the study from which the example was taken was uniformly well reported or well done; nor does it mean that its findings were reliable, in the sense that they were later confirmed by others: it only means that this particular item was well reported in that study. In addition to explanations and examples we included Boxes 1–8 with supplementary information. These are intended for readers who want to refresh their memories about some theoretical points, or be quickly informed about technical background details. A full understanding of these points may require studying the textbooks or methodological papers that are cited.

STROBE recommendations do not specifically address topics such as genetic linkage studies, infectious disease modelling or case reports and case series [11,12]. As many of the key elements in STROBE apply to these designs, authors who report such studies may nevertheless find our recommendations useful. For authors of observational studies that specifically address diagnostic tests, tumour markers and genetic associations, STARD [13], REMARK [14], and STREGA [15] recommendations may be particularly useful.

The Items in the STROBE Checklist

We now discuss and explain the 22 items in the STROBE checklist (Table 1), and give published examples for each item. Some examples have been edited by removing citations or spelling out abbreviations. Eighteen items apply to all three study designs whereas four are design-specific. Starred items (for example item 8*) indicate that the information should be given separately for cases and controls in case-control studies, or exposed and unexposed groups in cohort and cross-sectional studies. We advise authors to address all items somewhere in their paper, but we do not prescribe a precise location or order. For instance, we discuss the reporting of results under a number of separate items, while recognizing that authors might address several items within a single section of text or in a table.

The Items

TITLE AND ABSTRACT

1 (a). Indicate the study's design with a commonly used term in the title or the abstract.

Example

"Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study" [18].

Explanation

Readers should be able to easily identify the design that was used from the title or abstract. An explicit, commonly used term for the study design also helps ensure correct indexing of articles in electronic databases [19,20].

Table 1. The STROBE Statement—Checklist of Items That Should Be Addressed in Reports of Observational Studies

	Item number	Recommendation
TITLE and ABSTRACT	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
INTRODUCTION		
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
METHODS		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants (b) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
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
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why
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) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses
RESULTS		
Participants	13*	(a) Report the numbers of individuals at each stage of the study—e.g., 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
Descriptive data	14*	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study*—Summarise follow-up time (e.g., average and total amount)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *Case-control study*—Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study*—Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 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
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses
DISCUSSION		
Key results	18	Summarise key results with reference to study objectives
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
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
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

*Give such information separately for cases and controls in case-control studies, and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
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/). Separate versions of the checklist for cohort, case-control, and cross-sectional studies are available on the STROBE Web site at http://www.strobe-statement.org/.
doi:10.1371/journal.pmed.0040297.t001

Box 1. Main study designs covered by STROBE

Cohort, case-control, and cross-sectional designs represent different approaches of investigating the occurrence of health-related events in a given population and time period. These studies may address many types of health-related events, including disease or disease remission, disability or complications, death or survival, and the occurrence of risk factors.

In **cohort studies**, the investigators follow people over time. They obtain information about people and their exposures at baseline, let time pass, and then assess the occurrence of outcomes. Investigators commonly make contrasts between individuals who are exposed and not exposed or among groups of individuals with different categories of exposure. Investigators may assess several different outcomes, and examine exposure and outcome variables at multiple points during follow-up. Closed cohorts (for example birth cohorts) enrol a defined number of participants at study onset and follow them from that time forward, often at set intervals up to a fixed end date. In open cohorts the study population is dynamic: people enter and leave the population at different points in time (for example inhabitants of a town). Open cohorts change due to deaths, births, and migration, but the composition of the population with regard to variables such as age and gender may remain approximately constant, especially over a short period of time. In a closed cohort cumulative incidences (risks) and incidence rates can be estimated; when exposed and unexposed groups are compared, this leads to risk ratio or rate ratio estimates. Open cohorts estimate incidence rates and rate ratios.

In **case-control studies**, investigators compare exposures between people with a particular disease outcome (cases) and people without that outcome (controls). Investigators aim to collect cases and controls that are representative of an underlying cohort or a cross-section of a population. That population can be defined geographically, but also more loosely as the catchment area of health care facilities. The case sample may be 100% or a large fraction of available cases, while the control sample usually is only a small fraction of the people who do not have the pertinent outcome. Controls represent the cohort or population of people from which the cases arose. Investigators calculate the ratio of the odds of exposures to putative causes of the disease among cases and controls (see Box 7). Depending on the sampling strategy for cases and controls and the nature of the population studied, the odds ratio obtained in a case-control study is interpreted as the risk ratio, rate ratio or (prevalence) odds ratio [16,17]. The majority of published case-control studies sample open cohorts and so allow direct estimations of rate ratios.

In **cross-sectional studies**, investigators assess all individuals in a sample at the same point in time, often to examine the prevalence of exposures, risk factors or disease. Some cross-sectional studies are analytical and aim to quantify potential causal associations between exposures and disease. Such studies may be analysed like a cohort study by comparing disease prevalence between exposure groups. They may also be analysed like a case-control study by comparing the odds of exposure between groups with and without disease. A difficulty that can occur in any design but is particularly clear in cross-sectional studies is to establish that an exposure preceded the disease, although the time order of exposure and outcome may sometimes be clear. In a study in which the exposure variable is congenital or genetic, for example, we can be confident that the exposure preceded the disease, even if we are measuring both at the same time.

1 (b). Provide in the abstract an informative and balanced summary of what was done and what was found.

Example

"**Background:** The expected survival of HIV-infected patients is of major public health interest.

Objective: To estimate survival time and age-specific mortality rates of an HIV-infected population compared with that of the general population.

Design: Population-based cohort study.

Setting: All HIV-infected persons receiving care in Denmark from 1995 to 2005.

Patients: Each member of the nationwide Danish HIV Cohort Study was matched with as many as 99 persons from the general population according to sex, date of birth, and municipality of residence.

Measurements: The authors computed Kaplan–Meier life tables with age as the time scale to estimate survival from age 25 years. Patients with HIV infection and corresponding persons from the general population were observed from the date of the patient's HIV diagnosis until death, emigration, or 1 May 2005.

Results: 3990 HIV-infected patients and 379 872 persons from the general population were included in the study, yielding 22 744 (median, 5.8 y/person) and 2 689 287 (median, 8.4 years/person) person-years of observation. Three percent of participants were lost to follow-up. From age 25 years, the median survival was 19.9 years (95% CI, 18.5 to 21.3) among patients with HIV infection and 51.1 years (CI, 50.9 to 51.5) among the general population. For HIV-infected patients, survival increased to 32.5 years (CI, 29.4 to 34.7) during the 2000 to 2005 period. In the subgroup that excluded persons with known hepatitis C coinfection (16%), median survival was 38.9 years (CI, 35.4 to 40.1) during this same period. The relative mortality rates for patients with HIV infection compared with those for the general population decreased with increasing age, whereas the excess mortality rate increased with increasing age.

Limitations: The observed mortality rates are assumed to apply beyond the current maximum observation time of 10 years.

Conclusions: The estimated median survival is more than 35 years for a young person diagnosed with HIV infection in the late highly active antiretroviral therapy era. However, an ongoing effort is still needed to further reduce mortality rates for these persons compared with the general population" [21].

Explanation

The abstract provides key information that enables readers to understand a study and decide whether to read the article. Typical components include a statement of the research question, a short description of methods and results, and a conclusion [22]. Abstracts should summarize key details of studies and should only present information that is provided in the article. We advise presenting key results in a numerical form that includes numbers of participants, estimates of associations and appropriate measures of variability and uncertainty (e.g., odds ratios with confidence intervals). We regard it insufficient to state only that an exposure is or is not significantly associated with an outcome.

A series of headings pertaining to the background, design, conduct, and analysis of a study may help readers acquire the essential information rapidly [23]. Many journals require such structured abstracts, which tend to be of higher quality and more readily informative than unstructured summaries [24,25].

INTRODUCTION

The Introduction section should describe why the study was done and what questions and hypotheses it addresses. It should allow others to understand the study's context and judge its potential contribution to current knowledge.

2. Background/rationale: Explain the scientific background and rationale for the investigation being reported.

Example

"Concerns about the rising prevalence of obesity in children and adolescents have focused on the well documented associations between childhood obesity and increased cardiovascular risk and mortality in adulthood. Childhood obesity has considerable social and psychological consequences within childhood and adolescence, yet little is known about social, socioeconomic, and psychological consequences in adult life. A recent systematic review found no longitudinal studies on the outcomes of childhood obesity other than physical health outcomes and only two longitudinal studies of the socioeconomic effects of obesity in adolescence. Gortmaker et al. found that US women who had been obese in late adolescence in 1981 were less likely to be married and had lower incomes seven years later than women who had not been overweight, while men who had been overweight were less likely to be married. Sargent et al. found that UK women, but not men, who had been obese at 16 years in 1974 earned 7.4% less than their non-obese peers at age 23. (. . .) We used longitudinal data from the 1970 British birth cohort to examine the adult socioeconomic, educational, social, and psychological outcomes of childhood obesity" [26].

Explanation

The scientific background of the study provides important context for readers. It sets the stage for the study and describes its focus. It gives an overview of what is known on a topic and what gaps in current knowledge are addressed by the study. Background material should note recent pertinent studies and any systematic reviews of pertinent studies.

3. Objectives: State specific objectives, including any prespecified hypotheses.

Example

"Our primary objectives were to 1) determine the prevalence of domestic violence among female patients presenting to four community-based, primary care, adult medicine practices that serve patients of diverse socioeconomic background and 2) identify demographic and clinical differences between currently abused patients and patients not currently being abused " [27].

Explanation

Objectives are the detailed aims of the study. Well crafted objectives specify populations, exposures and outcomes, and parameters that will be estimated. They may be formulated as specific hypotheses or as questions that the study was designed to address. In some situations objectives may be less specific, for example, in early discovery phases. Regardless, the report should clearly reflect the investigators' intentions. For example, if important subgroups or additional analyses were not the original aim of the study but arose during data analysis, they should be described accordingly (see also items 4, 17 and 20).

METHODS

The Methods section should describe what was planned and what was done in sufficient detail to allow others to understand the essential aspects of the study, to judge whether the methods were adequate to provide reliable and valid answers, and to assess whether any deviations from the original plan were reasonable.

4. Study design: Present key elements of study design early in the paper.

Example

"We used a case-crossover design, a variation of a case-control design that is appropriate when a brief exposure (driver's phone use) causes a transient rise in the risk of a rare outcome (a crash). We compared a driver's use of a mobile phone at the estimated time of a crash with the same driver's use during another suitable time period. Because drivers are their own controls, the design controls for characteristics of the driver that may affect the risk of a crash but do not change over a short period of time. As it is important that risks during control periods and crash trips are similar, we compared phone activity during the hazard interval (time immediately before the crash) with phone activity during control intervals (equivalent times during which participants were driving but did not crash) in the previous week" [28].

Explanation

We advise presenting key elements of study design early in the methods section (or at the end of the introduction) so that readers can understand the basics of the study. For example, authors should indicate that the study was a cohort study, which followed people over a particular time period, and describe the group of persons that comprised the cohort and their exposure status. Similarly, if the investigation used a case-control design, the cases and controls and their source population should be described. If the study was a cross-sectional survey, the population and the point in time at which the cross-section was taken should be mentioned. When a study is a variant of the three main study types, there is an additional need for clarity. For instance, for a case-crossover study, one of the variants of the case-control design, a succinct description of the principles was given in the example above [28].

We recommend that authors refrain from simply calling a study 'prospective' or 'retrospective' because these terms are ill defined [29]. One usage sees cohort and prospective as synonymous and reserves the word retrospective for case-control studies [30]. A second usage distinguishes prospective and retrospective cohort studies according to the timing of data collection relative to when the idea for the study was developed [31]. A third usage distinguishes prospective and retrospective case-control studies depending on whether the data about the exposure of interest existed when cases were selected [32]. Some advise against using these terms [33], or adopting the alternatives 'concurrent' and 'historical' for describing cohort studies [34]. In STROBE, we do not use the words prospective and retrospective, nor alternatives such as concurrent and historical. We recommend that, whenever authors use these words, they define what they mean. Most importantly, we recommend that authors describe exactly how and when data collection took place.

The first part of the methods section might also be the place to mention whether the report is one of several from a study. If a new report is in line with the original aims of the study, this is usually indicated by referring to an earlier publication and by briefly restating the salient features of the study. However, the aims of a study may also evolve over time.

Researchers often use data for purposes for which they were not originally intended, including, for example, official vital statistics that were collected primarily for administrative purposes, items in questionnaires that originally were only included for completeness, or blood samples that were collected for another purpose. For example, the Physicians' Health Study, a randomized controlled trial of aspirin and carotene, was later used to demonstrate that a point mutation in the factor V gene was associated with an increased risk of venous thrombosis, but not of myocardial infarction or stroke [35]. The secondary use of existing data is a creative part of observational research and does not necessarily make results less credible or less important. However, briefly restating the original aims might help readers understand the context of the research and possible limitations in the data.

5. Setting: Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.

Example

"The Pasitos Cohort Study recruited pregnant women from Women, Infant and Child clinics in Socorro and San Elizario, El Paso County, Texas and maternal-child clinics of the Mexican Social Security Institute in Ciudad Juarez, Mexico from April 1998 to October 2000. At baseline, prior to the birth of the enrolled cohort children, staff interviewed mothers regarding the household environment. In this ongoing cohort study, we target follow-up exams at 6-month intervals beginning at age 6 months" [36].

Explanation

Readers need information on setting and locations to assess the context and generalisability of a study's results. Exposures such as environmental factors and therapies can change over time. Also, study methods may evolve over time. Knowing when a study took place and over what period participants were recruited and followed up places the study in historical context and is important for the interpretation of results.

Information about setting includes recruitment sites or sources (e.g., electoral roll, outpatient clinic, cancer registry, or tertiary care centre). Information about location may refer to the countries, towns, hospitals or practices where the investigation took place. We advise stating dates rather than only describing the length of time periods. There may be different sets of dates for exposure, disease occurrence, recruitment, beginning and end of follow-up, and data collection. Of note, nearly 80% of 132 reports in oncology journals that used survival analysis included the starting and ending dates for accrual of patients, but only 24% also reported the date on which follow-up ended [37].

6. Participants:

6 (a). Cohort study: Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.

Example

"Participants in the Iowa Women's Health Study were a random sample of all women ages 55 to 69 years derived from the state of Iowa automobile driver's license list in 1985, which represented approximately 94% of Iowa women in that age group. (...) Follow-up questionnaires were mailed in October 1987 and August 1989 to assess vital status and address changes. (...) Incident cancers, except for non-melanoma skin cancers, were ascertained by the State Health Registry of Iowa (...). The Iowa Women's Health Study cohort was matched to the registry with combinations of first, last, and maiden names, zip code, birthdate, and social security number" [38].

