

観察的疫学研究報告の質改善（STROBE）のための声明：

解説と詳細

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

Vandenvroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ,
Pool C, Schlesselman J, Egger M; STROBE Initiative.
(Epidemiology 2007;18:805-835)

-----日本語翻訳検討グループ-----

監修：

福原俊一（京都大学医学研究科医療疫学分野教授）

山口拓洋（東京大学医学部附属病院 臨床試験データ管理学講座准教授）

山崎新（京都大学医学研究科医療疫学分野准教授）

林野泰明（京都大学医学研究科医療疫学分野講師）

竹上未紗（京都大学医学研究科医療疫学分野助教）

翻訳担当：

横山葉子（京都大学医学研究科医療疫学分野研究員）

川口武彦（京都大学医学研究科医療疫学分野博士後期課程）

杉岡隆（京都大学医学研究科医療疫学分野博士課程）

小野玲（京都大学医学研究科医療疫学分野博士後期課程）

島田利彦（京都大学医学研究科医療疫学分野博士後期課程）

山本洋介（京都大学医学研究科医療疫学分野博士課程）

佐藤恵子（京都大学医学研究科医療疫学分野博士後期課程）

坂井加奈恵（京都大学医学研究科医療疫学分野博士後期課程）

2009年5月28日

目次

抄録	5
STROBE 声明とは	5
観察研究の領域.....	6
このドキュメント (document) の使用方法について.....	7
STROBE のチェックリスト項目	7
タイトル・抄録 (TITLE AND ABSTRACT)	10
1(a)タイトルまたは抄録の中で研究デザインを一般に用いられている用語で明示する。	10
1(b)抄録では研究で行われたことと明らかにされたことについて十分な情報を含み、かつ バランスのよい要約を記載する。	10
緒言 (INTRODUCTION)	11
2.背景 (background) / 論拠 (rationale) : 研究の科学的な背景と論拠を説明する。 11	
3.目的 (objectives) : 事前の仮説を含む形で研究の具体的な目的を明記する。	12
方法 (METHODS)	13
4.研究デザイン (study design) : 研究デザインの重要な要素を論文の早い段階で示す。	13
5.セッティング (setting) : セッティング、実施場所のほか、基準となる日付については、 登録、曝露、追跡、および、データ収集の期間を含めて明記する。	14
6.参加者 (participants) :	15
6(a)コホート研究 : 適格基準 (eligibility criteria)、および、対象者のソースと選定方 法を示す。追跡の方法についても記述する。	15
6(a)ケース・コントロール研究 : 適格基準、対象者のソース、ケースの確定方法、およ び、コントロールの選択方法を示す。ケースとコントロールの選択における論拠を示 す。	15
6(a)横断研究 : 適格基準、および、対象者のソースと選択方法を示す。	16
6(b)コホート研究 : マッチングを行った研究ではマッチングの基準と曝露群と非曝露群 の数を示す。	17
6(b)ケース・コントロール研究 : マッチングを行った研究ではマッチングの基準とケー スあたりのコントロールの数を示す。	17
7.変数 (variables) : すべてのアウトカム、曝露、予測因子 (predictor)、潜在的な交絡 因子 (potential confounder)、潜在的な効果指標の修飾因子 (potential effect modifier) を明確に定義する。該当する場合は、診断方法を示す。	20
8.データソース (data source) / 測定方法 (measurement) : 関連する各因子に対して、 データソース、評価・測定方法の詳細を示す。2つ以上の群がある場合は、測定方法の比 較可能性 (comparability) を明記する。	20

9.バイアス (bias) : バイアスが生じないようにとられた方策があればすべて示す。 ...	21
10.サンプルサイズ(study size) : サンプルサイズがどのように決められたかを説明する。	24
11.量的変数 (quantitative variable) : 量的変数の分析方法を説明する。また量的変数を グループに分けた場合は、どのように分けたかをその理由とともに記載する。	25
12.統計・分析方法 (statistical methods) :	27
12(a)交絡因子の調整に用いた方法を含め、すべての統計学的方法を示す。	27
12(b)サブグループと交互作用 (inter action) の検証に用いたすべての方法を示す。	29
12(c)欠損データ (missing data) をどのように扱ったかを説明する。	30
12(d)コホート研究 : 該当する場合は、脱落例 (loss to follow-up) をどのように扱った かを説明する。	31
12(d)ケース・コントロール研究 : 該当する場合は、ケースとコントロールのマッチン グをどのように行ったかを説明する。	32
12(d)横断研究 : 該当する場合は、サンプリング方法を考慮した分析法について記述す る。	33
12(e)あらゆる感度分析 (sensitivity analysis) の方法を示す。	34
結果 (RESULTS)	35
13.参加者 (participants) :	35
13(a)研究の各段階における人数を示す (例: 潜在的な適格 (potentially eligible) 者数、 適格性が調査された数、適格と確認された数、研究に組み入れられた数、フォローアッ プを完了した数、および、分析された数)。	35
13(b)研究の各段階での研究不参加の理由を記述する。	36
13(c)フローチャートによる記載を考慮する。	37
14.記述的データ (descriptive data) :	38
14(a) 参加者の特徴 (例 : 人口統計学的、臨床的、社会学的特徴) ならびに曝露および 潜在的な交絡因子の情報を示す。	38
14(b)それぞれの変数について、データが欠損した参加者数を記載する。	39
14(c)コホート研究:追跡期間を平均および合計で要約する。	40
15.アウトカム・データ (outcome data) :	40
コホート研究 : 経時的なアウトカム事象の発生数や要約指標 (summary measures) を報告する。	40
ケース・コントロール研究 : 各曝露カテゴリー内の対象者数または要約指標を示す。	41
横断研究 : アウトカム事象の発生数または要約指標を示す。	41
16. おもな結果 (main results) :	42
16(a)調整前 (unadjust) の推定値、該当する場合は交絡因子での調整後の推定値、お	

よび、それらの精度（例：95%信頼区間）を記述する。どの交絡因子がなぜ調整されたかを明確にする。	42
16(b)連続変数がカテゴリー化されているときはカテゴリーの境界値（category boundaries）を報告する。	43
16(c)意味のある（relevant）場合は、相対リスク（relative risk）を意義ある期間の絶対リスク（absolute risk）に換算することを考慮する。	44
17. 他の解析（other analyses）：その他に行われたすべての分析（例：サブグループと交互作用の解析や感度分析）の結果を報告する。	47
考察（DISCUSSION）	50
18. 鍵となる結果（key results）：研究目的に関しての鍵となる結果を要約する。	50
19. 限界（limitations）：潜在的なバイアスまたは精度の問題を考慮して研究の限界を議論する。潜在的なバイアスの方向性と大きさを議論する。	50
20. 解釈（interpretation）：目的、限界、解析の多重性（multiplicity）、同様の研究で得られた結果、および、その他の関連するエビデンスを考慮し、慎重で総合的な結果の解釈を記載する。	52
21. 一般化可能性：研究結果の一般化可能性（外的妥当性[external validity]）を議論する。	53
その他の情報（OTHER INFORMATION）	54
22. 研究の財源（funding）：研究の資金源、本研究における資金提供者の役割を示す。該当する場合には、現在の研究の元となる研究についても同様に示す。	54
結論	54
引用文献	55
日本語訳付録	66

抄録

多くの医学研究は観察研究である。観察研究には報告の質が不十分なものが散見され、研究の強み (strength) や弱み (weakness)、および、一般化可能性 (generalizability) の評価が妨げられている。そこで、方法論の専門家 (methodologist) や雑誌編集者が集い、経験上のエビデンスや理論を考慮した上で、STROBE (Strengthen the Reporting of Observational Studies in Epidemiology) 推奨 (recommendations) を作成した。これは観察研究の報告の質を改善するためのものである。

この STROBE 声明 (statement) は、論文の「タイトル (title)」「抄録 (abstract)」「緒言 (introduction)」「方法 (methods)」「結果 (results)」および「考察 (discussion)」に関連した 22 項目のチェックリストから成っている。うち 18 項目は、コホート研究 (cohort study)、ケース・コントロール研究 (case-control study)、および、横断研究 (cross-sectional study) に共通の項目であり、残りの 4 項目は上の 3 つの研究デザインのそれぞれに特有な項目である。STROBE 声明は、著者が観察研究の報告の質を向上させるために有用であり、また、雑誌の査読者、編集者、および、読者が論文の批判的評価 (critical appraisal) や解釈 (interpretation) を行うことを容易にするものである。