6 (a). Case-control study: Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls.

Example

"Cutaneous melanoma cases diagnosed in 1999 and 2000 were ascertained through the Iowa Cancer Registry (...). Controls, also identified through the Iowa Cancer Registry, were colorectal cancer patients diagnosed during the same time. Colorectal cancer controls were selected because they are common and have a relatively long survival, and because arsenic exposure has not been conclusively linked to the incidence of colorectal cancer" [39].

6 (a). Cross-sectional study: Give the eligibility criteria, and the sources and methods of selection of participants.

Example

"We retrospectively identified patients with a principal diagnosis of myocardial infarction (code 410) according to the International Classification of Diseases, 9th Revision, Clinical Modification, from codes designating discharge diagnoses, excluding the codes with a fifth digit of 2, which designates a subsequent episode of care (...) A random sample of the entire Medicare cohort with myocardial infarction from February 1994 to July 1995 was selected (...) To be eligible, patients had to present to the hospital after at least 30 minutes but less than 12 hours of chest pain and had to have ST-segment elevation of at least 1 mm on two contiguous leads on the initial electrocardiogram" [40].

Explanation

Detailed descriptions of the study participants help readers understand the applicability of the results. Investigators usually restrict a study population by defining clinical, demographic and other characteristics of eligible participants. Typical eligibility criteria relate to age, gender, diagnosis and comorbid conditions. Despite their importance, eligibility criteria often are not reported adequately. In a survey of observational stroke research, 17 of 49 reports (35%) did not specify eligibility criteria [5].

Eligibility criteria may be presented as inclusion and exclusion criteria, although this distinction is not always necessary or useful. Regardless, we advise authors to report all eligibility criteria and also to describe the group from which the study population was selected (e.g., the general population of a region or country), and the method of recruitment (e.g., referral or self-selection through advertisements).

Knowing details about follow-up procedures, including whether procedures minimized non-response and loss to follow-up and whether the procedures were similar for all participants, informs judgments about the validity of results. For example, in a study that used IgM antibodies to detect acute infections, readers needed to know the interval between blood tests for IgM antibodies so that they could judge whether some infections likely were missed because the

interval between blood tests was too long [41]. In other studies where follow-up procedures differed between exposed and unexposed groups, readers might recognize substantial bias due to unequal ascertainment of events or differences in non-response or loss to follow-up [42]. Accordingly, we advise that researchers describe the methods used for following participants and whether those methods were the same for all participants, and that they describe the completeness of ascertainment of variables (see also item 14).

In case-control studies, the choice of cases and controls is crucial to interpreting the results, and the method of their selection has major implications for study validity. In general, controls should reflect the population from which the cases arose. Various methods are used to sample controls, all with advantages and disadvantages: for cases that arise from a general population, population roster sampling, random digit dialling, neighbourhood or friend controls are used. Neighbourhood or friend controls may present intrinsic matching on exposure [17]. Controls with other diseases may have advantages over population-based controls, in particular for hospital-based cases, because they better reflect the catchment population of a hospital, have greater comparability of recall and ease of recruitment. However, they can present problems if the exposure of interest affects the risk of developing or being hospitalized for the control condition(s) [43,44]. To remedy this problem often a mixture of the best defensible control diseases is used [45].

6 (b). Cohort study: For matched studies, give matching criteria and number of exposed and unexposed.

Example

"For each patient who initially received a statin, we used propensity-based matching to identify one control who did not receive a statin according to the following protocol. First, propensity scores were calculated for each patient in the entire cohort on the basis of an extensive list of factors potentially related to the use of statins or the risk of sepsis. Second, each statin user was matched to a smaller pool of non-statin-users by sex, age (plus or minus 1 year), and index date (plus or minus 3 months). Third, we selected the control with the closest propensity score (within 0.2 SD) to each statin user in a 1:1 fashion and discarded the remaining controls." [46].

6 (b). Case-control study: For matched studies, give matching criteria and the number of controls per case.

Example

"We aimed to select five controls for every case from among individuals in the study population who had no diagnosis of autism or other pervasive developmental disorders (PDD) recorded in their general practice record and who were alive and registered with a participating practice on the date of the PDD diagnosis in the case. Controls were individually matched to cases by year of birth (up to 1 year older or younger), sex, and general practice. For each of 300 cases, five controls could be identified who met all the matching criteria. For the remaining 994, one or more controls was excluded..." [47].

Explanation

Matching is much more common in case-control studies, but occasionally, investigators use matching in cohort studies to make groups comparable at the start of follow-up. Matching in cohort studies makes groups directly comparable for potential confounders and presents fewer intricacies than with case-control studies. For example, it is not necessary to take the matching into account for the estimation of the relative risk [48]. Because matching in cohort studies may increase statistical precision investigators might allow for the matching in their analyses and thus obtain narrower confidence intervals.

In case-control studies matching is done to increase a study's efficiency by ensuring similarity in the distribution of variables between cases and controls, in particular the distribution of potential confounding variables [48,49]. Because matching can be done in various ways, with one or more controls per case, the rationale for the choice of matching variables and the details of the method used should be described. Commonly used forms of matching are frequency matching (also called group matching) and individual matching. In frequency matching, investigators choose controls so that the distribution of matching variables becomes identical or similar to that of cases. Individual matching involves matching one or several controls to each case. Although intuitively appealing and sometimes useful, matching in case-control studies has a number of disadvantages, is not always appropriate, and needs to be taken into account in the analysis (see Box 2).

Even apparently simple matching procedures may be poorly reported. For example, authors may state that controls were matched to cases 'within five years', or using 'five year age bands'. Does this mean that, if a case was 54 years old, the respective control needed to be in the five-year age band 50 to 54, or aged 49 to 59, which is within five years of age 54? If a wide (e.g., 10-year) age band is chosen, there is a danger of residual confounding by age (see also Box 4), for example because controls may then be younger than cases on average.

7. Variables: Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.

Example

"Only major congenital malformations were included in the analyses. Minor anomalies were excluded according to the exclusion list of European Registration of Congenital Anomalies (EUROCAT). If a child had more than one major congenital malformation of one organ system, those malformations were treated as one outcome in the analyses by organ system (...) In the statistical analyses, factors considered potential confounders were maternal age at delivery and number of previous parities. Factors considered potential effect modifiers were maternal age at reimbursement for antiepileptic medication and maternal age at delivery" [55].

Explanation

Authors should define all variables considered for and included in the analysis, including outcomes, exposures, predictors, potential confounders and potential effect modifiers. Disease outcomes require adequately detailed description of the diagnostic criteria. This applies to criteria for cases in a case-control study, disease events during follow-up in a cohort study and prevalent disease in a cross-sectional study. Clear definitions and steps taken to adhere to them are

Box 2. Matching in case-control studies

In any case-control study, sensible choices need to be made on whether to use matching of controls to cases, and if so, what variables to match on, the precise method of matching to use, and the appropriate method of statistical analysis. Not to match at all may mean that the distribution of some key potential confounders (e.g., age, sex) is radically different between cases and controls. Although this could be adjusted for in the analysis there could be a major loss in statistical efficiency.

The use of matching in case-control studies and its interpretation are fraught with difficulties, especially if matching is attempted on several risk factors, some of which may be linked to the exposure of prime interest [50,51]. For example, in a case-control study of myocardial infarction and oral contraceptives nested in a large pharmaco-epidemiologic data base, with information about thousands of women who are available as potential controls, investigators may be tempted to choose matched controls who had similar levels of risk factors to each case of myocardial infarction. One objective is to adjust for factors that might influence the prescription of oral contraceptives and thus to control for *confounding by indication*. However, the result will be a control group that is *no longer representative* of the oral contraceptive use in the source population: controls will be older than the source population because patients with myocardial infarction tend to be older. This has several implications. A crude analysis of the data will produce odds ratios that are usually biased towards unity if the matching factor is associated with the exposure. The solution is to perform a matched or stratified analysis (see item 12d). In addition, because the matched control group ceases to be representative for the population at large, the exposure distribution among the controls can no longer be used to estimate the population attributable fraction (see Box 7) [52]. Also, the effect of the matching factor can no longer be studied, and the search for well-matched controls can be cumbersome – making a design with a non-matched control group preferable because the non-matched controls will be easier to obtain and the control group can be larger. *Overmatching* is another problem, which may reduce the efficiency of matched case-control studies, and, in some situations, introduce bias. Information is lost and the power of the study is reduced if the matching variable is closely associated with the exposure. Then many individuals in the same matched sets will tend to have identical or similar levels of exposures and therefore not contribute relevant information. Matching will introduce irremediable bias if the matching variable is not a confounder but in the causal pathway between exposure and disease. For example, in vitro fertilization is associated with an increased risk of perinatal death, due to an increase in multiple births and low birth weight infants [53]. Matching on plurality or birth weight will bias results towards the null, and this cannot be remedied in the analysis.

Matching is intuitively appealing, but the complexities involved have led methodologists to advise against routine matching in case-control studies. They recommend instead a careful and judicious consideration of each potential matching factor, recognizing that it could instead be measured and used as an adjustment variable without matching on it. In response, there has been a reduction in the number of matching factors employed, an increasing use of frequency matching, which avoids some of the problems discussed above, and more case-control studies with no matching at all [54]. Matching remains most desirable, or even necessary, when the distributions of the confounder (e.g., age) might differ radically between the unmatched comparison groups [48,49].

particularly important for any disease condition of primary interest in the study.

For some studies, 'determinant' or 'predictor' may be appropriate terms for exposure variables and outcomes may be called 'endpoints'. In multivariable models, authors sometimes use 'dependent variable' for an outcome and 'independent variable' or 'explanatory variable' for exposure and confounding variables. The latter is not precise as it does not distinguish exposures from confounders.

If many variables have been measured and included in exploratory analyses in an early discovery phase, consider providing a list with details on each variable in an appendix, additional table or separate publication. Of note, the *International Journal of Epidemiology* recently launched a new section with 'cohort profiles', that includes detailed information on what was measured at different points in time in particular studies [56,57]. Finally, we advise that authors declare all 'candidate variables' considered for statistical analysis, rather than selectively reporting only those included in the final models (see also item 16a) [58,59].

8. Data sources/measurement: 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.

Example 1

"Total caffeine intake was calculated primarily using US Department of Agriculture food composition sources. In these calculations, it was assumed that the content of caffeine was 137 mg per cup of coffee, 47 mg per cup of tea, 46 mg per can or bottle of cola beverage, and 7 mg per serving of chocolate candy. This method of measuring (caffeine) intake was shown to be valid in both the NHS I cohort and a similar cohort study of male health professionals (...) Self-reported diagnosis of hypertension was found to be reliable in the NHS I cohort" [60].

Example 2

"Samples pertaining to matched cases and controls were always analyzed together in the same batch and laboratory personnel were unable to distinguish among cases and controls" [61].

Explanation

The way in which exposures, confounders and outcomes were measured affects the reliability and validity of a study. Measurement error and misclassification of exposures or outcomes can make it more difficult to detect cause-effect relationships, or may produce spurious relationships. Error in measurement of potential confounders can increase the risk of residual confounding [62,63]. It is helpful, therefore, if authors report the findings of any studies of the validity or reliability of assessments or measurements, including details of the reference standard that was used. Rather than simply citing validation studies (as in the first example), we advise that authors give the estimated validity or reliability, which can then be used for measurement error adjustment or sensitivity analyses (see items 12e and 17).

In addition, it is important to know if groups being compared differed with respect to the way in which the data were collected. This may be important for laboratory examinations (as in the second example) and other situations. For instance, if an interviewer first questions all the cases and then the controls, or vice versa, bias is possible because of the learning curve; solutions such as randomising the order of interviewing may avoid this problem. Information bias may also arise if the compared groups are not given the same diagnostic tests or if one group receives more tests of the same kind than another (see also item 9).

9. Bias: Describe any efforts to address potential sources of bias.

Example 1

"In most case-control studies of suicide, the control group comprises living individuals but we decided to have a control group of people who had died of other causes (...). With a

control group of deceased individuals, the sources of information used to assess risk factors are informants who have recently experienced the death of a family member or close associate - and are therefore more comparable to the sources of information in the suicide group than if living controls were used" [64].

Example 2

"Detection bias could influence the association between Type 2 diabetes mellitus (T2DM) and primary open-angle glaucoma (POAG) if women with T2DM were under closer ophthalmic surveillance than women without this condition. We compared the mean number of eye examinations reported by women with and without diabetes. We also recalculated the relative risk for POAG with additional control for covariates associated with more careful ocular surveillance (a self-report of cataract, macular degeneration, number of eye examinations, and number of physical examinations)" [65].

Explanation

Biased studies produce results that differ systematically from the truth (see also Box 3). It is important for a reader to know what measures were taken during the conduct of a study to reduce the potential of bias. Ideally, investigators carefully consider potential sources of bias when they plan their study. At the stage of reporting, we recommend that authors always assess the likelihood of relevant biases. Specifically, the direction and magnitude of bias should be discussed and, if possible, estimated. For instance, in case-control studies information bias can occur, but may be reduced by selecting an appropriate control group, as in the first example [64]. Differences in the medical surveillance of participants were a problem in the second example [65]. Consequently, the authors provide more detail about the additional data they collected to tackle this problem. When investigators have set up quality control programs for data collection to counter a possible "drift" in measurements of variables in longitudinal studies, or to keep variability at a minimum when multiple observers are used, these should be described.

Unfortunately, authors often do not address important biases when reporting their results. Among 43 case-control and cohort studies published from 1990 to 1994 that investigated the risk of second cancers in patients with a history of cancer, medical surveillance bias was mentioned in only 5 articles [66]. A survey of reports of mental health research published during 1998 in three psychiatric journals found that only 13% of 392 articles mentioned response bias [67]. A survey of cohort studies in stroke research found that 14 of 49 (28%) articles published from 1999 to 2003 addressed potential selection bias in the recruitment of study participants and 35 (71%) mentioned the possibility that any type of bias may have affected results [5].

Box 3. Bias

Bias is a systematic deviation of a study's result from a true value. Typically, it is introduced during the design or implementation of a study and cannot be remedied later. Bias and confounding are not synonymous. Bias arises from flawed information or subject selection so that a wrong association is found. Confounding produces relations that are factually right, but that cannot be interpreted causally because some underlying, unaccounted for factor is associated with both exposure and outcome (see Box 5). Also, bias needs to be distinguished from random error, a deviation from a true value caused by statistical fluctuations (in either direction) in the measured data. Many possible sources of bias have been described and a variety of terms are used [68,69]. We find two simple categories helpful: information bias and selection bias.

Information bias occurs when systematic differences in the completeness or the accuracy of data lead to differential misclassification of individuals regarding exposures or outcomes. For instance, if diabetic women receive more regular and thorough eye examinations, the ascertainment of glaucoma will be more complete than in women without diabetes (see item 9) [65]. Patients receiving a drug that causes non-specific stomach discomfort may undergo gastroscopy more often and have more ulcers detected than patients not receiving the drug – even if the drug does not cause more ulcers. This type of information bias is also called 'detection bias' or 'medical surveillance bias'. One way to assess its influence is to measure the intensity of medical surveillance in the different study groups, and to adjust for it in statistical analyses. In case-control studies information bias occurs if cases recall past exposures more or less accurately than controls without that disease, or if they are more or less willing to report them (also called 'recall bias'). 'Interviewer bias' can occur if interviewers are aware of the study hypothesis and subconsciously or consciously gather data selectively [70]. Some form of blinding of study participants and researchers is therefore often valuable.

Selection bias may be introduced in case-control studies if the probability of including cases or controls is associated with exposure. For instance, a doctor recruiting participants for a study on deep-vein thrombosis might diagnose this disease in a woman who has leg complaints and takes oral contraceptives. But she might not diagnose deep-vein thrombosis in a woman with similar complaints who is not taking such medication. Such bias may be countered by using cases and controls that were referred in the same way to the diagnostic service [71]. Similarly, the use of disease registers may introduce selection bias: if a possible relationship between an exposure and a disease is known, cases may be more likely to be submitted to a register if they have been exposed to the suspected causative agent [72]. 'Response bias' is another type of selection bias that occurs if differences in characteristics between those who respond and those who decline participation in a study affect estimates of prevalence, incidence and, in some circumstances, associations. In general, selection bias affects the internal validity of a study. This is different from problems that may arise with the selection of participants for a study in general, which affects the external rather than the internal validity of a study (also see item 21).

10. Study size: Explain how the study size was arrived at.

Example 1

"The number of cases in the area during the study period determined the sample size" [73].

Example 2

"A survey of postnatal depression in the region had documented a prevalence of 19.8%. Assuming depression in mothers with normal weight children to be 20% and an odds ratio of 3 for depression in mothers with a malnourished child we needed 72 case-control sets (one case to one control) with an 80% power and 5% significance" [74].

Explanation

A study should be large enough to obtain a point estimate with a sufficiently narrow confidence interval to meaningfully answer a research question. Large samples are needed to distinguish a small association from no association. Small studies often provide valuable information, but wide confidence intervals may indicate that they contribute less to current knowledge in comparison with studies providing estimates with narrower confidence intervals. Also, small studies that show 'interesting' or 'statistically significant' associations are published more frequently than small studies

that do not have 'significant' findings. While these studies may provide an early signal in the context of discovery, readers should be informed of their potential weaknesses.

The importance of sample size determination in observational studies depends on the context. If an analysis is performed on data that were already available for other purposes, the main question is whether the analysis of the data will produce results with sufficient statistical precision to contribute substantially to the literature, and sample size considerations will be informal. Formal, *a priori* calculation of sample size may be useful when planning a new study [75,76]. Such calculations are associated with more uncertainty than implied by the single number that is generally produced. For example, estimates of the rate of the event of interest or other assumptions central to calculations are commonly imprecise, if not guesswork [77]. The precision obtained in the final analysis can often not be determined beforehand because it will be reduced by inclusion of confounding variables in multivariable analyses [78], the degree of precision with which key variables can be measured [79], and the exclusion of some individuals.

Few epidemiological studies explain or report deliberations about sample size [4,5]. We encourage investigators to report pertinent formal sample size calculations if they were done. In other situations they should indicate the considerations that determined the study size (e.g., a fixed available sample, as in the first example above). If the observational study was stopped early when statistical significance was achieved, readers should be told. Do not bother readers with *post hoc* justifications for study size or retrospective power calculations [77]. From the point of view of the reader, confidence intervals indicate the statistical precision that was ultimately obtained. It should be realized that confidence intervals reflect statistical uncertainty only, and not all uncertainty that may be present in a study (see item 20).

11. Quantitative variables: Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.

Example

"Patients with a Glasgow Coma Scale less than 8 are considered to be seriously injured. A GCS of 9 or more indicates less serious brain injury. We examined the association of GCS in these two categories with the occurrence of death within 12 months from injury" [80].

Explanation

Investigators make choices regarding how to collect and analyse quantitative data about exposures, effect modifiers and confounders. For example, they may group a continuous exposure variable to create a new categorical variable (see Box 4). Grouping choices may have important consequences for later analyses [81,82]. We advise that authors explain why and how they grouped quantitative data, including the number of categories, the cut-points, and category mean or median values. Whenever data are reported in tabular form, the counts of cases, controls, persons at risk, person-time at risk, etc. should be given for each category. Tables should not consist solely of effect-measure estimates or results of model fitting.

Investigators might model an exposure as continuous in order to retain all the information. In making this choice, one needs to consider the nature of the relationship of the exposure to the outcome. As it may be wrong to assume a linear relation automatically, possible departures from linearity should be investigated. Authors could mention alternative models they explored during analyses (e.g., using log transformation, quadratic terms or spline functions). Several methods exist for fitting a non-linear relation between the exposure and outcome [82–84]. Also, it may be informative to present both continuous and grouped analyses for a quantitative exposure of prime interest.

In a recent survey, two thirds of epidemiological publications studied quantitative exposure variables [4]. In 42 of 50 articles (84%) exposures were grouped into several ordered categories, but often without any stated rationale for the choices made. Fifteen articles used linear associations to model continuous exposure but only two reported checking for linearity. In another survey, of the psychological literature, dichotomization was justified in only 22 of 110 articles (20%) [85].

Box 4. Grouping

There are several reasons why continuous data may be grouped [86]. When collecting data it may be better to use an ordinal variable than to seek an artificially precise continuous measure for an exposure based on recall over several years. Categories may also be helpful for presentation, for example to present all variables in a similar style, or to show a dose-response relationship.

Grouping may also be done to simplify the analysis, for example to avoid an assumption of linearity. However, grouping loses information and may reduce statistical power [87] especially when dichotomization is used [82,85,88]. If a continuous confounder is grouped, residual confounding may occur, whereby some of the variable's confounding effect remains unadjusted for (see Box 5) [62,89]. Increasing the number of categories can diminish power loss and residual confounding, and is especially appropriate in large studies. Small studies may use few groups because of limited numbers.

Investigators may choose cut-points for groupings based on commonly used values that are relevant for diagnosis or prognosis, for practicality, or on statistical grounds. They may choose equal numbers of individuals in each group using quantiles [90]. On the other hand, one may gain more insight into the association with the outcome by choosing more extreme outer groups and having the middle group(s) larger than the outer groups [91]. In case-control studies, deriving a distribution from the control group is preferred since it is intended to reflect the source population. Readers should be informed if cut-points are selected post hoc from several alternatives. In particular, if the cut-points were chosen to minimise a P value the true strength of an association will be exaggerated [81].

When analysing grouped variables, it is important to recognise their underlying continuous nature. For instance, a possible trend in risk across ordered groups can be investigated. A common approach is to model the rank of the groups as a continuous variable. Such linearity across group scores will approximate an actual linear relation if groups are equally spaced (e.g., 10 year age groups) but not otherwise. Il'yasova et al [92]. recommend publication of both the categorical and the continuous estimates of effect, with their standard errors, in order to facilitate meta-analysis, as well as providing intrinsically valuable information on dose-response. One analysis may inform the other and neither is assumption-free. Authors often ignore the ordering and consider the estimates (and P values) separately for each category compared to the reference category. This may be useful for description, but may fail to detect a real trend in risk across groups. If a trend is observed, a confidence interval for a slope might indicate the strength of the observation.

12. Statistical methods:

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

Example

"The adjusted relative risk was calculated using the Mantel-Haenszel 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." [93].

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) [94]. Often, the study design determines which type of regression analysis is chosen. For instance, Cox proportional hazard regression is commonly used in cohort studies [95]. whereas logistic regression is often the method of choice in case-control studies [96,97]. Analysts should fully describe specific procedures for variable selection and not only present results from the final model [98,99]. 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. Non-standard 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" [100].

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 [101]. Another study found that among 67 articles in which statistical analyses were adjusted for confounders, it was mostly unclear how confounders were chosen [4].

12 (b). 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^-, a^-b^+, a^+b^-, and a^+b^+) 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:

$$\mathrm{RERI} = \mathrm{RR}(a^+b^+) - \mathrm{RR}(a^-b^+) - \mathrm{RR}(a^+b^-) - 1$$

where $\mathrm{RR}(a^+b^+)$ denotes RR among those exposed to both factors where $\mathrm{RR}(a^-b^-)$ is used as reference category (RR = 1.0). Ninety-five percent CIs were calculated as proposed by Hosmer and Lemeshow. RERI of 0 means no interaction" [103].

Explanation

As discussed in detail under item 17, many debate the use and value of analyses restricted to subgroups of the study population [4,104]. 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

Box 5. Confounding

Confounding literally means confusion of effects. A study might seem to show either an association or no association between an exposure and the risk of a disease. In reality, the seeming association or lack of association is due to another factor that determines the occurrence of the disease but that is also associated with the exposure. The other factor is called the confounding factor or confounder. Confounding thus gives a wrong assessment of the potential 'causal' association of an exposure. For example, if women who approach middle age and develop elevated blood pressure are less often prescribed oral contraceptives, a simple comparison of the frequency of cardiovascular disease between those who use contraceptives and those who do not, might give the wrong impression that contraceptives protect against heart disease.

Investigators should think beforehand about potential confounding factors. This will inform the study design and allow proper data collection by identifying the confounders for which detailed information should be sought. Restriction or matching may be used. In the example above, the study might be restricted to women who do not have the confounder, elevated blood pressure. Matching on blood pressure might also be possible, though not necessarily desirable (see Box 2). In the analysis phase, investigators may use stratification or multivariable analysis to reduce the effect of confounders. Stratification consists of dividing the data in strata for the confounder (e.g., strata of blood pressure), assessing estimates of association within each stratum, and calculating the combined estimate of association as a weighted average over all strata. Multivariable analysis achieves the same result but permits one to take more variables into account simultaneously. It is more flexible but may involve additional assumptions about the mathematical form of the relationship between exposure and disease.

Taking confounders into account is crucial in observational studies, but readers should not assume that analyses adjusted for confounders establish the 'causal part' of an association. Results may still be distorted by residual confounding (the confounding that remains after unsuccessful attempts to control for it [102]), random sampling error, selection bias and information bias (see Box 3).

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 [105]. Whatever view is taken, this should be clearly presented to the reader, as is done in the example above [103]. 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" [106].

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 [5,107]. 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).

12 (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 μL and 123 cells per μL), 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 μL and 123 cells per μL). (...) 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.

Box 6. Missing data: problems and possible solutions

A common approach to dealing with missing data is to restrict analyses to individuals with complete data on all variables required for a particular analysis. Although such 'complete-case' analyses are unbiased in many circumstances, they can be biased and are always inefficient [108]. Bias arises if individuals with missing data are not typical of the whole sample. Inefficiency arises because of the reduced sample size for analysis.

Using the last observation carried forward for repeated measures can distort trends over time if persons who experience a foreshadowing of the outcome selectively drop out [109]. Inserting a missing category indicator for a confounder may increase residual confounding [107]. Imputation, in which each missing value is replaced with an assumed or estimated value, may lead to attenuation or exaggeration of the association of interest, and without the use of sophisticated methods described below may produce standard errors that are too small.

Rubin developed a typology of missing data problems, based on a model for the probability of an observation being missing [108,110]. Data are described as missing completely at random (MCAR) if the probability that a particular observation is missing does not depend on the value of any observable variable(s). Data are missing at random (MAR) if, given the observed data, the probability that observations are missing is independent of the actual values of the missing data. For example, suppose younger children are more prone to missing spirometry measurements, but that the probability of missing is unrelated to the true unobserved lung function, after accounting for age. Then the missing lung function measurement would be MAR in models including age. Data are missing not at random (MNAR) if the probability of missing still depends on the missing value even after taking the available data into account. When data are MNAR valid inferences require explicit assumptions about the mechanisms that led to missing data.

Methods to deal with data missing at random (MAR) fall into three broad classes [108,111]: likelihood-based approaches [112], weighted estimation [113] and multiple imputation [111,114]. Of these three approaches, multiple imputation is the most commonly used and flexible, particularly when multiple variables have missing values [115]. Results using any of these approaches should be compared with those from complete case analyses, and important differences discussed. The plausibility of assumptions made in missing data analyses is generally unverifiable. In particular it is impossible to prove that data are MAR, rather than MNAR. Such analyses are therefore best viewed in the spirit of sensitivity analysis (see items 12e and 17).

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.

12 (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" [117].

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 [50]. 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.

12 (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" [118].

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 [119]. 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.

12 (e). Describe any sensitivity analyses.

Example

"Because we had a relatively higher proportion of 'missing' dead patients with insufficient data (38/148=25.7%) as compared to live patients (15/437=3.4%) (...), 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" [120].

Explanation

Sensitivity analyses are useful to investigate whether or not the main results are consistent with those obtained with alternative analysis strategies or assumptions [121]. Issues that may be examined include the criteria for inclusion in analyses, the definitions of exposures or outcomes [122], which confounding variables merit adjustment, the handling of missing data [120,123], 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 [124–126].

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 [127]. 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 [128]. 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.

RESULTS

The Results section should give a factual account of what was found, from the recruitment of study participants, the description of the study population to the main results and ancillary analyses. It should be free of interpretations and discursive text reflecting the authors' views and opinions.

13. Participants:

13 (a). Report the numbers of individuals at each stage of the study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.

Example

"Of the 105 freestanding bars and taverns sampled, 13 establishments were no longer in business and 9 were located in restaurants, leaving 83 eligible businesses. In 22 cases, the owner could not be reached by telephone despite 6 or more attempts. The owners of 36 bars declined study participation. (...) The 25 participating bars and taverns employed 124 bartenders, with 67 bartenders working at least 1 weekly daytime shift. Fifty-four of the daytime bartenders (81%) completed baseline interviews and spirometry; 53 of these subjects (98%) completed follow-up" [129].