STROBE の説明と詳細について記したこのドキュメント (document) は、STROBE 声明の使用、理解、および、普及を促すために作成されたものである。各チェックリストの項目に対して、その意義と理論的根拠を記し、また出版されている論文の中から 1 つ以上の例を用い、可能な限り、関連した実証的研究 (empirical studies) や方法論に関する文献 (methodological literature) を記した参考資料を載せた。その中には、有用なフローチャート (flow diagrams) の例も含まれている。STROBE 声明、このドキュメント、および、それに関連するウェブサイト (<http://www.strobe-statement.org>) はともに、観察研究の報告を向上させるのに有用であるものとする。 (Epidemiology 2007;18:805-835)

医療を正しく実践していくためには、疾患の原因やその作用様式に関する学問および研究 (etiology) についての知識と、疾患の病態、診断、予後、および、治療方法についての知識が必要である。ランダム化試験 (randomized trial) は治療や他の介入に関する価値のあるエビデンスを提供する。しかし、多くの臨床上・公衆衛生上の知見は観察研究から得られる¹。臨床専門の学術雑誌に出版された論文のうち約 10 分の 9 は、観察研究による研究を記したものである^{2,3}。

STROBE 声明とは

観察研究による報告は、その研究の強み (strength) と弱み (weakness) を評価するには十分詳細かつ明確に記されていないことが多い^{4,5}。そこでわれわれは、観察研究の報告の質を改善するために、報告すべき項目のチェックリストとして、Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) 声明 (statement) を作成した (TABLE 1)。

チェックリストの各項目は、論文の「タイトル (title)」「抄録 (abstract)」「緒言 (introduction)」「方法 (methods)」「結果 (results)」および「考察 (discussion)」に関するものである。この STROBE 声明は最近いくつかの雑誌に掲載されている⁶。この声明の目的は、観察研究で何が計画され、実際に何が行われ、そして、何が発見されたのかに関わる明確な報告を強化することである。この声明で推奨していることは、研究の計画や実施における規定ではないし、また、研究の方法論を説明したり発表の仕方を画一的に提示したりするものではないことを強調しておく。

STROBE は、記述的な観察研究や、曝露と健康に関するアウトカム (outcomes) との関連を検討する研究の報告において、一般的に含めるべき内容について言及している。また STROBE では、観察研究における 3 つの主要な研究デザイン (study design) である、コホート研究 (cohort study)、ケース・コントロール研究 (case-control study)、横断研究 (cross-sectional study) について言及している。これらの研究デザインを示すとき、著者は雑多な用語を使用しており、例えば「フォローアップ研究 (follow-up study)」や「縦断研究 (longitudinal study)」という用語が「コホート研究」の同義語として、また「有病割合調査 (あるいは有病率調査 [prevalence study])」が「横断研究」と同義語として用いられることがある。われわれは、一般的に使用されているという理由だけでこのような用語を選んでいく。残念ながら、往々にしてその使い方は的確でなかったり (incorrectly)⁷、あるいは厳密でなかったり (imprecisely) する⁸。Box 1 に 3 つの研究デザインの特徴をまとめた。

観察研究の領域

観察研究は、疾患の潜在的な原因の最初の手がかりを報告するものから先行研究で報告された関連性の程度を検証するものまで、幅広い目的に対応している。研究のアイデアは、臨床での観察や生物学的視点から生じるものかもしれない。また、事前の計画がなされていない非公式のデータの検討からも、さらに進んだ研究を導くアイデアが得られることがあるかもしれない。診察した何千人もの患者の中から自分の関心を引くたった一人の患者を見つけ出す医師のように、研究者はデータの中に特別な何かを見出すかもしれない。多重性の観点からデータを調整することが可能ではない場合や望ましくない場合があるが⁹、ある観察された事柄が正しいのかあるいは正しくないのかについて、さらなる研究での検討が必要となる場合がよくある¹⁰。すでにあるデータを用いて、原因として考えられる因子についての新たなアイデアを検証することができるであろうし、それにより、そのアイデアを十分に否定あるいは肯定することもできるであろう。また、先行研究における潜在的な問題点を克服するように研究を特別にデザインすることもできるであろう。後者の研究では、既存のデータを分析する場合と比べ、その目的を達成するために新しいデータを収集し、計画を練る必要が生じるであろう。したがって、後者のような研究では、例えば、サブグループを調べる利点やサンプルサイズ (sample size) を事前に設定することの重要性といった、多様な視点が生じる。STROBE は様々な観察研究—探索的な研究から検証的な

研究まで一に対応するものである。また、必要に応じてどのような場合に特別な内容を含めるべきかについても言及するものである。

このドキュメント (document) の使用方法について

このドキュメントは STROBE 声明の中心を成すものである。また、いくつかの雑誌に掲載されたチェックリスト項目を紹介する簡易版 STROBE 論文とリンクしている⁶。われわれの意図は、よい研究報告の仕方を説明することであり、研究をどのように実施するべきかを説明することではない。このドキュメントでは、チェックリストの各項目について詳細な説明を加えている。それぞれの説明の前には、われわれがわかりやすい報告と考える実際の記載例を挙げてある。ただし、これらの記載例を含む研究論文全体としての報告内容が一様に優れているとっているわけではなく、また、後に他の研究においてその結果が確認されたという意味でその研究の信頼性が高いとっているわけでもない。ここで挙げられている記載例は、その研究において特定の項目についての報告が優れているということしか意味しない。また、各項目についての説明や記載例に加え、補足説明のための囲み記事 (Box) も載せた。これらは、いくつかの理論上のポイントの新しい知識を吸収したい読者や、技術的な背景についての詳細を迅速に知りたい読者の要望に応えるものである。これらの補足事項について十分に理解したければ、引用した教科書や方法論についての論文を参照するとよい。

STROBE では、遺伝的連鎖研究 (genetic linkage study)、感染症モデルの研究 (infectious disease modeling)、あるいは、症例報告 (case reports) や症例集積 (case series) についての特別な言及はない^{11,12}。しかし、STROBE の軸となる構成要素の多くを、これらの研究にあてはめることはできるので、このような研究を報告する著者にとっても STROBE は有用であろう。診断方法、腫瘍マーカー、および、遺伝子関連について特別に焦点をおいた研究を報告する場合には、それぞれ、STARD 声明¹³、REMARK 声明¹⁴、および、STREGA 声明¹⁵を参考にするとよい。

STROBE のチェックリスト項目

これより STROBE のチェックリスト (TABLE 1) の 22 項目について説明を行うとともに、それぞれの項目ごとに、実際に出版されている論文を例に挙げる。いくつかの記載例においては、先行研究論文の引用に関する記述を編集により省いたものもあり、また、略語を用いずにそのまま表記したものもある。22 項目のうち 18 項目は 3 つの研究デザインに共通するものであり、残りの 4 項目は研究デザインに特有なものである。*印の付いている項目 (例えば項目 8*) については、ケース・コントロール研究であればケース群とコントロール群に分けて、あるいは、コホート研究や横断研究であれば曝露群と非曝露群に分けて情報を提供すべきであるということを示す。われわれは、論文の著者にこれらすべての項

目を論文中に記載することを推奨するが、厳密な位置や順序について指定することはしない。例えば、われわれは、著者が1つのセクション、文章、あるいは、表に対して複数の項目の適用があることを想定し、多くの項目で結果に関する報告方法について論じている。

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) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —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 <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —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) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —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—eg, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed (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 (eg, 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) <i>Cohort study</i> —Summarize follow-up time (eg, average and total amount)
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg 95% confidence intervals). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorised (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg, analyses of subgroups and interactions, and sensitivity analyses
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
Generalizability	21	Discuss the generalizability (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. Separate versions of the checklist for cohort, case-control and cross-sectional studies are available on the STROBE website at www.strobe-statement.org.