Explanation

Detailed information on the process of recruiting study participants is important for several reasons. Those included in a study often differ in relevant ways from the target population to which results are applied. This may result in estimates of prevalence or incidence that do not reflect the experience of the target population. For example, people who agreed to participate in a postal survey of sexual behaviour attended church less often, had less conservative sexual attitudes and earlier age at first sexual intercourse, and were more likely to smoke cigarettes and drink alcohol than people who refused [130]. These differences suggest that postal surveys may overestimate sexual liberalism and activity in the population. Such response bias (see Box 3) can distort exposure-disease associations if associations differ between those eligible for the study and those included in the study. As another example, the association between young maternal age and leukaemia in offspring, which has been observed in some case-control studies [131,132], was explained by differential participation of young women in case and control groups. Young women with healthy children were less likely to participate than those with unhealthy children [133]. Although low participation does not necessarily compromise the validity of a study, transparent information on participation and reasons for non-participation is essential. Also, as there are no universally agreed definitions for participation, response or follow-up rates, readers need to understand how authors calculated such proportions [134].

Ideally, investigators should give an account of the numbers of individuals considered at each stage of recruiting study participants, from the choice of a target population to the inclusion of participants' data in the analysis. Depending on the type of study, this may include the number of individuals considered to be potentially eligible, the number assessed for eligibility, the number found to be eligible, the number included in the study, the number examined, the number followed up and the number included in the analysis. Information on different sampling units may be required, if sampling of study participants is carried out in two or more stages as in the example above (multistage sampling). In case-control studies, we advise that authors describe the flow of participants separately for case and control groups [135]. Controls can sometimes be selected from several sources, including, for example, hospitalised patients and community dwellers. In this case, we recommend a separate account of the numbers of participants for each type of control group. Olson and colleagues proposed useful reporting guidelines for controls recruited through random-digit dialling and other methods [136].

A recent survey of epidemiological studies published in 10 general epidemiology, public health and medical journals found that some information regarding participation was provided in 47 of 107 case-control studies (59%), 49 of 154 cohort studies (32%), and 51 of 86 cross-sectional studies (59%) [137]. Incomplete or absent reporting of participation and non-participation in epidemiological studies was also documented in two other surveys of the literature [4,5]. Finally, there is evidence that participation in epidemiological studies may have declined in recent decades [137,138], which underscores the need for transparent reporting [139].

13 (b). Give reasons for non-participation at each stage.

Example

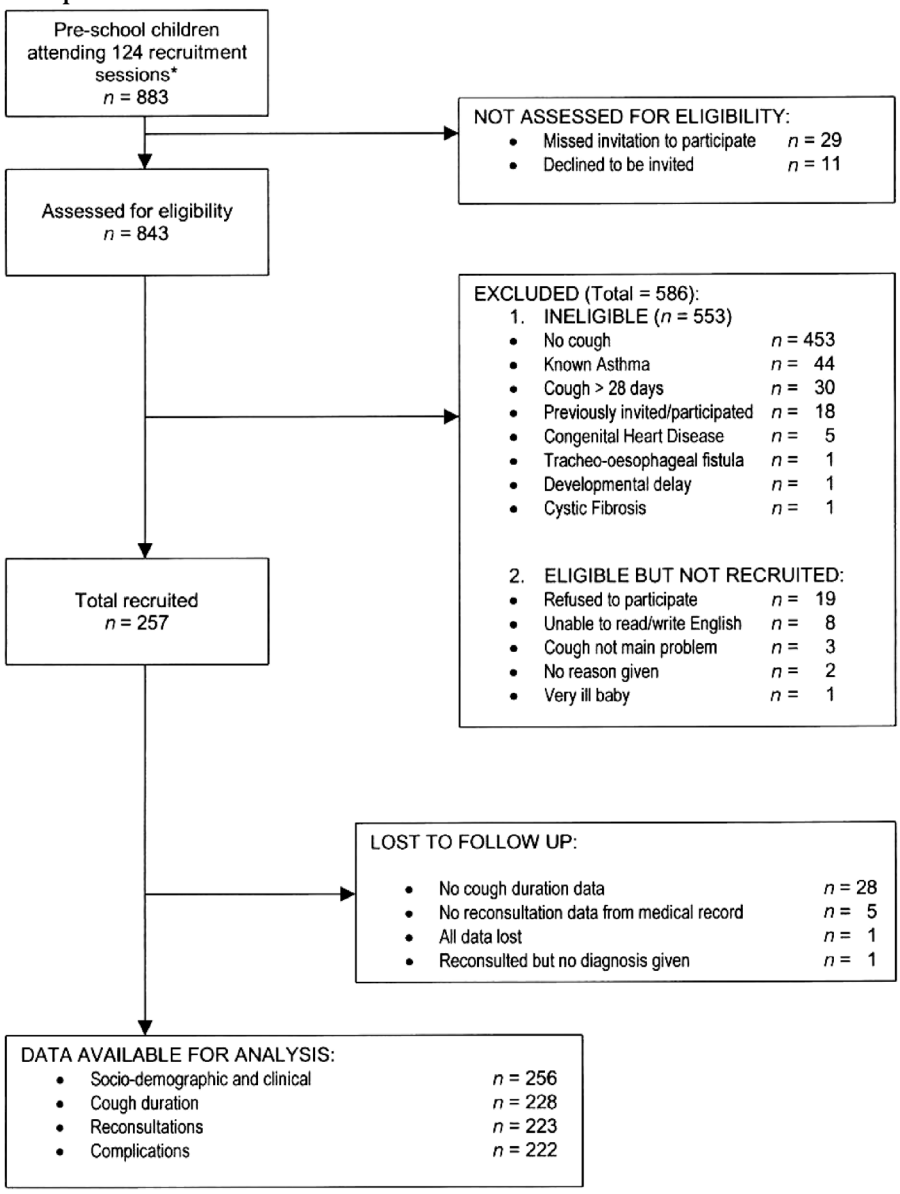
"The main reasons for non-participation were the participant was too ill or had died before interview (cases 30%, controls $< 1\%$), nonresponse (cases 2%, controls 21%), refusal (cases 10%, controls 29%), and other reasons (refusal by consultant or general practitioner, non-English speaking, mental impairment) (cases 7%, controls 5%)" [140].

Explanation

Explaining the reasons why people no longer participated in a study or why they were excluded from statistical analyses helps readers judge whether the study population was representative of the target population and whether bias was possibly introduced. For example, in a cross-sectional health survey, non-participation due to reasons unlikely to be related to health status (for example, the letter of invitation was not delivered because of an incorrect address) will affect the precision of estimates but will probably not introduce bias. Conversely, if many individuals opt out of the survey because of illness, or perceived good health, results may underestimate or overestimate the prevalence of ill health in the population.

13 (c). Consider use of a flow diagram.

Example



*Denominator data missing from one session at which at least 3 attended with cough, 2 recruited
Flow diagram from Hay et al. [141].

Explanation

An informative and well-structured flow diagram can readily and transparently convey information that might otherwise require a lengthy description [142], as in the example above. The diagram may usefully include the main results, such as the number of events for the primary outcome. While we recommend the use of a flow diagram, particularly for complex observational studies, we do not propose a specific format for the diagram.

14. Descriptive data:

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

Example

Table. Characteristics of the Study Base at Enrolment, Castellana G (Italy), 1985–1986

	HCV-Negative *n* = 1458	HCV-Positive *n* = 511	Unknown *n* = 513
Sex (%)			
Male	936 (64%)	296 (58%)	197 (39%)
Female	522 (36%)	215 (42%)	306 (61%)
Mean age at enrolment (SD)	45.7 (10.5)	52.0 (9.7)	52.5 (9.8)
Daily alcohol intake (%)			
None	250 (17%)	129 (25%)	119 (24%)
Moderate[a]	853 (59%)	272 (53%)	293 (58%)
Excessive[b]	355 (24%)	110 (22%)	91 (18%)

HCV, Hepatitis C virus.
[a]Males <60 g ethanol/day, females <30 g ethanol/day.
[b]Males >60 g ethanol/day, females >30 g ethanol/day.
Table adapted from Osella et al. [143].

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 [144,145].

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 [146]. 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 [121,147].

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

Example

Table. Symptom End Points Used in Survival Analysis

	Cough	Short of Breath	Sleeplessness
Symptom resolved	201 (79%)	138 (54%)	171 (67%)
Censored	27 (10%)	21 (8%)	24 (9%)
Never symptomatic	0	46 (18%)	11 (4%)
Data missing	28 (11%)	51 (20%)	50 (20%)
Total	256 (100%)	256 (100%)	256 (100%)

Table adapted from Hay et al. [141].

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) [148]. We advise authors to use their tables and figures to enumerate amounts of missing data.

14 (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" [149].

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 [37].

15. Outcome data: Cohort study: Report numbers of outcome events or summary measures over time.

Example

Table. Rates of HIV-1 Seroconversion by Selected Sociodemographic Variables: 1990–1993

Variable	Person-Years	No. Seroconverted	Rate/1000 Person-Years (95% CI)
Calendar year			
1990	2197.5	18	8.2 (4.4–12.0)
1991	3210.7	22	6.9 (4.0–9.7)
1992	3162.6	18	5.7 (3.1–8.3)
1993	2912.9	26	8.9 (5.5–12.4)
1994	1104.5	5	4.5 (0.6–8.5)
Tribe			
Bagandan	8433.1	48	5.7 (4.1–7.3)
Other Ugandan	578.4	9	15.6 (5.4–25.7)
Rwandese	2318.6	16	6.9 (3.5–10.3)
Other tribe	866.0	12	13.9 (6.0–21.7)
Religion			
Muslim	3313.5	9	2.7 (0.9–4.5)
Other	8882.7	76	8.6 (6.6–10.5)

CI, confidence interval.
Table adapted from Kengeya-Kayondo et al. [150].

Case-control study: Report numbers in each exposure category, or summary measures of exposure.

Example

Table. Exposure among Liver Cirrhosis Cases and Controls

	Cases ($n = 40$)	Controls ($n = 139$)
Vinyl chloride monomer (cumulative exposure: ppm × years)		
<160	7 (18%)	38 (27%)
160–500	7 (18%)	40 (29%)
500-2500	9 (23%)	37 (27%)
>2500	17 (43%)	24 (17%)
Alcohol consumption (g/day)		
<30	1 (3%)	82 (59%)
30–60	7 (18%)	46 (33%)
>60	32 (80%)	11 (8%)
HBsAG/HCV		
Negative	33 (83%)	136 (98%)
Positive	7 (18%)	3 (2%)

HBsAG, hepatitis B surface antigen; HCV, hepatitis C virus.
Table adapted from Mastrangelo et al. [151].

Cross-sectional study: Report numbers of outcome events or summary measures.

Example

Table. Prevalence of Current Asthma and Diagnosed Hay Fever by Average *Alternaria alternata* Antigen Level in the Household

Categorized *Alternaria* Level*	Current Asthma		Diagnosed Hay Fever	
	N	Prevalence† (95% CI)	*N*	Prevalence† (95% CI)
1st tertile	40	4.8 (3.3–6.9)	93	16.4 (13.0–20.5)
2nd tertile	61	7.5 (5.2–10.6)	122	17.1 (12.8–22.5)
3rd tertile	73	8.7 (6.7–11.3)	93	15.2 (12.1–18.9)

*1st tertile < 3.90 μg/g; 2nd tertile 3.90–6.27 μg/g; 3rd tertile ≥ 6.28 μg/g.
†Percentage (95% CI) weighted for the multistage sampling design of the National Survey of Lead and Allergens in Housing.
Table adapted from Salo et al. [152].

Explanation

Before addressing the possible association between exposures (risk factors) and outcomes, authors should report relevant descriptive data. It may be possible and meaningful to present measures of association in the same table that presents the descriptive data (see item 14a). In a cohort study with events as outcomes, report the numbers of events for each outcome of interest. Consider reporting the event rate per person-year of follow-up. If the risk of an event changes over follow-up time, present the numbers and rates of events in appropriate intervals of follow-up or as a Kaplan-Meier life table or plot. It might be preferable to show plots as cumulative incidence that go up from 0% rather than down from 100%, especially if the event rate is lower than, say, 30% [153]. Consider presenting such information separately for participants in different exposure categories of interest. If a cohort study is investigating other time-related outcomes (e.g., quantitative disease markers such as blood pressure), present appropriate summary measures (e.g., means and standard deviations) over time, perhaps in a table or figure.

For cross-sectional studies, we recommend presenting the same type of information on prevalent outcome events or summary measures. For case-control studies, the focus will be on reporting exposures separately for cases and controls as frequencies or quantitative summaries [154]. For all designs, it may be helpful also to tabulate continuous outcomes or exposures in categories, even if the data are not analysed as such.

16. Main results:

16 (a). Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.

Example 1

"We initially considered the following variables as potential confounders by Mantel-Haenszel stratified analysis: (...) The variables we included in the final logistic regression models were those (...) that produced a 10% change in the odds ratio after the Mantel-Haenszel adjustment" [155].

Example 2

Table. Relative Rates of Rehospitalisation by Treatment in Patients in Community Care after First Hospitalisation due to Schizophrenia and Schizoaffective Disorder

Treatment	No. of Relapses	Person-Years	Crude Relative Rate (95% CI)	Adjusted Relative Rate (95% CI)	Fully Adjusted Relative Rate (95% CI)
Perphenazine	53	187	0.41 (0.29 to 0.59)	0.45 (0.32 to 0.65)	0.32 (0.22 to 0.49)
Olanzapine	329	822	0.59 (0.45 to 0.75)	0.55 (0.43 to 0.72)	0.54 (0.41 to 0.71)
Clozapine	336	804	0.61 (0.47 to 0.79)	0.53 (0.41 to 0.69)	0.64 (0.48 to 0.85)
Chlorprothixene	79	146	0.79 (0.58 to 1.09)	0.83 (0.61 to 1.15)	0.64 (0.45 to 0.91)
Thioridazine	115	201	0.84 (0.63 to 1.12)	0.82 (0.61 to 1.10)	0.70 (0.51 to 0.96)
Perphenazine	155	327	0.69 (0.58 to 0.82)	0.78 (0.59 to 1.03)	0.85 (0.63 to 1.13)
Risperidone	343	651	0.77 (0.60 to 0.99)	0.80 (0.62 to 1.03)	0.89 (0.69 to 1.16)
Haloperidol	73	107	1.00	1.00	1.00
Chlorpromazine	82	127	0.94 (0.69 to 1.29)	0.97 (0.71 to 1.33)	1.06 (0.76 to 1.47)
Levomepromazine	52	63	1.21 (0.84 to 1.73)	0.82 (0.58 to 1.18)	1.09 (0.76 to 1.57)
No antipsychotic drugs	2248	3362	0.98 (0.77 to 1.23)	1.01 (0.80 to 1.27)	1.16 (0.91 to 1.47)

Adjusted for sex, calendar year, age at onset of follow-up, number of previous relapses, duration of first hospitalisation, and length of follow-up (adjusted column) and additionally for a score of the propensity to start a treatment other than haloperidol (fully adjusted column).
Table adapted from Tiihonen et al. [156].

Explanation

In many situations, authors may present the results of unadjusted or minimally adjusted analyses and those from fully adjusted analyses. We advise giving the unadjusted

analyses together with the main data, for example the number of cases and controls that were exposed or not. This allows the reader to understand the data behind the measures of association (see also item 15). For adjusted analyses, report the number of persons in the analysis, as this number may differ because of missing values in covariates (see also item 12c). Estimates should be given with confidence intervals.

Readers can compare unadjusted measures of association with those adjusted for potential confounders and judge by how much, and in what direction, they changed. Readers may think that 'adjusted' results equal the causal part of the measure of association, but adjusted results are not necessarily free of random sampling error, selection bias, information bias, or residual confounding (see Box 5). Thus, great care should be exercised when interpreting adjusted results, as the validity of results often depends crucially on complete knowledge of important confounders, their precise measurement, and appropriate specification in the statistical model (see also item 20) [157,158].

Authors should explain all potential confounders considered, and the criteria for excluding or including variables in statistical models. Decisions about excluding or including variables should be guided by knowledge, or explicit assumptions, on causal relations. Inappropriate decisions may introduce bias, for example by including variables that are in the causal pathway between exposure and disease (unless the aim is to asses how much of the effect is carried by the intermediary variable). If the decision to include a variable in the model was based on the change in the estimate, it is important to report what change was considered sufficiently important to justify its inclusion. If a 'backward deletion' or 'forward inclusion' strategy was used to select confounders, explain that process and give the significance level for rejecting the null hypothesis of no confounding. Of note, we and others do not advise selecting confounders based solely on statistical significance testing [147,159,160].

Recent studies of the quality of reporting of epidemiological studies found that confidence intervals were reported in most articles [4]. However, few authors explained their choice of confounding variables [4,5].

16 (b). Report category boundaries when continuous variables were categorised.

Example

Table. Polychlorinated Biphenyls in Cord Serum

Quartile	Range (ng/g)	Number
1	0.07–0.24	180
2	0.24–0.38	181
3	0.38–0.60	181
4	0.61–18.14	180

Table adapted from Sagiv et al. [161].

Explanation

Categorizing continuous data has several important implications for analysis (see Box 4) and also affects the presentation of results. In tables, outcomes should be given for each exposure category, for example as counts of persons at risk, person-time at risk, if relevant separately for each group (e.g., cases and controls). Details of the categories used may aid comparison of studies and meta-analysis. If data were grouped using conventional cut-points, such as body mass index thresholds [162], group boundaries (i.e., range of values) can be derived easily, except for the highest and lowest categories. If quantile-derived categories are used, the category boundaries cannot be inferred from the data. As a minimum, authors should report the category boundaries; it is helpful also to report the range of the data and the mean or median values within categories.

16 (c). If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.

Example

"10 years' use of HRT [hormone replacement therapy] is estimated to result in five (95% CI 3–7) additional breast cancers per 1000 users of oestrogen-only preparations and 19 (15–23) additional cancers per 1000 users of oestrogen-progestagen combinations" [163].

Explanation

The results from studies examining the association between an exposure and a disease are commonly reported in relative terms, as ratios of risks, rates or odds (see Box 8). Relative measures capture the strength of the association between an exposure and disease. If the relative risk is a long way from 1 it is less likely that the association is due to confounding [164,165]. Relative effects or associations tend to be more consistent across studies and populations than absolute measures, but what often tends to be the case may be irrelevant in a particular instance. For example, similar relative risks were obtained for the classic cardiovascular risk factors for men living in Northern Ireland, France, the USA and Germany, despite the fact that the underlying risk of coronary heart disease varies substantially between these countries [166,167]. In contrast, in a study of hypertension as a risk factor for cardiovascular disease mortality, the data were more compatible with a constant rate difference than with a constant rate ratio [168].

Widely used statistical models, including logistic [169] and proportional hazards (Cox) regression [170] are based on ratio measures. In these models, only departures from constancy of ratio effect measures are easily discerned. Nevertheless, measures which assess departures from additivity of risk differences, such as the Relative Excess Risk from Interaction (RERI, see item 12b and Box 8), can be estimated in models based on ratio measures.

In many circumstances, the absolute risk associated with an exposure is of greater interest than the relative risk. For example, if the focus is on adverse effects of a drug, one will want to know the number of additional cases per unit time of use (e.g., days, weeks, or years). The example gives the additional number of breast cancer cases per 1000 women who used hormone-replacement therapy for 10 years [163]. Measures such as the attributable risk or population attributable fraction may be useful to gauge how much disease can be prevented if the exposure is eliminated. They should preferably be presented together with a measure of statistical uncertainty (e.g., confidence intervals as in the

Box 7. Measures of association, effect and impact

Observational studies may be solely done to describe the magnitude and distribution of a health problem in the population. They may examine the number of people who have a disease at a particular time (prevalence), or that develop a disease over a defined period (incidence). The incidence may be expressed as the proportion of people developing the disease (cumulative incidence) or as a rate per person-time of follow-up (incidence rate). Specific terms are used to describe different incidences; amongst others, mortality rate, birth rate, attack rate, or case fatality rate. Similarly, terms like point prevalence and period, annual or lifetime prevalence are used to describe different types of prevalence [30].

Other observational studies address cause-effect relationships. Their focus is the comparison of the risk, rate or prevalence of the event of interest between those exposed and those not exposed to the risk factor under investigation. These studies often estimate a 'relative risk', which may stand for risk ratios (ratios of cumulative incidences) as well as rate ratios (ratios of incidence rates). In case-control studies only a fraction of the source population (the controls) are included. Results are expressed as the ratio of the odds of exposure among cases and controls. This odds ratio provides an estimate of the risk or rate ratio depending on the sampling of cases and controls (see also Box 1) [175,176]. The prevalence ratio or prevalence odds ratio from cross-sectional studies may be useful in some situations [177].

Expressing results both in relative and absolute terms may often be helpful. For example, in a study of male British doctors the incidence rate of death from lung cancer over 50 years of follow-up was 249 per 100,000 per year among smokers, compared to 17 per 100,000 per year among non-smokers: a rate ratio of 14.6 (249/17) [178]. For coronary heart disease (CHD), the corresponding rates were 1001 and 619 per 100,000 per year, for a rate ratio of 1.61 (1001/619). The effect of smoking on death appears much stronger for lung cancer than for CHD. The picture changes when we consider the absolute effects of smoking. The difference in incidence rates was 232 per 100,000 per year (249 − 17) for lung cancer and 382 for CHD (1001 − 619). Therefore, among doctors who smoked, smoking was more likely to cause death from CHD than from lung cancer.

How much of the disease burden in a population could be prevented by eliminating an exposure? Global estimates have been published for smoking: according to one study 91% of all lung cancers, 40% of CHD and 33% of all deaths among men in 2000 were attributed to smoking [179]. The population attributable fraction is generally defined as the proportion of cases caused by a particular exposure, but several concepts (and no unified terminology) exist, and incorrect approaches to adjust for other factors are sometimes used [172,180]. What are the implications for reporting? The relative measures emphasise the strength of an association, and are most useful in etiologic research. If a causal relationship with an exposure is documented and associations are interpreted as *effects*, estimates of relative risk may be translated into suitable measures of absolute risk in order to gauge the possible impact of public health policies (see item 16c) [181]. However, authors should be aware of the strong assumptions made in this context [171]. Care is needed in deciding which concept and method is appropriate for a particular situation.

example). Authors should be aware of the strong assumptions made in this context, including a causal relationship between a risk factor and disease (also see Box 7) [171]. Because of the semantic ambiguity and complexities involved, authors should report in detail what methods were used to calculate attributable risks, ideally giving the formulae used [172].

A recent survey of abstracts of 222 articles published in leading medical journals found that in 62% of abstracts of randomised trials including a ratio measure absolute risks were given, but only in 21% of abstracts of cohort studies [173]. A free text search of Medline 1966 to 1997 showed that 619 items mentioned attributable risks in the title or abstract, compared to 18,955 using relative risk or odds ratio, for a ratio of 1 to 31 [174].

17. Other analyses: Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses.

Example 1

Table. Analysis of Oral Contraceptive Use, Presence of Factor V Leiden Allele, and Risk for Venous Thromboembolism

Factor V Leiden	Oral Contraceptives	No. of Patients	No. of Controls	Odds Ratio
Yes	Yes	25	2	34.7
Yes	No	10	4	6.9
No	Yes	84	63	3.7
No	No	36	100	1 (Reference)

Table modified from Vandenbroucke et al. [182] by Botto et al. [183].

Example 2

Table. Sensitivity of the Rate Ratio for Cardiovascular Outcome to an Unmeasured Confounder

Prevalence of Unmeasured Binary Confounder in the Exposed Group, %	Prevalence of Unmeasured Binary Confounder in the Comparator Group, %	Unmeasured Binary Confounder Rate Ratio	High Exposure Rate Ratio (95% CI)*
90	10	1.5	1.20 (1.01–1.42)
90	50	1.5	1.43 (1.22–1.67)
50	10	1.5	1.39 (1.18–1.63)
90	10	2	0.96 (0.81–1.13)
90	50	2	1.27 (1.11–1.45)
50	10	2	1.21 (1.03–1.42)
90	50	3	1.18 (1.01–1.38)
50	10	3	0.99 (0.85–1.16)
90	50	5	1.08 (0.85–1.26)

CI, confidence interval.
*Adjusted for age, sex, cardiovascular drug use, and unmeasured binary confounder.
Table adapted from Wei et al. [184].

Explanation

In addition to the main analysis other analyses are often done in observational studies. They may address specific subgroups, the potential interaction between risk factors, the calculation of attributable risks, or use alternative definitions of study variables in sensitivity analyses.

There is debate about the dangers associated with subgroup analyses, and multiplicity of analyses in general [4,104]. In our opinion, there is too great a tendency to look for evidence of subgroup-specific associations, or effect-measure modification, when overall results appear to suggest little or no effect. On the other hand, there is value in exploring whether an overall association appears consistent across several, preferably pre-specified subgroups especially when a study is large enough to have sufficient data in each subgroup. A second area of debate is about interesting subgroups that arose during the data analysis. They might be important findings, but might also arise by chance. Some argue that it is neither possible nor necessary to inform the reader about all subgroup analyses done as future analyses of other data will tell to what extent the early exciting findings stand the test of time [9]. We advise authors to report which

analyses were planned, and which were not (see also items 4, 12b and 20). This will allow readers to judge the implications of multiplicity, taking into account the study's position on the continuum from discovery to verification or refutation.

A third area of debate is how joint effects and interactions between risk factors should be evaluated: on additive or multiplicative scales, or should the scale be determined by the statistical model that fits best (see also item 12b and Box 8)? A sensible approach is to report the separate effect of each exposure as well as the joint effect—if possible in a table, as in the first example above [183], or in the study by Martinelli et al. [185]. Such a table gives the reader sufficient information to evaluate additive as well as multiplicative interaction (how these calculations are done is shown in Box 8). Confidence intervals for separate and joint effects may help the reader to judge the strength of the data. In addition, confidence intervals around measures of interaction, such as the Relative Excess Risk from Interaction (RERI) relate to tests of interaction or homogeneity tests. One recurrent problem is that authors use comparisons of P-values across subgroups, which lead to erroneous claims about an effect modifier. For instance, a statistically significant association in one category (e.g., men), but not in the other (e.g., women) does not in itself provide evidence of effect modification. Similarly, the confidence intervals for each point estimate are sometimes inappropriately used to infer that there is no interaction when intervals overlap. A more valid inference is achieved by directly evaluating whether the magnitude of an association differs across subgroups.

Sensitivity analyses are helpful to investigate the influence of choices made in the statistical analysis, or to investigate the robustness of the findings to missing data or possible biases (see also item 12b). Judgement is needed regarding the level of reporting of such analyses. If many sensitivity analyses were performed, it may be impractical to present detailed findings for them all. It may sometimes be sufficient to report that sensitivity analyses were carried out and that they were consistent with the main results presented. Detailed presentation is more appropriate if the issue investigated is of major concern, or if effect estimates vary considerably [59,186].

Pocock and colleagues found that 43 out of 73 articles reporting observational studies contained subgroup analyses. The majority claimed differences across groups but only eight articles reported a formal evaluation of interaction (see item 12b) [4].

Box 8. Interaction (effect modification): the analysis of joint effects

Interaction exists when the association of an exposure with the risk of disease differs in the presence of another exposure. One problem in evaluating and reporting interactions is that the effect of an exposure can be measured in two ways: as a relative risk (or rate ratio) or as a risk difference (or rate difference). The use of the relative risk leads to a multiplicative model, while the use of the risk difference corresponds to an additive model [187,188]. A distinction is sometimes made between 'statistical interaction' which can be a departure from either a multiplicative or additive model, and 'biologic interaction' which is measured by departure from an additive model [189]. However, neither additive nor multiplicative models point to a particular biologic mechanism.

Regardless of the model choice, the main objective is to understand how the joint effect of two exposures differs from their separate effects (in the absence of the other exposure). The Human Genomic Epidemiology Network (HuGENet) proposed a lay-out for transparent presentation of separate and joint effects that permits evaluation of different types of interaction [183]. Data from the study on oral contraceptives and factor V Leiden mutation [182] were used to explain the proposal, and this example is also used in item 17. Oral contraceptives and factor V Leiden mutation each increase the risk of venous thrombosis; their separate and joint effects can be calculated from the 2 by 4 table (see example 1 for item 17) where the odds ratio of 1 denotes the baseline of women without Factor V Leiden who do not use oral contraceptives.