Box 1 – STROBE で扱う主要なデザインについて

コホート研究、ケース・コントロール研究、および、横断研究の各デザインでは、ある

集団において一定期間に発症する健康関連イベントについての調査方法がそれぞれ異なる。これらの研究では、疾病（の有病や発生）や疾病の寛解、廃疾、合併症、死亡、生存、および、リスク因子の存在を含む様々な健康関連イベントについて取り扱う。

コホート研究では、研究者は時間を追って対象者を追跡する。彼らは対象者とベースラインにおける曝露状況についての情報を得た上で、しばらく時間をおき、それからアウトカムの発生について評価する。研究者は通常、曝露を受けた者と受けていない者の間、あるいは、曝露状況の異なるいくつかのカテゴリー間で比較を行う。研究者は複数の異なるアウトカムについて評価し、追跡期間中の複数の時点における曝露変数とアウトカム変数の関連について調べる。閉じたコホート（closed cohort）（例：出生コホート）では、研究開始時点で決められた参加者数を登録し、それ以降の経過を通常決められた日付までの一定の期間追跡する。一方、開いたコホート（open cohort）では、研究対象者は流動的であり、時期によって対象者が出入りする（例：地域の住民）。開いたコホートの集団は、死亡、出生、転居により変化するが、年齢・性といった変数に関しての人口構成は、特に短期間では、概ね一定に保たれている。閉じたコホートでは累積罹患（cumulative incidences [risks]）や罹患率（incidence rates）が計算され、曝露群と非曝露群を比較することで、リスク比（risk ratio）や率比（rate ratio）の推定値が得られる。開いたコホートでは罹患率や率比が推定できる。

ケース・コントロール研究では、研究者は特定の疾患アウトカムを有する者（ケース）と有していない者（コントロール）の曝露状況について比較する。研究者は、実際にあると考えられるコホート集団や横断的な一般集団を代表するようにケースとコントロールを集めることを計画する。その対象集団は地理的に定義することができるが、医療圏など医療施設の管轄区域として大まかに定義することも可能である。ケースとなるサンプルは、研究に利用できるケースの100%あるいは大部分であるかもしれないが、一方、コントロールとなるサンプルは、一般に、着目しているアウトカムを有していない者のうちのごく一部である。コントロールはケースが生じるコホートや集団を代表するものである。研究者は、ケースとコントロールの間で疾患の原因として予想される曝露のオッズ比（odds ratio）を算出する（Box 7を参照）。ケースやコントロールのサンプリング方法や研究の対象集団の特性によって、ケース・コントロール研究で得られたオッズ比はリスク比、率比、あるいは、有病オッズ比（prevalence odds ratio）として解釈される^{16,17}。出版されているケース・コントロール研究の大半は、開いたコホートからサンプリングを行っており、率比を直接推定することができる。

横断研究では、研究者は同じ時点でサンプリングした個々の対象者を評価し、多くの場合、曝露あるいは疾患のリスク要因に特徴づけられる集団に占める有病割合（あるいは、有病率[prevalence]）について検討する。横断研究によっては分析的なものもあり、曝露と疾患の間に見込まれる因果関係を定量化する目的のものもある。このような研究では、コホート研究のように、曝露群の間での有病割合を比較する解析が行われることもあり、また、ケース・コントロール研究のように、疾患を有する群と有しない群の間で曝露のオッ

ズを比較する解析が行われることもある。どの研究デザインにおいても考える問題のうち、とりわけ横断研究で困難であることがはっきりしているのは、想定している曝露が疾患に先行して存在していることを証明することである。もちろん研究によっては曝露とアウトカムの時間的順序が明らかな場合もあるかもしれない。例えば、曝露変数が先天的あるいは遺伝的なものである研究では、その曝露と疾患について同時に測定しているとしても、曝露が疾患に先行して存在することを確信してよいだろう。

タイトル・抄録 (TITLE AND ABSTRACT)

1(a)タイトルまたは抄録の中で研究デザインを一般に用いられている用語で明示する。

記載例

製靴業に従事する労働者における白血病の発症：ケース・コントロール研究¹⁸

解説

読者がタイトル・抄録を見て、容易にその研究デザインについてわかるようにすべきである。通常用いられる語句で明確に記すことによって、電子データベース上の論文のインデックスを正しく作成することにも役立つ^{19,20}。

1(b)抄録では研究で行われたことと明らかにされたことについて十分な情報を含み、かつバランスのよい要約を記載する。

記載例

背景：HIV 感染患者の平均生存期間 (expected survival) については、公衆衛生上大きな関心もたれている。

目的：一般住民と比較し、HIV 感染患者の生存時間と年齢による死亡率について推定する。

デザイン：地域住民を対象としたコホート研究 (population-based cohort study)。

セッティング：1995 年から 2005 年までにデンマークで治療を受けた全ての HIV 感染患者。

患者：デンマーク国内 HIV コホート研究 (The Nationwide Danish HIV Cohort Study) に登録された患者 1 名に対し、性別、生年月日、および、居住地によって、99 名の一般住民をマッチさせた。

測定：著者らは、25 歳からの生存時間を推定するために、年齢をタイム・スケールとした Kaplan-Meier 法により計算した生命表 (Kaplan-Meier life tables) を作成した。HIV 感染患者とそれに対応した一般住民からの対象者を、HIV 患者が診断された日から観察し、死亡日、移動日、あるいは、2005 年 5 月 1 日まで追跡した。

結果：この研究では HIV 感染患者 3,990 名と一般住民からの 379,872 名を対象としており、それぞれ 22,744 人年 (中央値 5.8 年/人) と 2,689,287 人年 (中央値 8.4 年/人) 観察された。参加者の 3%が追跡不可能であった。25 歳時からの生存期間の中央値は、HIV

感染患者で 19.9 年 (95%信頼区間[confidential interval: CI]: 18.5-21.3) であり、一般住民で 51.1 年 (95%CI: 50.9-51.5) であった。HIV 感染患者においては、2000 年から 2005 年の間で生存期間が 32.5 年 (95%CI: 29.4-34.7) にまで延びた。既知の C 型肝炎合併例 (16%) を除外したサブグループでは、上と同様の期間中の生存期間の中央値は 38.9 年 (95%CI: 35.4-40.1) であった。一般住民に対して HIV 感染患者の相対死亡率 (relative mortality rate) は年齢の増加に伴って減少し、超過死亡率 (excess mortality rate) は年齢の上昇に伴い増加した。

研究の限界：今回得られた死亡率は、この研究で最長観察期間である 10 年を越えても当てはまることを前提としている。

結論：強力な抗レトロウイルス治療が行われる現代においては、HIV 感染を診断された若年者の推定生存期間の中央値は 35 年を越えている。しかし、一般住民と比較すると、これらの患者の死亡率をさらに低下するような努力が引き続き必要である²¹。

解説

抄録は、読者が研究の内容を把握し、さらにその論文を読み進めるべきかどうかを判断する重要な情報を提供する。標準的な構成要素として、リサーチ・クエスション (research question) の明示、方法および結果の簡潔な記述、そして、結論が抄録には含まれる²²。抄録は研究の主要部分を要約し、その論文に記載されている情報のみを提示したものでなければならない。われわれは、抄録には、研究の参加者数や、関連性についての推定値とそれに伴うばらつきや不確実性 (例：オッズ比とその信頼区間) といった、数値で表せる主要な結果を記載することを推奨する。ある曝露がアウトカムに対して統計学的に有意に関連する、あるいは関連しないことを述べるだけでは不十分であるとわれわれは考えている。

研究の背景やデザイン、運営、および、解析に関する一連の見出しをつけることによって、読者はすぐに要点となる情報を把握することができる²³。多くの雑誌では、このような構造化抄録 (structured abstract) の形で記載することになっており、構造化されていない抄録に比べてより質が高く、より迅速に情報が提供できるようになっている^{24,25}。

緒言 (INTRODUCTION)

緒言のセクションには、報告する観察研究について、なぜこの研究を行ったのか、および、どのような疑問や仮説について取り組んだのかを記載すべきである。それによって読者は研究のコンテキスト (背景[context]) を理解することができ、その研究の現在の知見に加えることが可能な潜在的な貢献について判断することができる。

2.背景 (background) / 論拠 (rationale) : 研究の科学的な背景と論拠を説明する。

記載例

小児期と青年期の肥満の割合 (prevalence) が上昇していることが懸念されることに伴い、

小児期の肥満と成人における心血管系のリスクおよび死亡率との関連についての関心が高まっている。小児期の肥満は、小児期あるいは青年期にかなりの社会的および心理的影響をもたらすが、それが成人になってからどのように社会的、社会経済的、および、心理的な影響を与えるかについてはほとんど知られていない。最近のシステマティック・レビュー (systematic review) によれば、身体的な健康に関するアウトカムに関わる研究以外に、小児期の肥満によるアウトカムに関わる縦断研究 (longitudinal studies) はなされておらず、たった 2 つの縦断研究において、青年期の肥満が与える社会経済的な影響について言及されているのみである。Gortmaker らは、1981 年に青年期後期 (late adolescence) に肥満であったアメリカ人女性が、そうでない人と比べて、7 年後の時点で結婚していないことが多く、また、収入も少ない一方で、肥満の男性もそうでない人と比べてより結婚していない事実を発見した。Sargent らは、1974 年の時点で 16 歳であったイギリス人女性では、肥満でなかった同僚と比べ、23 歳時での収入が 7.4%低い (が、男性ではそのようなことはなかった) ことを示した。(中略) われわれは 1970 年の英国出生コホート (British Birth Cohort) から縦断的なデータを用いて、小児期の肥満が与える成人期の社会経済的、教育的、社会的、および、心理的アウトカムについて検討した²⁶。