A difficulty is that some study designs, such as case-control studies, and several statistical models, such as logistic or Cox regression models, estimate relative risks (or rate ratios) and intrinsically lead to multiplicative modelling. In these instances, relative risks can be translated to an additive scale. In example 1 of item 17, the separate odds ratios are 3.7 and 6.9; the joint odds ratio is 34.7. When these data are analysed under a multiplicative model, a joint odds ratio of 25.7 is expected (3.7×6.9). The observed joint effect of 34.7 is 1.4 times greater than expected on a multiplicative scale (34.7/25.7). This quantity (1.4) is the odds ratio of the multiplicative interaction. It would be equal to the antilog of the estimated interaction coefficient from a logistic regression model. Under an additive model the joint odds ratio is expected to be 9.6 ($3.7 + 6.9 - 1$). The observed joint effect departs strongly from additivity: the difference is 25.1 (34.7 – 9.6). When odds ratios are interpreted as relative risks (or rate ratios), the latter quantity (25.1) is the Relative Excess Risk from Interaction (RERI) [190]. This can be understood more easily when imagining that the reference value (equivalent to OR=1) represents a baseline incidence of venous thrombosis of, say, 1/10 000 women-years, which then increases in the presence of separate and joint exposures.

DISCUSSION

The discussion section addresses the central issues of validity and meaning of the study [191]. Surveys have found that discussion sections are often dominated by incomplete or biased assessments of the study's results and their implications, and rhetoric supporting the authors' findings [192,193]. Structuring the discussion may help authors avoid unwarranted speculation and over-interpretation of results while guiding readers through the text [194,195]. For example, *Annals of Internal Medicine* [196] recommends that authors structure the discussion section by presenting the following: (1) a brief synopsis of the key findings; (2) consideration of possible mechanisms and explanations; (3) comparison with relevant findings from other published studies; (4) limitations of the study; and (5) a brief section that summarizes the implications of the work for practice and research. Others have made similar suggestions [191,194]. The section on research recommendations and the section on limitations of the study should be closely linked to each other. Investigators should suggest ways in which subsequent research can improve on their studies rather than blandly stating 'more research is needed' [197,198]. We recommend that authors structure their discussion sections, perhaps also using suitable subheadings.

18. Key results: Summarise key results with reference to study objectives.

Example

"We hypothesized that ethnic minority status would be associated with higher levels of cardiovascular disease (CVD) risk factors, but that the associations would be explained substantially by socioeconomic status (SES). Our hypothesis was not confirmed. After adjustment for age and SES, highly significant differences in body mass index, blood pressure, diabetes, and physical inactivity remained between white women and both black and Mexican American women. In addition, we found large differences in CVD risk factors by SES, a finding that illustrates the high-risk status of both ethnic minority women as well as white women with low SES" [199].

Explanation

It is good practice to begin the discussion with a short summary of the main findings of the study. The short summary reminds readers of the main findings and may help them assess whether the subsequent interpretation and implications offered by the authors are supported by the findings.

19. Limitations: Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.

Example

"Since the prevalence of counseling increases with increasing levels of obesity, our estimates may overestimate the true prevalence. Telephone surveys also may overestimate the true prevalence of counseling. Although persons without telephones have similar levels of overweight as persons with telephones, persons without telephones tend to be less educated, a factor associated with lower levels of counseling in our study. Also, of concern is the potential bias caused by those who refused to participate as well as those who refused to respond to questions about weight. Furthermore, because data were collected cross-sectionally, we cannot infer that counseling preceded a patient's attempt to lose weight" [200].

Explanation

The identification and discussion of the limitations of a study are an essential part of scientific reporting. It is important not only to identify the sources of bias and confounding that could have affected results, but also to discuss the relative importance of different biases, including the likely direction and magnitude of any potential bias (see also item 9 and Box 3).

Authors should also discuss any imprecision of the results. Imprecision may arise in connection with several aspects of a study, including the study size (item 10) and the measurement of exposures, confounders and outcomes (item 8). The inability to precisely measure true values of an exposure tends to result in bias towards unity: the less precisely a risk factor is measured, the greater the bias. This effect has been described as 'attenuation' [201,202], or more recently as 'regression dilution bias' [203]. However, when correlated risk factors are measured with different degrees of imprecision, the adjusted relative risk associated with them can be biased towards or away from unity [204–206].

When discussing limitations, authors may compare the study being presented with other studies in the literature in terms of validity, generalisability and precision. In this approach, each study can be viewed as contribution to the literature, not as a stand-alone basis for inference and action [207]. Surprisingly, the discussion of important limitations of a study is sometimes omitted from published reports. A survey of authors who had published original research articles in *The Lancet* found that important weaknesses of the study were reported by the investigators in the survey questionnaires, but not in the published article [192].

20. Interpretation: Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.

Example

"Any explanation for an association between death from myocardial infarction and use of second generation oral contraceptives must be conjectural. There is no published evidence to suggest a direct biologic mechanism, and there are no other epidemiologic studies with relevant results. (. . .) The increase in absolute risk is very small and probably applies predominantly to smokers. Due to the lack of corroborative evidence, and because the analysis is based on relatively small numbers, more evidence on the subject is needed. We would not recommend any change in prescribing practice on the strength of these results" [120].

Explanation

The heart of the discussion section is the interpretation of a study's results. Over-interpretation is common and human: even when we try hard to give an objective assessment, reviewers often rightly point out that we went too far in some respects. When interpreting results, authors should consider the nature of the study on the discovery to verification continuum and potential sources of bias, including loss to follow-up and non-participation (see also items 9, 12 and 19). Due consideration should be given to confounding (item 16a), the results of relevant sensitivity analyses, and to the issue of multiplicity and subgroup analyses (item 17). Authors should also consider residual confounding due to unmeasured variables or imprecise measurement of confounders. For example, socioeconomic status (SES) is associated with many health outcomes and often differs between groups being compared. Variables used to measure SES (income, education, or occupation) are surrogates for other undefined and unmeasured exposures, and the true confounder will by definition be measured with error [208]. Authors should address the real range of uncertainty in estimates, which is larger than the statistical uncertainty reflected in confidence intervals. The latter do not take into account other uncertainties that arise from a study's design, implementation, and methods of measurement [209].

To guide thinking and conclusions about causality, some may find criteria proposed by Bradford Hill in 1965 helpful [164]. How strong is the association with the exposure? Did it precede the onset of disease? Is the association consistently observed in different studies and settings? Is there supporting evidence from experimental studies, including laboratory and animal studies? How specific is the exposure's putative effect, and is there a dose-response relationship? Is the association biologically plausible? These criteria should not, however, be applied mechanically. For example, some have argued that relative risks below 2 or 3 should be ignored [210,211]. This is a reversal of the point by Cornfield et al. about the strength of large relative risks (see item 12b) [127]. Although a causal effect is more likely with a relative risk of 9, it does not follow that one below 3 is necessarily spurious. For instance, the small increase in the risk of childhood leukaemia after intrauterine irradiation is credible because it concerns an adverse effect of a medical procedure for which no alternative explanations are obvious [212]. Moreover, the carcinogenic effects of radiation are well established. The doubling in the risk of ovarian cancer associated with eating 2 to 4 eggs per week is not immediately credible, since dietary

habits are associated with a large number of lifestyle factors as well as SES [213]. In contrast, the credibility of much debated epidemiologic findings of a difference in thrombosis risk between different types of oral contraceptives was greatly enhanced by the differences in coagulation found in a randomised cross-over trial [214]. A discussion of the existing external evidence, from different types of studies, should always be included, but may be particularly important for studies reporting small increases in risk. Further, authors should put their results in context with similar studies and explain how the new study affects the existing body of evidence, ideally by referring to a systematic review.

21. Generalisability: Discuss the generalisability (external validity) of the study results.

Example

"How applicable are our estimates to other HIV-1-infected patients? This is an important question because the accuracy of prognostic models tends to be lower when applied to data other than those used to develop them. We addressed this issue by penalising model complexity, and by choosing models that generalised best to cohorts omitted from the estimation procedure. Our database included patients from many countries from Europe and North America, who were treated in different settings. The range of patients was broad: men and women, from teenagers to elderly people were included, and the major exposure categories were well represented. The severity of immunodeficiency at baseline ranged from not measureable to very severe, and viral load from undetectable to extremely high" [215].

Explanation

Generalisability, also called external validity or applicability, is the extent to which the results of a study can be applied to other circumstances [216]. There is no external validity *per se*; the term is meaningful only with regard to clearly specified conditions [217]. Can results be applied to an individual, groups or populations that differ from those enrolled in the study with regard to age, sex, ethnicity, severity of disease, and co-morbid conditions? Are the nature and level of exposures comparable, and the definitions of outcomes relevant to another setting or population? Are data that were collected in longitudinal studies many years ago still relevant today? Are results from health services research in one country applicable to health systems in other countries?

The question of whether the results of a study have external validity is often a matter of judgment that depends on the study setting, the characteristics of the participants, the exposures examined, and the outcomes assessed. Thus, it is crucial that authors provide readers with adequate information about the setting and locations, eligibility criteria, the exposures and how they were measured, the definition of outcomes, and the period of recruitment and follow-up. The degree of non-participation and the proportion of unexposed participants in whom the outcome develops are also relevant. Knowledge of the absolute risk and prevalence of the exposure, which will often vary across populations, are helpful when applying results to other settings and populations (see Box 7).

OTHER INFORMATION

22. Funding: 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.

Explanation

Some journals require authors to disclose the presence or absence of financial and other conflicts of interest [100,218]. Several investigations show strong associations between the source of funding and the conclusions of research articles [219–222]. The conclusions in randomised trials recommended the experimental drug as the drug of choice much more often (odds ratio 5.3) if the trial was funded by for-profit organisations, even after adjustment for the effect size [223]. Other studies document the influence of the tobacco and telecommunication industries on the research they funded [224–227]. There are also examples of undue influence when the sponsor is governmental or a non-profit organisation.

Authors or funders may have conflicts of interest that influence any of the following: the design of the study [228]; choice of exposures [228,229], outcomes [230], statistical methods [231], and selective publication of outcomes [230] and studies [232]. Consequently, the role of the funders should be described in detail: in what part of the study they took direct responsibility (e.g., design, data collection, analysis, drafting of manuscript, decision to publish) [100]. Other sources of undue influence include employers (e.g., university administrators for academic researchers and government supervisors, especially political appointees, for government researchers), advisory committees, litigants, and special interest groups.

Concluding Remarks

The STROBE Statement aims to provide helpful recommendations for reporting observational studies in epidemiology. Good reporting reveals the strengths and weaknesses of a study and facilitates sound interpretation and application of study results. The STROBE Statement may also aid in planning observational studies, and guide peer reviewers and editors in their evaluation of manuscripts.

We wrote this explanatory article to discuss the importance of transparent and complete reporting of observational studies, to explain the rationale behind the different items included in the checklist, and to give examples from published articles of what we consider good reporting. We hope that the material presented here will assist authors and editors in using STROBE.

We stress that STROBE and other recommendations on the reporting of research [13,233,234] should be seen as evolving documents that require continual assessment, refinement, and, if necessary, change [235,236]. For example, the CONSORT Statement for the reporting of parallel-group randomized trials was first developed in the mid 1990s [237]. Since then members of the group have met regularly to review the need to revise the recommendations; a revised version appeared in 2001 [233] and a further version is in development. Similarly, the principles presented in this article and the STROBE checklist are open to change as new evidence and critical comments accumulate. The STROBE Web site (http://www.strobe-statement.org/) provides a forum for discussion and suggestions for improvements of the checklist, this explanatory document and information about the good reporting of epidemiological studies.

Several journals ask authors to follow the STROBE Statement in their instructions to authors (see http://www.strobe-statement.org/ for current list). We invite other journals to adopt the STROBE Statement and contact us through our Web site to let us know. The journals publishing the STROBE recommendations provide open access. The STROBE Statement is therefore widely accessible to the biomedical community.

Acknowledgments

We are grateful to Gerd Antes, Kay Dickersin, Shah Ebrahim and Richard Lilford for supporting the STROBE Initiative. We are grateful to the following institutions that have hosted working meetings: Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland; Department of Social Medicine, University of Bristol, UK; London School of Hygiene & Tropical Medicine, London, UK; Nordic Cochrane Centre, Copenhagen, Denmark; and Centre for Statistics in Medicine, Oxford, UK. We are grateful to four anonymous reviewers who provided helpful comments on a previous draft of this paper.

Contributors to the STROBE Initiative. The following persons have contributed to the content and elaboration of the STROBE Statement: Douglas G. Altman, Maria Blettner, Paolo Boffetta, Hermann Brenner, Geneviève Chêne, Cyrus Cooper, George Davey-Smith, Erik von Elm, Matthias Egger, France Gagnon, Peter C. Gøtzsche, Philip Greenland, Sander Greenland, Claire Infante-Rivard, John Ioannidis, Astrid James, Giselle Jones, Bruno Ledergerber, Julian Little, Margaret May, David Moher, Hooman Momen, Alfredo Morabia, Hal Morgenstern, Cynthia D. Mulrow, Fred Paccaud, Stuart J. Pocock, Charles Poole, Martin Röösli, Dietrich Rothenbacher, Kenneth Rothman, Caroline Sabin, Willi Sauerbrei, Lale Say, James J. Schlesselman, Jonathan Sterne, Holly Syddall, Jan P. Vandenbroucke, Ian White, Susan Wieland, Hywel Williams, Guang Yong Zou.

Author contributions. All authors contributed to the writing of the paper. JPV, EvE, DGA, PCG, SJP, and ME wrote the first draft of different sections of the paper. EvE takes care of most of the practical coordination of STROBE. ME initiated STROBE and, together with EvE, organised the first workshop.

References

1. Glasziou P, Vandenbroucke JP, Chalmers I (2004) Assessing the quality of research. BMJ 328: 39–41.
2. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G (2001) Distribution of study designs in four major US journals of obstetrics and gynecology. Gynecol Obstet Invest 51: 8–11.
3. Scales CD Jr., Norris RD, Peterson BL, Preminger GM, Dahm P (2005) Clinical research and statistical methods in the urology literature. J Urol 174: 1374–1379.
4. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, et al. (2004) Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ 329: 883.
5. Tooth L, Ware R, Bain C, Purdie DM, Dobson A (2005) Quality of reporting of observational longitudinal research. Am J Epidemiol 161: 280–288.
6. von Elm E, Altman DG, Egger M, Gøtzsche PC, Pocock SJ, et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. PLoS Med e296. doi:10.1371/journal.pmed.0040296.
7. Mihailovic A, Bell CM, Urbach DR (2005) Users' guide to the surgical literature. Case-control studies in surgical journals. Can J Surg 48: 148–151.
8. Rushton L (2000) Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. Occup Environ Med 57: 1–9.
9. Rothman KJ (1990) No adjustments are needed for multiple comparisons. Epidemiology 1: 43–46.
10. Moonesinghe R, Khoury MJ, Janssens AC (2007) Most published research findings are false—but a little replication goes a long way. PLoS Med 4: e28. doi:10.1371/journal.pmed.0040028
11. Jenicek M (1999) Clinical Case Reporting. Evidence-Based Medicine. Oxford: Butterworth-Heinemann. 117 pp.
12. Vandenbroucke JP (2001) In defense of case reports and case series. Ann Intern Med 134: 330–334.
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 138: 40–44.
14. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, et al. (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 93: 387–391.
15. Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, et al. (2006) A road map for efficient and reliable human genome epidemiology. Nat Genet 38: 3–5.
16. Rodrigues L, Kirkwood BR (1990) Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 19: 205–213.
17. Rothman KJ, Greenland S (1998) Case-Control Studies. In: Rothman KJ, Greenland S editors. Modern epidemiology. 2nd ed: Lippincott Raven. pp. 93–114.
18. Forand SP (2004) Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. Environ Health 3: 7.
19. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials. N Engl J Med 342: 1878–1886.
20. Gøtzsche PC, Harden A Searching for non-randomised studies. Draft chapter 3. Cochrane Non-Randomised Studies Methods Group, 26 July 2002. Available: http://www.cochrane.dk/nrsmg. Accessed 10 September 2007.
21. Lohse N, Hansen AB, Pedersen G, Kronborg G, Gerstoft J, et al. (2007) Survival of persons with and without HIV infection in Denmark, 1995–2005. Ann Intern Med 146: 87–95.
22. American Journal of Epidemiology (2007) Information for authors. Available: http://www.oxfordjournals.org/aje/for__authors/index.html. Accessed 10 September 2007.
23. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ (1990) More informative abstracts revisited. Ann Intern Med 113: 69–76.
24. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, et al. (1994) Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. CMAJ 150: 1611–1615.
25. Hartley J, Sydes M (1996) Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. J Inform Sci 22: 27–37.
26. Viner RM, Cole TJ (2005) Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. BMJ 330: 1354.
27. McCauley J, Kern DE, Kolodner K, Dill L, Schroeder AF, et al. (1995) The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. Ann Intern Med 123: 737–746.
28. McEvoy SP, Stevenson MR, McCartt AT, Woodward M, Haworth C, et al. (2005) Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. BMJ 331: 428.
29. Vandenbroucke JP (1991) Prospective or retrospective: what's in a name? BMJ 302: 249–250.
30. Last JM (2000) A Dictionary of Epidemiology. New York: Oxford University Press.
31. Miettinen OS (1985) Theoretical Epidemiology: principles of occurrence research in medicine. New York: Wiley. pp. 64–66.
32. Rothman KJ, Greenland S (1998) Types of Epidemiologic Studies. In: Rothman KJ, Greenland S editors. Modern epidemiology. 2nd ed: Lippincott Raven. pp. 74–75.
33. MacMahon B, Trichopoulos D (1996) Epidemiology, principles and methods. 2nd ed. Boston: Little, Brown. 81 p.
34. Lilienfeld AM (1976) Foundations of Epidemiology. New York: Oxford University Press.
35. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, et al. (1995) Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. N Engl J Med 332: 912–917.
36. Goodman KJ, O'Rourke K, Day RS, Wang C, Nurgalieva Z, et al. (2005) Dynamics of Helicobacter pylori infection in a US-Mexico cohort during the first two years of life. Int J Epidemiol 34: 1348–1355.
37. Altman DG, De Stavola BL, Love SB, Stepniewska KA (1995) Review of survival analyses published in cancer journals. Br J Cancer 72: 511–518.
38. Cerhan JR, Wallace RB, Folsom AR, Potter JD, Munger RG, et al. (1993) Transfusion history and cancer risk in older women. Ann Intern Med 119: 8–15.
39. Beane Freeman LE, Dennis LK, Lynch CF, Thorne PS, Just CL (2004) Toenail arsenic content and cutaneous melanoma in Iowa. Am J Epidemiol 160: 679–687.
40. Canto JG, Allison JJ, Kiefe CI, Fincher C, Farmer R, et al. (2000) Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. N Engl J Med 342: 1094–1100.
41. Metzkor-Cotter E, Kletter Y, Avidor B, Varon M, Golan Y, et al. (2003) Long-term serological analysis and clinical follow-up of patients with cat scratch disease. Clin Infect Dis 37: 1149–1154.
42. Johnson ES (1990) Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. J Occup Med 32: 250–254.

43. Berkson J (1946) Limitations of the application of fourfold table analysis to hospital data. Biom Bull 2: 53.
44. Feinstein AR, Walter SD, Horwitz RI (1986) An analysis of Berkson's bias in case-control studies. J Chronic Dis 39: 495–504.
45. Jick H, Vessey MP (1978) Case-control studies in the evaluation of drug-induced illness. Am J Epidemiol 107: 1–7.
46. Hackam DG, Mamdani M, Li P, Redelmeier DA (2006) Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. Lancet 367: 413–418.
47. Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, et al. (2004) MMR vaccination and pervasive developmental disorders: a case-control study. Lancet 364: 963–969.
48. Costanza MC (1995) Matching. Prev Med 24: 425–433.
49. Sturmer T, Brenner H (2002) Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. Am J Epidemiol 155: 593–602.
50. Rothman KJ, Greenland S (1998) Matching. In: Rothman KJ, Greenland S, editors. 2nd ed. Modern epidemiology. Lippincott Raven. pp. 147–161.
51. Szklo MF, Nieto J (2000). Epidemiology, Beyond the Basics. Sudbury (MA): Jones and Bartlett. pp. 40–51.
52. Cole P, MacMahon B (1971) Attributable risk percent in case-control studies. Br J Prev Soc Med 25: 242–244.
53. Gissler M, Hemminki E (1996) The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. Eur J Obstet Gynecol Reprod Biol 69: 73–75.
54. Gefeller O, Pfahlberg A, Brenner H, Windeler J (1998) An empirical investigation on matching in published case-control studies. Eur J Epidemiol 14: 321–325.
55. Artama M, Ritvanen A, Gissler M, Isojarvi J, Auvinen A (2006) Congenital structural anomalies in offspring of women with epilepsy—a population-based cohort study in Finland. Int J Epidemiol 35: 280–287.
56. Ebrahim S (2004) Cohorts, infants and children. Int J Epidemiol 33: 1165–1166.
57. Walker M, Whincup PH, Shaper AG (2004) The British Regional Heart Study 1975–2004. Int J Epidemiol 33: 1185–1192.
58. Wieland S, Dickersin K (2005) Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. J Clin Epidemiol 58: 560–567.
59. Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L (2005) Ambient particulate matter and health effects: publication bias in studies of short-term associations. Epidemiology 16: 155–163.
60. Winkelmayer WC, Stampfer MJ, Willett WC, Curhan GC (2005) Habitual caffeine intake and the risk of hypertension in women. JAMA 294: 2330–2335.
61. Lukanova A, Soderberg S, Kaaks R, Jellum E, Stattin P (2006) Serum adiponectin is not associated with risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 15: 401–402.
62. Becher H (1992) The concept of residual confounding in regression models and some applications. Stat Med 11: 1747–1758.
63. Brenner H, Blettner M (1997) Controlling for continuous confounders in epidemiologic research. Epidemiology 8: 429–434.
64. Phillips MR, Yang G, Zhang Y, Wang L, Ji H, et al. (2002) Risk factors for suicide in China: a national case-control psychological autopsy study. Lancet 360: 1728–1736.
65. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, et al. (2006) Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. Ophthalmology 113: 1081–1086.
66. Craig SL, Feinstein AR (1999) Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. Am J Clin Oncol 22: 51–56.
67. Rogler LH, Mroczek DK, Fellows M, Loftus ST (2001) The neglect of response bias in mental health research. J Nerv Ment Dis 189: 182–187.
68. Murphy EA (1976) The logic of medicine. Baltimore: Johns Hopkins University Press.
69. Sackett DL (1979) Bias in analytic research. J Chronic Dis 32: 51–63.
70. Johannes CB, Crawford SL, McKinlay JB (1997) Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. Am J Epidemiol 146: 429–438.
71. Bloemenkamp KW, Rosendaal FR, Buller HR, Helmerhorst FM, Colly LP, et al. (1999) Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. Arch Intern Med 159: 65–70.
72. Feinstein AR (1985) Clinical epidemiology: the architecture of clinical research. Philadelphia: W.B. Saunders.
73. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA (2003) Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. Am J Trop Med Hyg 68: 519–526.
74. Anoop S, Saravanan B, Joseph A, Cherian A, Jacob KS (2004) Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. Arch Dis Child 89: 325–329.
75. Carlin JB, Doyle LW (2002) Sample size. J Paediatr Child Health 38: 300–304.
76. Rigby AS, Vail A (1998) Statistical methods in epidemiology. II: A commonsense approach to sample size estimation. Disabil Rehabil 20: 405–410.
77. Schulz KF, Grimes DA (2005) Sample size calculations in randomised trials: mandatory and mystical. Lancet 365: 1348–1353.
78. Drescher K, Timm J, Jockel KH (1990) The design of case-control studies: the effect of confounding on sample size requirements. Stat Med 9: 765–776.
79. Devine OJ, Smith JM (1998) Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. Stat Med 17: 1375–1389.
80. Linn S, Levi L, Grunau PD, Zaidise I, Zarka S (2007) Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. Ann Epidemiol 17: 142–147.
81. Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994) Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 86: 829–835.
82. Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 25: 127–141.
83. Greenland S (1995) Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. Epidemiology 6: 450–454.
84. Royston P, Ambler G, Sauerbrei W (1999) The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 28: 964–974.
85. MacCallum RC, Zhang S, Preacher KJ, Rucker DD (2002) On the practice of dichotomization of quantitative variables. Psychol Methods 7: 19–40.
86. Altman DG (2005) Categorizing continuous variables. In: Armitage P, Colton T editors. Encyclopedia of biostatistics. 2nd ed. Chichester: John Wiley. pp. 708–711.
87. Cohen J (1983) The cost of dichotomization. Applied Psychological Measurement 7: 249–253.
88. Zhao LP, Kolonel LN (1992) Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. Am J Epidemiol 136: 464–474.
89. Cochran WG (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics 24: 295–313.
90. Clayton D, Hills M (1993) Models for dose-response (Chapter 25). Statistical Models in Epidemiology. Oxford: Oxford University Press. pp. 249–260.
91. Cox DR (1957) Note on grouping. J Am Stat Assoc 52: 543–547.
92. Il'yasova D, Hertz-Picciotto I, Peters U, Berlin JA, Poole C (2005) Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. Cancer Causes Control 16: 383–388.
93. Berglund A, Alfredsson L, Cassidy JD, Jensen I, Nygren A (2000) The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. J Clin Epidemiol 53: 1089–1094.
94. Slama R, Werwatz A (2005) Controlling for continuous confounding factors: non- and semiparametric approaches. Rev Epidemiol Sante Publique 53 Spec No 2: 2S65–80.
95. Greenland S (1998) Introduction to regression modelling (Chapter 21). In: Rothman KJ, Greenland S editors. Modern epidemiology. 2nd ed: Lippincott Raven. pp. 401–432.
96. Thompson WD (1994) Statistical analysis of case-control studies. Epidemiol Rev 16: 33–50.
97. Schlesselman JJ (1982) Logistic regression for case-control studies (Chapter 8.2). Case-control studies Design, conduct, analysis. New York, Oxford: Oxford University Press. pp. 235–241.
98. Clayton D, Hills M (1993) Choice and interpretation of models (Chapter 27). Statistical Models in Epidemiology. Oxford: Oxford University Press. pp. 271–281.
99. Altman DG, Gore SM, Gardner MJ, Pocock SJ (1983) Statistical guidelines for contributors to medical journals. Br Med J 286: 1489–1493.
100. International Committee of Medical Journal Editors (1997) Uniform requirements for manuscripts submitted to biomedical journals [Electronic version updated February 2006, available at http://www.icmje.org/] N Engl J Med 336: 309–315.
101. Mullner M, Matthews H, Altman DG (2002) Reporting on statistical methods to adjust for confounding: a cross-sectional survey. Ann Intern Med 136: 122–126.
102. Olsen J, Basso O (1999) Re: Residual confounding. Am J Epidemiol 149: 290.
103. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarod K, et al. (2006) Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? Am J Kidney Dis 47: 396–405.
104. Gotzsche PC (2006) Believability of relative risks and odds ratios in abstracts: cross sectional study. BMJ 333: 231–234.
105. Szklo MF, Nieto J (2000) Communicating Results of Epidemiologic Studies (Chapter 9). Epidemiology, Beyond the Basics. Sudbury (MA): Jones and Bartlett. pp. 408–430.
106. Chandola T, Brunner E, Marmot M (2006) Chronic stress at work and the metabolic syndrome: prospective study. BMJ 332: 521–525.
107. Vach W, Blettner M (1991) Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. Am J Epidemiol 134: 895–907.