解説

研究の科学的な背景は、読者にその研究の重要なコンテキストを提供する。また、それはその研究がどういった段階にあり、何に焦点を置いているのかを明確にする。さらに、そのトピックで既知のことと、この研究によって示そうとすることは既知の知見とどのような差異があるのかの概観を与える。背景では、最近の関連する研究およびそれらについてのシステマティック・レビューについて言及するべきである。

3.目的 (objectives) : 事前の仮説を含む形で研究の具体的な目的を明記する。

記載例

われわれの研究の主要な目的は、1)4 つの地域のプライマリ・ケア (primary care) で成人に対する医療に着目し、様々な社会経済的な背景を有する患者を対象とし、そこで見られる女性患者の家庭内暴力の現状 (prevalence) を把握すること、および、2)現在虐待を受けている患者と虐待を受けていない患者との間の人口統計学的 (demographic) および臨床的な差異について明らかにすることである²⁷。

解説

目的は、その研究の詳細な狙い (aims) である。よくまとめた目的には、対象、曝露、アウトカム、そして、評価したパラメーターが明確にされている。研究の目的は、具体的な仮説、あるいは、その研究によって明らかにされるべきリサーチ・クエスチョンとして明確に示されるだろう。いくつかの状況、例えば、初期の発見の段階では、研究の目的は具体的でない場合もある。そのような場合でも、研究者の意図を明確に示した形で報告さ

れるべきである。例えば、もしサブグループ解析や追加解析が研究の本来の目的ではなく、データ解析の過程で生じたものであれば、それに応じた説明がなされるべきである（項目 4、17、20 を参照）。

方法（METHODS）

方法のセクションには、第三者が研究の重要な側面（aspect）を理解し、その方法が信頼性と妥当性のある結論を導くために適切かを判断し、そして、元々の計画からのずれが正当なものかを判断するために、どのように綿密に計画を立て、何を実施したかについて十分に詳細に記載するべきである。

4. 研究デザイン（study design）：研究デザインの重要な要素を論文の早い段階で示す。

記載例

われわれはケース・クロスオーバー研究（case-crossover design）を用いたが、これはケース・コントロール研究（case-control design）のバリエーション（変化形[variation]）であり、短時間の曝露（運転者の携帯電話の使用）が稀なアウトカム（交通事故）のリスクを一時的に上昇させる場合に適用可能な方法である。ここでは事故が推測される時間の携帯電話の使用と、同じ運転者の別の適切な時間帯の携帯電話の使用を比較する。運転者自身がコントロールであるため、事故のリスクに影響する運転者の特性が（短時間で変わらないため）制御可能なデザインとなっている。コントロール期間と事故を起こした時のリスクが類似していることが重要であるため、われわれはハザード期間（事故直前）の電話の使用頻度とその前週のコントロール期間（ハザード期間に相当する、参加者が運転しているが事故を起こさなかった期間）の電話の使用頻度を比較した²⁸。

解説

方法のセクションの早い段階（あるいは緒言の最後）で研究デザインの主要な要素について述べることをわれわれは勧める。これによって読者は研究の骨格（basics）を知ることが出来る。例えば、コホート研究で実施された研究であれば、著者は自分の研究はコホート研究であり、対象が一定期間追跡されることを示すべきであるし、コホートを形成するのはどのような集団かということとその曝露状況を示すべきである。同様に、もしケース・コントロール研究で実施された研究であれば、ケースとコントロールおよびそれらの源泉（ソース）集団（source population）について記載するべきである。もし横断研究で実施された研究であれば、その集団と研究を行った時点について示すべきである。研究が主たる3つの研究デザインのバリエーションである場合、研究デザインを明確にするためにさらに説明を付け加える必要がある。例えば、上記の記載例では、ケース・コントロール研究の発展形であるケース・クロスオーバー研究についての簡潔な原理が記載された²⁸。

われわれは研究を単純に「前向き（prospective）」とか「後ろ向き（retrospective）」と呼ぶ

ことを勧めない。なぜなら、これらの用語の定義が明確ではないからである²⁹。第1の使われ方では、前向きという用語はコホート研究を意味し、後ろ向きという用語はケース・コントロール研究を意味する³⁰。第2の使われ方は、前向きコホート研究と後ろ向きコホート研究の区別であり、研究のアイデアが出た時期とデータ収集のタイミングにより区別している³¹。第3の使われ方は、前向きケース・コントロール研究と後ろ向きのケース・コントロール研究の区別であり、ケースが選択された時点で関心のある曝露についてのデータが存在するかどうかにより区別している³²。これらの用語の使われ方に対して³³、コホート研究を記述する際には「同時的 (concurrent)」と「歴史的 (historical)」で代用しようという意見もある³⁴。STROBE では「前向き」あるいは「後ろ向き」という用語は使わず、さらに、その代用である「同時的」や「歴史的」という用語も用いていない。われわれは、著者がこれらの用語を使う際にはいつでも、これらの用語をどのような意味で用いたのかを定義して用いることを推奨する。われわれが推奨するもっとも重要なことは、著者がいつどのようにデータ収集を行ったのかを明確に記載するということである。

方法のセクションの最初の部分では、その報告が、ある1つの研究をもとに発表されたいくつかの論文のうちの1つかどうかということについても触れておいた方がよい。新しい報告が研究の当初の目的と一致している場合は、通常当初出版された論文を引用し、研究の主要な特徴を簡潔に示しておく。しかし、一般に、研究の目的は時間と共に発展する。研究者は時として、元々は意図していなかった目的でデータを用いることがある。例えば、本来行政上の目的で収集された人口動態統計、調査票にはあつた質問項目であるが特に解析には使用する予定のなかった項目、あるいは、別の目的で集められた血液サンプルなどである。例えば、The Physicians' Health Study は、アスピリンとカロチンについてのランダム化試験として行われたが、のちに第 V 因子遺伝子の点突然変異 (point mutation) が心筋梗塞や脳梗塞ではなく静脈血栓症のリスクを上昇させていることを示すために用いられた³⁵。既存データの2次利用は観察研究の創造的な部分であり、必ずしも結果の信頼性や重要性を落とすものではない。しかし、簡潔に当初の研究の目的を述べることで、読者の研究のコンテキストやデータの潜在的な限界についての理解が深まるであろう。

5.セッティング (setting) : セッティング、実施場所のほか、基準となる日付については、登録、曝露、追跡、および、データ収集の期間を含めて明記する。

記載例

Pasitos Cohort Study では、1998年4月から2000年10月までテキサス・El Paso の Socorro と San Elizario の女性乳児幼児クリニックとメキシコ・Ciudad Juarez のメキシコ社会保障協会の母性小児医療クリニックで妊婦を募集した。ベースラインとして、コホートに登録される新生児の誕生に先立って、家庭環境について母親にインタビューをおこなった。現在実施中のこのコホート研究は、生後6ヶ月から開始して6ヶ月毎の追跡調査を目指している³⁶。

解説

読者は研究結果のコンテキストや一般化可能性 (generalizability) を評価するために、セッティングや場所についての情報を必要としている。環境因子や治療のような曝露は時と共に変化する。研究方法も同じように発展する可能性がある。いつ研究が行われ、どの期間参加者が募集されて追跡されたかを知ることにより、その研究の歴史的な位置づけが可能となり、結果を解釈する際に有用な情報となる。

セッティングについての情報は、募集を行った場所やそのソース (source) (例：選挙人名簿、外来クリニック、がん登録、3次医療センター) を含む。その情報では、調査が行われた地方、街、病院、あるいは、開業場所について言及することがある。単に期間の長さを述べるだけでなく、日付を述べることを勧めている。曝露、疾患発生、登録、追跡の開始と終了、および、データ収集について異なる日付の組み合わせがあり得るであろう。しかし、現状としては、がん専門の医学雑誌で生存解析を行った 132 の報告のうち、約 80% で患者発生についての開始日と終了日について述べているにも関わらず、わずか 24% でしか追跡の終了日について報告していなかったというデータがある³⁷。

6.参加者 (participants) :

6(a)コホート研究：適格基準 (eligibility criteria)、および、対象者のソースと選定方法を示す。追跡の方法についても記述する。

記載例

Iowa Women's Health Study の参加者は、1985 年のアイオワ運転免許者リストの 55 歳から 69 歳の全女性からランダムに抽出したサンプルである。このリストはこの年齢グループのアイオワ女性のおおよそ 94% を代表している。(中略) 健康状態 (vital status) と住所変更を把握するため、追跡質問紙は 1987 年 10 月と 1989 年 8 月に郵送された。(中略) 非メラノーマの皮膚がんを除くがん罹患がアイオワ州健康登録で確認された。(中略) Iowa Women's Health Study コホートでは、名、姓、旧姓、郵便番号、誕生日、および、社会保障番号の組み合わせによりデータの照合がされた³⁸。

6(a)ケース・コントロール研究：適格基準、対象者のソース、ケースの確定方法、および、コントロールの選択方法を示す。ケースとコントロールの選択における論拠を示す。

記載例

1999 年から 2000 年までに診断された皮膚メラノーマの患者が、アイオワがん登録で確認された。(中略) コントロールも同様にアイオワがん登録より、同時期に結腸直腸がんが診断された患者から得た。結腸直腸がんがコントロールに選ばれたのは、これがよくあるがんであり、比較的予後が良く、ヒ素への曝露と結腸直腸がん発生の関連が確実なものではないためである³⁹。

6(a)横断研究：適格基準、および、対象者のソースと選択方法を示す。

記載例

われわれは退院時診断に関するコードから主要診断が ICD9 (International Classification of Diseases, 9th Revision, Clinical Modification) の心筋梗塞 (コード 410) のものを選び、退院に引き続く介護エピソードに関する 5 桁目が 2 のコードを除外した。(中略) 1994 年 2 月から 1995 年 6 月までのメディケア・コホート全体からの心筋梗塞患者のサンプリングを行った。(中略) 適格基準は、少なくとも 30 分以上かつ 12 時間以下の胸痛のあとで来院し、最初の心電図で 2 つの連続するリードで少なくとも 1mm の ST 上昇があった患者とした⁴⁰。

解説

研究の参加者について詳しく記載することは、読者が結果の妥当性について理解することに役立つ。通常、研究者は研究集団を、一定の臨床的、人口統計学的、あるいは、その他の特徴で限定する。典型的な適格基準は年齢、性別、診断そして合併症に関わるものである。これらの重要性にも関わらず、適格基準が適切に報告されないことがよくある。脳卒中の観察研究の調査によると、49 の報告のうち 17 報 (35%) でしか適格基準を特定していなかった⁵。

適格基準は、組み入れ基準 (inclusion criteria) と除外基準 (exclusion criteria) と表現されることがあるが、この区別は必ずしも必要でなく、また有用でもない。しかし、われわれは著者らに、全ての適格基準を報告すること、および、どのような集団から選択されたグループなのか (例：地域や地方の一般集団) ということと、その募集の方法 (例：紹介あるいは広告を通じての自己応募) について記述することを勧める。

非回答者 (nonresponse) や追跡不能者が最小限になるような手段や全ての参加者に均一な手段を講じたかどうかなどの詳細な追跡方法を知ることが、結果の妥当性の判断に役立つ。例えば、急性感染を見つけるために IgM 抗体を用いた研究では、読者らは IgM 抗体の血液検査の間隔を知る必要があり、これによって血液検査の間隔が長すぎることで何らかの感染が見逃されていないかを判断することができる⁴¹。曝露群と非曝露群の間で追跡方法が異なる別の研究では、読者らは非回答者あるいは追跡不能者において、イベント発生を確認する間隔が異なっていたり、確認方法に違いがあることでかなりのバイアス (bias) が生じていることに気付くであろう⁴²。したがって、われわれは研究者に、追跡に用いた方法についてとそれらが参加者全てで同等であったか、および、変数が完全に測定されたか (項目 14 も参照) を記載するように勧めている。

ケース・コントロール研究では、ケースとコントロールの選択が結果の解釈に極めて重要であり、それらの選択方法は研究の妥当性に大きな影響を及ぼす。一般に、コントロールはケースが発生した集団を反映するべきである。コントロールのサンプリングには様々な方法があり、それぞれに一長一短がある：ケースが一般集団に由来する場合は、名簿からのサンプリング (population roster sampling)、ランダム番号ダイアル法 (random digit dialing)、近隣者コントロール (neighborhood control)、あるいは、友人コントロール (friend

control) が用いられる。近隣者あるいは友人のコントロールを用いることで、内在する曝露に関するマッチング (matching) を行うことになる¹⁷。他の疾患を持つコントロールは、特に病院ベースのケースにとって、集団ベースのコントロールよりも利点を持つことがある。なぜなら、他の疾患を持つ病院ベースのコントロールは病院で集積された集団を良く反映し、思い出し方についてケースと似通っており、さらに、募集も容易であるためである。しかし、関心のある曝露がコントロールとなった患者の疾患の悪化や入院状況のリスクに影響を与える場合には問題になりうるであろう^{43,44}。この問題を改善するために、もっとも問題のない疾患患者を取り混ぜてコントロールとして用いられることがよくある⁴⁵。

6(b)コホート研究：マッチングを行った研究ではマッチングの基準と曝露群と非曝露群の数を示す。

記載例

われわれは以下のプロトコール (protocol) に従い、最初にスタチンの投与を受けた患者それぞれについて、傾向スコアに基づいたマッチング (propensity-based matching) によりスタチン投与を受けていない非投与例 1 名を同定した。第 1 に、スタチンの使用あるいは敗血症のリスクと関連がありえる要因の広汎なリストに基づいてコホート全体の全ての患者について傾向スコア (propensity scores) が計算された。第 2 に、各スタチン投与例がスタチン非投与例のプール (集積された候補者集団[pool]) と性、年齢 (前後 1 歳)、および、登録日 (前後 3 ヶ月) によりマッチされた。第 3 に、各スタチン投与例に最も近い傾向スコア (0.2 標準偏差以内) を持つコントロールをスタチン投与例に対して 1 対 1 となるように選択し、残ったコントロールは用いなかった⁴⁶。

6(b)ケース・コントロール研究：マッチングを行った研究ではマッチングの基準とケースあたりのコントロールの数を示す。

記載例

われわれはケース 1 例あたり 5 人のコントロールを研究対象集団から選択することを目標とした。この研究集団は、一般診療カルテ上で自閉症あるいは広汎性発達障害と診断されておらず、ケースが広汎性発達障害と診断された日に生存し診療に関わり登録されている患者の集団である。コントロールは誕生年 (前後 1 歳)、性、および、一般診療状況によりそれぞれのケースにマッチされた。ケース 300 例については、すべてのマッチング基準を満たす 5 人のコントロールを同定可能であった。残りの 994 例については、1 人以上のコントロールが除外された⁴⁷。

解説

マッチングはケース・コントロール研究でよく用いられるが、研究者は時に追跡開始時点で両群を比較可能な集団とするためにコホート研究でマッチングを用いることがある。コホート研究でのマッチングはケース・コントロール研究ほど複雑ではない。例えば相対

リスクの推定にマッチングを考慮することは必ずしも必要でない⁴⁸。コホート研究でマッチングを行うと統計学的な精度が上昇するので、研究者は解析でマッチングを用いることにより信頼区間を狭めようとすることもある。

ケース・コントロール研究では、ケース群とコントロール群における変数の分布、特に交絡（confounding）の可能性のある変数の分布を揃えて研究の効率（efficiency）を高めるためにマッチングが行われる^{48,49}。マッチングはケース 1 例あたり 1 人、あるいはそれ以上のコントロールを用いることも含め様々な方法で行うことが可能であるため、マッチした変数選択の合理性と用いた方法の詳細について記載すべきである。よく用いられるマッチングの方法は頻度マッチング（frequency matching）（グループ・マッチング[group matching]とも呼ばれる）と個別マッチング（individual matching）である。頻度マッチングではマッチした変数の分布がケース群のそれと同一あるいは類似するようにコントロール群を選択する。個別マッチングでは各ケースについて、1 人から数人のコントロールをマッチする。ケース・コントロール研究でのマッチングは、直感的には魅力的であり時として有効であるが、多くの欠点を抱えており、常に適切とは言えず、解析でも考慮にいれる必要がある（Box 2 を参照）。