108. Little RJ, Rubin DB (2002) A taxonomy of missing-data methods (Chapter 1.4.) In: Statistical Analysis with Missing Data. New York: Wiley. pp. 19–23.
109. Ware JH (2003) Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 348: 2136–2137.
110. Rubin DB (1976) Inference and missing data. Biometrika 63: 581–592.
111. Schafer JL (1997) Analysis of Incomplete Multivariate Data. London: Chapman & Hall.
112. Lipsitz SR, Ibrahim JG, Chen MH, Peterson H (1999) Non-ignorable missing covariates in generalized linear models. Stat Med 18: 2435–2448.
113. Rotnitzky A, Robins J (1997) Analysis of semi-parametric regression models with non-ignorable non-response. Stat Med 16: 81–102.
114. Rubin DB (1987) Multiple Imputation for Nonresponse in Surveys. New York: John Wiley.
115. Barnard J, Meng XL (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. Stat Methods Med Res 8: 17–36.
116. Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boulle A, et al. (2006) Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. Lancet 367: 817–824.
117. Purandare N, Burns A, Daly KJ, Hardicre J, Morris J, et al. (2006) Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. BMJ 332: 1119–1124.
118. Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J (2001) Hypertension in South African adults: results from the Demographic and Health Survey, 1998. J Hypertens 19: 1717–1725.
119. Lohr SL (1999) Design Effects (Chapter 7.5). Sampling: Design and Analysis. Pacific Grove (CA): Duxbury Press.
120. Dunn NR, Arscott A, Thorogood M (2001) The relationship between use of oral contraceptives and myocardial infarction in young women with fatal outcome, compared to those who survive: results from the MICA case-control study. Contraception 63: 65–69.
121. Rothman KJ, Greenland S (1998) Basic Methods for Sensitivity Analysis and External Adjustment. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed: Lippincott Raven. pp. 343–357.
122. Custer B, Longstreth WT Jr., Phillips LE, Koepsell TD, Van Belle G (2006) Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. BMC Cancer 6: 152.
123. Wakefield MA, Chaloupka FJ, Kaufman NJ, Orleans CT, Barker DC, et al. (2000) Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. BMJ 321: 333–337.
124. Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. J Am Stat Assoc 98: 47–54.
125. Lash TL, Fink AK (2003) Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology 14: 451–458.
126. Phillips CV (2003) Quantifying and reporting uncertainty from systematic errors. Epidemiology 14: 459–466.
127. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, et al. (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst 22: 173–203.
128. Langholz B (2001) Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? Bioelectromagnetics 5: S19–31.
129. Eisner MD, Smith AK, Blanc PD (1998) Bartenders' respiratory health after establishment of smoke-free bars and taverns. JAMA 280: 1909–1914.
130. Dunne MP, Martin NG, Bailey JM, Heath AC, Bucholz KK, et al. (1997) Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. Int J Epidemiol 26: 844–854.
131. Schuz J, Kaatsch P, Kaletsch U, Meinert R, Michaelis J (1999) Association of childhood cancer with factors related to pregnancy and birth. Int J Epidemiol 28: 631–639.
132. Cnattingius S, Zack M, Ekbom A, Gunnarskog J, Linet M, et al. (1995) Prenatal and neonatal risk factors for childhood myeloid leukemia. Cancer Epidemiol Biomarkers Prev 4: 441–445.
133. Schuz J (2003) Non-response bias as a likely cause of the association between young maternal age at the time of delivery and the risk of cancer in the offspring. Paediatr Perinat Epidemiol 17: 106–112.
134. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD (1995) Response rates among control subjects in case-control studies. Ann Epidemiol 5: 245–249.
135. Schulz KF, Grimes DA (2002) Case-control studies: research in reverse. Lancet 359: 431–434.
136. Olson SH, Voigt LF, Begg CB, Weiss NS (2002) Reporting participation in case-control studies. Epidemiology 13: 123–126.
137. Morton LM, Cahill J, Hartge P (2006) Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol 163: 197–203.
138. Olson SH (2001) Reported participation in case-control studies: changes over time. Am J Epidemiol 154: 574–581.
139. Sandler DP (2002) On revealing what we'd rather hide: the problem of describing study participation. Epidemiology 13: 117.
140. Hepworth SJ, Schoemaker MJ, Muir KR, Swerdlow AJ, van Tongeren MJ, et al. (2006) Mobile phone use and risk of glioma in adults: case-control study. BMJ 332: 883–887.
141. Hay AD, Wilson A, Fahey T, Peters TJ (2003) The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. Fam Pract 20: 696–705.
142. Egger M, Juni P, Bartlett C (2001) Value of flow diagrams in reports of randomized controlled trials. JAMA 285: 1996–1999.
143. Osella AR, Misciagna G, Guerra VM, Chiloiro M, Cuppone R, et al. (2000) Hepatitis C virus (HCV) infection and liver-related mortality: a population-based cohort study in southern Italy. The Association for the Study of Liver Disease in Puglia. Int J Epidemiol 29: 922–927.
144. Dales LG, Ury HK (1978) An improper use of statistical significance testing in studying covariables. Int J Epidemiol 7: 373–375.
145. Maldonado G, Greenland S (1993) Simulation study of confounder-selection strategies. Am J Epidemiol 138: 923–936.
146. Tanis BC, van den Bosch MA, Kemmeren JM, Cats VM, Helmerhorst FM, et al. (2001) Oral contraceptives and the risk of myocardial infarction. N Engl J Med 345: 1787–1793.
147. Rothman KJ, Greenland S (1998) Precision and Validity in Epidemiologic Studies. In: Rothman KJ, Greenland S editors. Modern epidemiology. 2nd ed: Lippincott Raven. pp. 120–125.
148. Clark TG, Altman DG, De Stavola BL (2002) Quantification of the completeness of follow-up. Lancet 359: 1309–1310.
149. Qiu C, Fratiglioni L, Karp A, Winblad B, Bellander T (2004) Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. Epidemiology 15: 687–694.
150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, Ruberantwari A, Wagner HU, et al. (1996) Incidence of HIV-1 infection in adults and socio-demographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. Int J Epidemiol 25: 1077–1082.
151. Mastrangelo G, Fedeli U, Fadda E, Valentini F, Agnesi R, et al. (2004) Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. Environ Health Perspect 112: 1188–1192.
152. Salo PM, Arbes SJ Jr., Sever M, Jaramillo R, Cohn RD, et al. (2006) Exposure to Alternaria alternata in US homes is associated with asthma symptoms. J Allergy Clin Immunol 118: 892–898.
153. Pocock SJ, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 359: 1686–1689.
154. Sasieni P (1992) A note on the presentation of matched case-control data. Stat Med 11: 617–620.
155. Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA (2002) A nested case-control study of residential and personal magnetic field measures and miscarriages. Epidemiology 13: 21–31.
156. Tiihonen J, Walhbeck K, Lonnqvist J, Klaukka T, Ioannidis JP, et al. (2006) Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. BMJ 333: 224.
157. Christenfeld NJ, Sloan RP, Carroll D, Greenland S (2004) Risk factors, confounding, and the illusion of statistical control. Psychosom Med 66: 868–875.
158. Smith GD, Phillips A (1990) Declaring independence: why we should be cautious [erratum: J Epidemiol Community Health 1991 Mar;45(1);88]. J Epidemiol Community Health 44: 257–258.
159. Greenland S, Neutra R (1980) Control of confounding in the assessment of medical technology. Int J Epidemiol 9: 361–367.
160. Robins JM (2001) Data, design, and background knowledge in etiologic inference. Epidemiology 12: 313–320.
161. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA (2007) Organochlorine exposures during pregnancy and infant size at birth. Epidemiology 18: 120–129.
162. World Health Organization (2007) Body Mass Index (BMI). Available: http://www.euro.who.int/nutrition/20030507_1. Accessed 10 September 2007.
163. Beral V (2003) Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet 362: 419–427.
164. Hill AB (1965) The environment and disease: Association or causation? Proc R Soc Med 58: 295–300.
165. Vineis P (2003) Causality in epidemiology. Soz Praventivmed 48: 80–87.
166. Empana JP, Ducimetiere P, Arveiler D, Ferrieres J, Evans A, et al. (2003) Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. Eur Heart J 24: 1903–1911.
167. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, et al. (1999) Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. Lancet 353: 1547–1557.
168. Cambien F, Chretien JM, Ducimetiere P, Guize L, Richard JL (1985) Is the

relationship between blood pressure and cardiovascular risk dependent on body mass index? Am J Epidemiol 122: 434–442.
169. Hosmer DW, Taber S, Lemeshow S (1991) The importance of assessing the fit of logistic regression models: a case study. Am J Public Health 81: 1630–1635.
170. Tibshirani R (1982) A plain man's guide to the proportional hazards model. Clin Invest Med 5: 63–68.
171. Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. Am J Public Health 88: 15–19.
172. Uter W, Pfahlberg A (2001) The application of methods to quantify attributable risk in medical practice. Stat Methods Med Res 10: 231–237.
173. Schwartz LM, Woloshin S, Dvorin EL, Welch HG (2006) Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. BMJ 333: 1248.
174. Nakayama T, Zaman MM, Tanaka H (1998) Reporting of attributable and relative risks, 1966–97. Lancet 351: 1179.
175. Cornfield J (1951) A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst 11: 1269–1275.
176. Pearce N (1993) What does the odds ratio estimate in a case-control study? Int J Epidemiol 22: 1189–1192.
177. Rothman KJ, Greenland S (1998) Measures of Disease Frequency. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. Lippincott Raven. pp. 44–45.
178. Doll R, Hill AB (2004) The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. BMJ 328: 1529–1533; discussion 1533.
179. Ezzati M, Lopez AD (2003) Estimates of global mortality attributable to smoking in 2000. Lancet 362: 847–852.
180. Greenland S (1998) Applications of Stratified Analysis Methods. In: Rothman KJ, Greenland S, editors. Modern epidemiology. 2nd ed. Lippincott Raven. pp. 295–297.
181. Rose G (2001) Sick individuals and sick populations. Int J Epidemiol 30: 427–432; discussion 433–424.
182. Vandenbroucke JP, Koster T, Briet E, Reitsma PH, Bertina RM, et al. (1994) Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. Lancet 344: 1453–1457.
183. Botto LD, Khoury MJ (2001) Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. Am J Epidemiol 153: 1016–1020.
184. Wei L, MacDonald TM, Walker BR (2004) Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. Ann Intern Med 141: 764–770.
185. Martinelli I, Taioli E, Battaglioli T, Podda GM, Passamonti SM, et al. (2003) Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. Arch Intern Med 163: 2771–2774.
186. Kyzas PA, Loizou KT, Ioannidis JP (2005) Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 97: 1043–1055.
187. Rothman KJ, Greenland S, Walker AM (1980) Concepts of interaction. Am J Epidemiol 112: 467–470.
188. Saracci R (1980) Interaction and synergism. Am J Epidemiol 112: 465–466.
189. Rothman KJ (2002) Epidemiology. An introduction. Oxford: Oxford University Press. pp. 168–180.
190. Rothman KJ (1986) Interactions Between Causes. Modern epidemiology. Boston: Little Brown. pp. 311–326.
191. Hess DR (2004) How to write an effective discussion. Respir Care 49: 1238–1241.
192. Horton R (2002) The hidden research paper. JAMA 287: 2775–2778.
193. Horton R (1995) The rhetoric of research. BMJ 310: 985–987.
194. Docherty M, Smith R (1999) The case for structuring the discussion of scientific papers. BMJ 318: 1224–1225.
195. Perneger TV, Hudelson PM (2004) Writing a research article: advice to beginners. Int J Qual Health Care 16: 191–192.
196. Annals of Internal Medicine. Information for authors. Available at: http://www.annals.org/shared/author__info.html. Accessed 10 September 2007.
197. Maldonado G, Poole C (1999) More research is needed. Ann Epidemiol 9: 17–18.
198. Phillips CV (2001) The economics of 'more research is needed'. Int J Epidemiol 30: 771–776.
199. Winkleby MA, Kraemer HC, Ahn DK, Varady AN (1998) Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988–1994. JAMA 280: 356–362.
200. Galuska DA, Will JC, Serdula MK, Ford ES (1999) Are health care professionals advising obese patients to lose weight? JAMA 282: 1576–1578.
201. Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15: 72–101.
202. Fuller WA, Hidiroglou MA (1978) Regression estimates after correcting for attenuation. J Am Stat Assoc 73: 99–104.
203. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, et al. (1990) Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 335: 765–774.
204. Phillips AN, Smith GD (1991) How independent are "independent" effects? Relative risk estimation when correlated exposures are measured imprecisely. J Clin Epidemiol 44: 1223–1231.
205. Phillips AN, Smith GD (1992) Bias in relative odds estimation owing to imprecise measurement of correlated exposures. Stat Med 11: 953–961.
206. Greenland S (1980) The effect of misclassification in the presence of covariates. Am J Epidemiol 112: 564–569.
207. Poole C, Peters U, Il'yasova D, Arab L (2003) Commentary: This study failed? Int J Epidemiol 32: 534–535.
208. Kaufman JS, Cooper RS, McGee DL (1997) Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. Epidemiology 8: 621–628.
209. Greenland S (1990) Randomization, statistics, and causal inference. Epidemiology 1: 421–429.
210. Taubes G (1995) Epidemiology faces its limits. Science 269: 164–169.
211. Temple R (1999) Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. JAMA 281: 841–844.
212. Greenberg RS, Shuster JL Jr. (1985) Epidemiology of cancer in children. Epidemiol Rev 7: 22–48.
213. Kushi LH, Mink PJ, Folsom AR, Anderson KE, Zheng W, et al. (1999) Prospective study of diet and ovarian cancer. Am J Epidemiol 149: 21–31.
214. Kemmeren JM, Algra A, Meijers JC, Tans G, Bouma BN, et al. (2004) Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V Leiden mutation: a randomized trial. Blood 103: 927–933.
215. Egger M, May M, Chene G, Phillips AN, Ledergerber B, et al. (2002) Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet 360: 119–129.
216. Campbell DT (1957) Factors relevant to the validity of experiments in social settings. Psychol Bull 54: 297–312.
217. Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. Ann Intern Med 130: 515–524.
218. Krimsky S, Rothenberg LS (2001) Conflict of interest policies in science and medical journals: editorial practices and author disclosures. Sci Eng Ethics 7: 205–218.
219. Bekelman JE, Li Y, Gross CP (2003) Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 289: 454–465.
220. Davidson RA (1986) Source of funding and outcome of clinical trials. J Gen Intern Med 1: 155–158.
221. Stelfox HT, Chua G, O'Rourke K, Detsky AS (1998) Conflict of interest in the debate over calcium-channel antagonists. N Engl J Med 338: 101–106.
222. Lexchin J, Bero LA, Djulbegovic B, Clark O (2003) Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 326: 1167–1170.
223. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL (2003) Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 290: 921–928.
224. Barnes DE, Bero LA (1998) Why review articles on the health effects of passive smoking reach different conclusions. JAMA 279: 1566–1570.
225. Barnes DE, Bero LA (1996) Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. J Health Polit Policy Law 21: 515–542.
226. Glantz SA, Barnes DE, Bero L, Hanauer P, Slade J (1995) Looking through a keyhole at the tobacco industry. The Brown and Williamson documents. JAMA 274: 219–224.
227. Huss A, Egger M, Hug K, Huwiler-Muntener K, Roosli M (2007) Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. Environ Health Perspect 115: 1–4.
228. Safer DJ (2002) Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. J Nerv Ment Dis 190: 583–592.
229. Aspinall RL, Goodman NW (1995) Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. BMJ 311: 844–846.
230. Chan AW, Hrobjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 291: 2457–2465.
231. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326: 1171–1173.
232. Scherer RW, Langenberg P, von Elm E (2005) Full publication of results initially presented in abstracts. Cochrane Database of Systematic Reviews, Issue 2. Art. No.: MR000005. Available: http://www.cochrane.org/reviews/en/mr000005.html. Accessed 10 September 10, 2007.

233. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357: 1191–1194.
234. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, et al. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283: 2008–2012.
235. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, et al. (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134: 663–694.
236. Moher D (1998) CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. JAMA 279: 1489–1491.
237. Begg C, Cho M, Eastwood S, Horton R, Moher D, et al. (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 276: 637–639.