明らかに単純なマッチング方法ですら、不十分にしか報告されていないこともある。例えば、著者は「5 歳以内で」あるいは「5 歳の年齢幅で」ケースにコントロールをマッチさせたとして記したとする。これではケースが 54 歳だった場合に対応するコントロールは 50 歳から 54 歳までの 5 歳の年齢幅を指すのか、それとも 49 歳から 59 歳までなのか、いったいどちらが 54 歳から 5 歳以内であるのか不明である。もし広い年齢幅（例：10 歳）を採用した場合であれば、例えば、コントロールを平均するとケースより若くなるといったような年齢による残差交絡（residual confounding）が生じる危険性がある（Box 4 も参照）。

Box 2 – ケース・コントロール研究におけるマッチング

いかなるケース・コントロール研究においても、ケースに対するコントロールのマッチングを用いるのかどうか、もしそうなら、どの要因をマッチさせるのか、マッチングに用いる正確な方法はどうか、そして、統計解析の方法は適切か、ということについて正しい選択が必要である。全く要因をマッチさせない研究もあるが、その場合には、いくつかの重要な潜在的な交絡因子（年齢や性別）の分布が、ケースとコントロールで全く異なるということが起こりうる。これは解析で調整可能であるが、統計学的効率（statistical efficiency）を大きく損なう可能性がある。

ケース・コントロール研究におけるマッチングの使用とその解釈には難しい問題があり、特に、マッチングをいくつかのリスク要因について行い、それらが主に関心のある曝露と関連している可能性があるときに問題となる^{50,51}。例えば、大規模な薬剤疫学データベースを用いた心筋梗塞と経口避妊薬のケース・コントロール研究（訳者注：ネステッド・ケース・コントロール研究[nested case-control studies]）においては、コントロールとして用いる

ことが可能な何千人もの女性の情報があり、研究者らは心筋梗塞の各ケースに対して似たようなレベルのリスク要因を持つコントロールをマッチさせるであろう。その目的の一つは経口避妊薬処方に与える要因を調整し、処方交絡（適応による交絡[confounding by indication]）を制御することにある。しかし、結果としてそのようなコントロール群はもはや源泉（ソース）集団で経口避妊薬を用いている集団を代表する集団ではなくなる。例えば、心筋梗塞は高齢で発症しやすいため、コントロール群は元の集団よりも高齢となる。これにはいくつかの意味が含まれている。マッチされた要因が曝露と関連しているときには、データの粗解析（交絡要因で調整してない解析[crude analysis]）の結果は通常オッズ比が 1 に近づくようなバイアスが働く。その解決方法はマッチングを考慮した解析、あるいは、層別解析（stratified analysis）を行うことである（項目 12d を参照）。さらに、マッチされたコントロール群は源泉（ソース）集団を代表した集団ではなくなるため、コントロール群の中での曝露の分布をもはや集団の寄与割合の推測に用いることはできない（Box 7 を参照）⁵²。さらに、マッチされた要因の効果はもはや推計することはできないということや、マッチングのための適切なコントロールを探すことは困難であるという問題がある。マッチングを行わないコントロールは得られやすく、また、コントロール群の規模を大きくできるという利点があるため、マッチングを行わないコントロールを用いたデザインが好まれる。

別の問題としてはオーバー・マッチング（over matching）が挙げられる。これはマッチングを用いたケース・コントロール研究の効率を損ない、また別の状況ではバイアスを招くものである。マッチされた要因が曝露と密接に関連している場合には情報は失われ、研究の検出力は低下する。このとき同じマッチされたセット内の個々人は、同一あるいは類似した曝露レベルを持つ傾向となり、したがって意味のある情報とならない。マッチされた要因が交絡ではなく、曝露と疾患の因果の中間因子である場合、マッチングは取り返しのつかないバイアスを招いてしまう。例えば、人工授精は、多重産の増加と低体重出生幼児が増加するために周産期死亡と関連している⁵³。出産数と出生体重をマッチすることで 0 に近づくバイアスが働き、これを解析で解決することはできない。

マッチングは直感的に魅力的なものであるが内容的に複雑なため、方法論の専門家はケース・コントロール研究でルーチンにマッチングを用いることを勧めていない。彼らは代わりに各々の可能性のあるマッチング因子について注意深く慎重に考慮し、マッチングするのではなく、調整因子として測定し用いることを勧めている。これに対応して、採用されたマッチング因子の数は減少してきており、頻度マッチングを用いることが増えてきている。これは頻度マッチングが上記で述べたようないくつかの問題点を回避するからであり、さらにまったくマッチングを行わないケース・コントロール研究も増えてきている⁵⁴。交絡の分布（例：年齢）が、マッチしていない比較群間で極端に異なるかもしれないときマッチングは未だ魅力があり、また必要でもある^{48,49}。

7.変数 (variables) : すべてのアウトカム、曝露、予測因子 (predictor)、潜在的な交絡因子 (potential confounder)、潜在的な効果指標の修飾因子 (potential effect modifier) を明確に定義する。該当する場合は、診断方法を示す。

記載例

先天奇形の内、大奇形のみが解析に含まれた。小奇形はヨーロッパ先天奇形登録 (EUROCAT) の除外リストにしたがって除外された。一人の小児が2つ以上の大奇形を有していた場合は、器官毎に1つのアウトカムとして扱われた (中略)。統計解析において、交絡因子の可能性があると考えられた因子は、出産時の母親の年齢と過去の出産回数であった。効果修飾の可能性があると考えられた因子は、抗てんかん薬処方償還請求書 (reimbursement) に記載のある母親の年齢と出産時の母親の年齢であった⁵⁵。

解説

著者は、アウトカム、曝露、予測因子、潜在的な交絡因子、および、潜在的な効果指標の修飾因子を含め、解析において考慮し投入された全ての変数について定義づけを行うべきである。疾患のアウトカムに関しては、十分に詳細な診断基準を提示する必要がある。これはケース・コントロール研究におけるケースの適格基準や、コホート研究の追跡期間における疾患イベントの基準や、横断研究において関心のある疾病の定義についても言えることである。明確な定義、あるいは、その定義に従って手順を踏むことは、研究において最も知りたいあらゆる疾患の状況を把握する上で特に重要である。

いくつかの研究においては、曝露変数の用語として「決定因子 (determinant)」や「予測因子 (predictor)」という用語が適している場合があるだろうし、またアウトカムは「エンドポイント (end points)」と呼ばれる場合もあるだろう。多変量モデル (multivariate model) では、著者はしばしばアウトカムに「従属変数 (dependent variable)」という用語を、曝露や交絡要因に「独立変数 (independent variable)」や「説明変数 (explanatory variable)」という用語を用いる。ただし、後者は交絡と曝露を区別しないため正確な用法ではない。

初期の探索段階 (discovery phase) における探索的な解析で多くの変数が測定され扱われている場合は、それぞれの変数の詳細について書かれたリストを、付録 (appendix)、追加の表、あるいは、別の発行物で提示することを考慮すべきである。参考までに、The International Journal of Epidemiology は最近、「コホート・プロフィール (cohort profile)」という新しいセクションを立ち上げたが、そこには、ある特定の研究についての、種々の時点においてそれぞれ測定された事柄の詳しい情報が掲載されている^{56,57}。最後に、われわれは、著者は最終的なモデルに投入された変数だけを選んで報告するのではなく、統計解析において考慮された全ての「候補となった変数」について言及することを推奨する (項目 16a も参照)^{58,59}。

8.データソース (data source) /測定方法 (measurement) : 関連する各因子に対して、データソース、評価・測定方法の詳細を示す。2つ以上の群がある場合は、測定方法の比較可

能性 (comparability) を明記する。

記載例 1

カフェインの総摂取量は初めに米国農務省 (Department of Agriculture) の食品成分表 (food composition sources) を用いて計算した。この計算で、コーヒー1杯にはカフェインが 137mg、紅茶 1 杯には 47mg、コーラ 1 ビンまたは 1 缶には 46mg、チョコレートキャンディ 1 つかみ (serving) には 7mg 含まれると想定された。このカフェイン摂取量の測定法は、NHS I コホートと、男性の保健医療専門職 (health professionals) を対象にした同じようなコホートでも妥当性が示された (中略)。高血圧の自己診断の信頼性は米国 NHS I コホートにおいて確認された⁶⁰。

記載例 2

マッチされたケースとコントロールのサンプルは常に同じバッチ (batch) 内で分析され、実験者 (laboratory personnel) がケースとコントロールを区別できないようにした⁶¹。

解説

曝露、交絡因子、および、アウトカムの測定方法は研究の信頼性 (reliability) や妥当性 (validity) に影響を与える。曝露やアウトカムの測定誤差 (measurement error) や誤分類 (misclassification) によって因果関係を検出することが困難になる可能性があり、また、反対に偽の関係が作り出されてしまう可能性もある。交絡の可能性のある要因の測定誤差によって残差交絡が生じる可能性が高くなる^{62,63}。よって、参照基準の詳細も含めて、著者が使用した評価法や測定法の信頼性あるいは妥当性に関する研究結果を報告することは有用である。記載例 1 のように妥当性研究 (validation study) を単純に引用するよりも、推定された妥当性や信頼性の指標を示すことをわれわれは勧める。そうすれば、それらの値を測定誤差の調整や感度分析 (sensitivity analysis) に用いることが出来る (項目 12e, 17 を参照)。

加えて、データ収集の方法が比較する各グループ間で異なっていないかどうかを知ることが重要である。これは記載例 2 のような実験的研究等の状況において重要であろう。例えば、あるインタビュアーがまず全てのケースに質問してからその後にコントロールに質問した場合、あるいはその反対でも、学習効果によるバイアスを生じうる。この問題はインタビューの順番をランダムに決める等の方法で解決できる。情報バイアス (information bias) は、比較する各グループが同じ診断検査を受けていなかった場合、あるいは同じ検査でも一方だけが検査回数を多く受けていた場合でも生じる (項目 9 を参照)。

9. バイアス (bias) : バイアスが生じないようにとられた方策があればすべて示す。

記載例 1

自殺に関するケース・コントロール研究のほとんどにおいて現在生存している人がコントロール群になっているが、われわれは自殺以外の原因で死亡した人をコントロール群とした (中略)。死亡した人からなるコントロール群の場合、そのリスクファクターの評価に用

いる情報源は、最近家族や近親者の死亡を経験した人から提供される。したがって生存している人をコントロールとした場合よりも自殺者に関する情報源との比較可能性が高い⁶⁴。

記載例 2

2 型糖尿病 (T2DM) の女性はそうでない女性と比べて眼科的検査をより綿密に受けているとすると、2 型糖尿病と開放隅角緑内障 (POAG) との関連に発見徴候バイアス (detection bias) が影響を与えてしまう。われわれは糖尿病の女性と糖尿病ではない女性のそれぞれの情報から眼科的検査の平均回数を比較した。また、より精密な眼科的検査と関連する共変量 (白内障や黄斑変性の有無、眼科診察の回数、および、身体診察の回数についての患者自己報告情報) のさらなる調整を行った上で、開放隅角緑内障の相対リスクの再計算も行った⁶⁵。

解説

バイアスを含んだ研究は、真実とは系統的に偏った結果を導く (Box 3 も参照)。研究の過程でバイアスを減らすためにどのような方策がとられているかを読者が知ることは重要である。理想的には、研究者は研究を計画する段階においてバイアスの可能性について注意深く考慮する。報告の段階で著者が、問題となるバイアスが尤もらしいかどうかを常に評価することをわれわれは推奨する。具体的に言えば、バイアスの方向とその大きさについて議論し、可能であればその程度を見積もるべきである。例えば、ケース・コントロール研究では情報バイアスを生じうるが、記載例 1 のように適切にコントロール群を選択すればそのバイアスを減らすことができる⁶⁴。研究参加者の検査状況が違っていたことが記載例 2 における問題点であった⁶⁵。そのため、著者は、この問題に対処するために集めた追加のデータについてさらに詳細に示した。縦断研究において変数測定が「変化」する可能性に対応する、あるいは、多くの観察者が変数測定を行う場合のばらつきを最小限に食い止めるといった、データ収集方法の品質管理プログラムを研究者が設定した時には、そのことについて説明をするべきである。

残念なことに、著者が結果を報告する際に、重大なバイアスについて示されていないことがよくある。1990 年から 1994 年までに出版された、がんの既往歴がある患者における新たながん発症リスクを調査した 43 報のケース・コントロール研究やコホート研究において、医学検査バイアス (medical surveillance bias) について言及されたのは 5 報に過ぎなかった⁶⁶。1998 年に精神科領域の 3 つのジャーナルに掲載されたメンタル・ヘルスに関する研究では応答バイアス (response bias) について言及されたのは 392 報中 13% に過ぎなかった⁶⁷。1999 年から 2003 年に掲載された脳卒中に関する 49 報のコホート研究の報告の内 14 報 (28%) で参加者のリクルートにおいて選択バイアス (selection bias) の可能性について述べており、35 報 (71%) で何らかのバイアスによって結果に影響を受けている可能性が言及されていた⁵。

Box 3 – バイアス

バイアスとは研究結果の真の値からの系統的な偏りのことである。バイアスは研究デザイン上で、あるいは、研究実施の際に生じるのが典型的であり、データ収集後には修正することはできない。バイアスと交絡は同義語ではない。バイアスとは偏りのある情報や対象の選択方法によって生じ、誤った関連を導き出すものである。交絡とは、曝露とアウトカムとの間に実際に関連が認められたとしても、想定されていない他の要因がいくつか潜在しそれが曝露とアウトカムの両者に関連するために、因果関係として解釈ができないことをいう (Box 5 を参照)。また、バイアスはランダム誤差 (random error) とも区別されなければならない。ランダム誤差とは測定されたデータのランダムなばらつき (statistical fluctuations) (いずれの方向でも) によって生じた真の値からのズレのことを指す。バイアスを生じる要因については数多く記載され様々な用語が用いられているが^{68,69}、われわれは2つのシンプルなカテゴリー (情報バイアスと選択バイアス) に大別する方がわかりやすいと考える。

情報バイアスは、データが完全にあるいは正確に測定されずに系統的な誤差が生じ、曝露やアウトカムに関する誤分類につながる場合に生じる。例えば、糖尿病の女性患者は糖尿病ではない女性よりも頻回に綿密な眼科的検査を受けているとすれば、それにより完全に緑内障を発見することになるであろう (項目 9 を参照)。非特異的な胃の不快感をもたらし薬剤を服用している患者はそうでない患者よりも多く胃カメラの検査を受け、たとえその薬剤が潰瘍の原因でなかったとしても、それにより多くの潰瘍が見つかるであろう。このタイプの情報バイアスは「発見徴候バイアス」とか「医学検査バイアス」と呼ばれている。情報バイアスの影響を評価する一つの方法としては、異なる研究グループでの医学検査 (medical surveillance) の強度 (intensity) を測定し、それを統計解析によって調整することである。ケース・コントロール研究における情報バイアスは、ケースがその疾患を持たないコントロールと比べて過去の曝露をより正確に、または、反対に、より不正確に思い出す場合に生じるし、より積極的に過去の曝露のことを報告しようとするか、反対に、より消極的に報告しようとする場合にも生じる (思い出しバイアス[recall bias]とも呼ばれている)。インタビュアー・バイアス (interviewer bias) はインタビュアーが研究仮説に気づき、意識的あるいは無意識の内に、選択的な情報収集を行った時に生じる⁷⁰。それゆえ、研究参加者や研究者に対する盲検化 (マスク化[blinding]) が有効なことがよくある。

選択バイアスはケース・コントロール研究において、曝露に影響されて (関連して) ケースとコントロールが対象として抽出されている場合に生じる。例えば、深部静脈血栓症の患者をリクルートしている医師は、下肢症状を有し避妊薬を服用している女性を深部静脈血栓症と診断するだろう。しかし、同様の下肢症状があっても避妊薬を服用していない女性はそう診断されないかもしれない。このようなバイアスは一定の診断方法 (diagnostic service) に照らし合わされたケースとコントロールを用いれば相殺されるであろう⁷¹。同様に、疾患登録も選択バイアスをもたらす可能性がある。すなわち、ある薬剤の曝露と疾患

の関係がわかっている場合、その疑わしい薬剤を使用している人の方をより多くケースに登録する傾向を生じる⁷²。応答バイアス (response bias) は別のタイプの選択バイアスで、研究参加者と非参加者との間での特徴の違いが発生割合や有病割合、あるいは、何らかの状況や関連性に影響を与える場合に生じる。一般的に、選択バイアスは研究の内的妥当性 (internal validity) に影響を与える。これは研究参加者全体を選択する際に起こる一般的な問題点 (その場合は内的妥当性よりもむしろ外的妥当性に影響を与えることになる) とは異なる (項目 21 を参照)。

10. サンプルサイズ (study size) : サンプルサイズがどのように決められたかを説明する。

記載例 1

研究期間内のその地域におけるケースの数からサンプルサイズ (sample size) を決定した⁷³。

記載例 2

ある調査では、その地域における産後うつの有病割合が 19.8%であることが示された。これを参考に正常出生体重児の母親におけるうつの割合を 20%として、また、栄養不良の子供を持つ母親のうつのオッズ比を 3 倍、検出力 80%、有意差を 5%として、72 組のケース・コントロールのセット (ケース 1 例に対してコントロール 1 人) を必要とした⁷⁴。

解説

リサーチ・クエスチョンに対して意味のある回答を導くために十分に狭い信頼区間を持つ推定値を得るためには、研究は十分な規模をもって行われるべきである。小さな関連がある場合を関連のない場合と区別するためにはサンプル数を大きくする必要がある。小規模研究により価値のある情報を提供される場合がよくある。しかし、信頼区間の幅が広い研究結果は、信頼区間の幅が狭い推定値が得られた研究結果に比べて、現在の知見に対する貢献の度合いが少ないことを示している。また、「興味深い」関連、あるいは、「統計的に有意な」関連を示している小規模研究は、有意差を示していない小規模研究に比べて数多く出版されている。これらの小規模研究は、新たな発見の前兆を示したものかもしれないが、読者にはその考えられる弱点を伝えるべきである。

観察研究においてサンプルサイズを決めることの重要度は状況によって変わる。すでに他の目的に使用されたデータを用いた解析の場合は、その解析が十分な統計的な精度を持ってその論文 (literature) に本質的な貢献をもたらすかどうかの主たる課題であって、サンプルサイズを考慮するかどうかについては形式上のものではなく、必ずしも必要ではない (will be informal)。形式上の事前のサンプルサイズの計算は、新しい研究を計画する場合に有用であろう^{75,76}。この場合のサンプルサイズの計算は、一般的な計算で求められる 1 つの数値が有する意味以上に不確実性に関連している。例えば、サンプルサイズの計算におい

て、主な関心のあるイベントの発生率の推定や他の推定は、全くのあてずっぽうではないものの、通常不正確である⁷⁷。最終的な解析で得られる精度は事前には決まらないことが多い。なぜなら多変量解析で交絡要因を入れること⁷⁸、重要な変数の測定精度の程度⁷⁹、あるいは、一部の対象者を除外することによって、その精度が減弱する場合があるからである。

疫学研究ではサンプルサイズの算出方法について説明・報告されることがほとんど無い^{4,5}。研究者がもし適切かつ定型的なサンプルサイズの計算方法を用いているのであれば、それを報告することを推奨する。そうでない場合は、研究者はサンプルサイズを決定した思考プロセスを提示すべきである（例えば記載例 1 で提示したような、実現可能なサンプルサイズが決められてしまう場合のように）。観察研究が統計的有意差に達して早めに打ち切られた時には、そのことを読者に伝えるべきである。サンプルサイズを後付けで正当化したり、後ろ向きにパワー計算を行ったりして読者を混乱させてはいけない⁷⁷。読者の視点からすれば得られた統計的な精度は信頼区間によって示される。信頼区間は統計的な不確実性を示すだけであって、研究に存在しうるあらゆる不確実性を示すものではないことを理解すべきである（項目 20 を参照）。

11.量的変数（quantitative variable）：量的変数の分析方法を説明する。また量的変数をグループに分けた場合は、どのように分けたかをその理由とともに記載する。

記載例

グラスゴー昏睡尺度（GCS）が 8 未満の患者は重度の外傷を受けていると考えられる。9 以上であれば脳外傷はそれよりは重症ではないことを示している。われわれは GCS におけるこの 2 つのカテゴリー（categories）と外傷後 12 時間以内の死亡との関連を調べた⁸⁰。

解説

研究者は、曝露因子、効果指標の修飾因子、および、交絡因子に関する量的変数の取得方法や解析方法に関して選択を行う。例えば、曝露に関する連続変数をグループ化して新たなカテゴリー変数（categorical variable）を作る場合がある（Box 4 を参照）。この場合、どのようにグループ化したかは、後の解析に重要な影響を与えるであろう^{81,82}。なぜ、そして、どのように量的データをグループ化したかについて、カテゴリーの数やカット・ポイント（cut-points）を、カテゴリー内での平均値や中央値などとともに説明することを推奨する。データを表で示す場合は常にケース人数、コントロール人数、リスク集団の人数、リスク人年数等を加えるべきである。表に示されているものが、効果の推定量やモデルの適合性の結果だけであってはならない。

研究者は全情報を保持するために曝露を連続値としてモデル化する場合がある。この方法を選択する場合、曝露とアウトカムとの関係を考慮する必要がある。両者の関連を単に線形的（linearity）な関係と仮定するのは誤りかもしれないので、非線形性の可能性を調べる必要がある。著者は解析において（対数変換[log transformation]や二次項[quadratic terms]やスプライン関数[spline function]などの）線形性以外のモデルについても検討したことにつ

いて言及すべきである。曝露とアウトカムが非線形関係であった場合、それにフィットするモデルは複数存在する⁸²⁻⁸⁴。また、主に関心のある量的変数の曝露については、連続値とグループ化の両方で提示するとより豊富な情報が得られるかもしれない。

最近の調査では、雑誌に掲載された疫学研究の内 3 分の 2 が量的な曝露変数を用いていた。50 報の論文中 42 報 (84%) はいくつかの順序カテゴリー (ordered categories) にグループ化されていたが、多くはそのグループ化の根拠を示していなかった。15 報の論文で連続値の曝露に線形モデルを用いていたが、そのうち線形性の確認をしていたのは 2 つだけであった。心理学分野における他の調査では、2 つのグループに分けた根拠が示されていたのは 110 報の論文中わずか 22 (20%) 報だけであった⁸⁵。

Box 4 – グループ化

連続変数のデータをグループ化 (カテゴリー変数化) することにはいくつかの理由がある⁸⁶。データを収集するときに、記憶に基づき数年にわたる曝露の状況を連続変数として厳密に測定しようとするよりも、順序変数を用いて測定したほうが良いかもしれない。カテゴリーはプレゼンテーションの際にも役に立つ。例えば全ての変数を同じような形式で発表したいとき、あるいは用量反応関係 (dose-response relationship) を示したいときなどにおいてである。

また、グループ化は、例えば、線形性の仮定を避けることができたりするなど、解析を単純化することができる。しかし、グループ化は情報を失うことにつながり、また、特に 2 値化した場合^{82,85,88}には統計学的な検出力を減じうる⁸⁷。もし、連続変数の交絡因子をグループ化したとすると、交絡による影響がいくらか未調整のまま残る残差交絡が起こりうる (Box 5 を参照)^{62,89}。カテゴリー数を多くすることは検出力の低下を防ぐことや残差交絡を減らすことができるので、大規模研究においては特に適切である。小規模研究においては対象者数が限定されるため、少数のグループ化しか適用することができないであろう。

研究者は診断および予後の観点、臨床の観点、あるいは、統計学的な観点で適切な通常用いられる値に基づいたカット・ポイントにおいてグループ化を行っても良い。また、等分位を用いて各グループ同じ数になるようなグループ化を選択しても良い⁹⁰。一方、等分位の外側のグループよりも極端に外側のグループを設け、中間のグループを大きくすることで、アウトカムとの関連において、より多くの知見を得られる場合がある⁹¹。ケース・コントロール研究においては、母集団を反映している必要があるため、コントロール群の分布をもとにグループ化することが推奨される。カット・ポイントがいくつかの候補の中から事後に選択されたものであるのなら、読者にその情報を与えるべきである。とりわけ、P 値が最も小さくなるようにカット・ポイントが選択されているなら関連の強さは過大に評価されてしまうであろう⁸¹。

グループ化された変数を解析に用いる際には、それらがもともと有している連続的な性質を認識することが重要である。例えば、順序グループ間におけるリスクの傾向性などの

検討である。一般的な手法としては、順位カテゴリーにグループ化された変数を連続変数としてモデル化することである。もし10年おきの年齢グループなどの形でグループ化が等間隔でなされていれば、グループの値における線形性は実際の値の分布における線形性とほぼ同じになるであろう。しかし、グループ化が等間隔でない場合は、その限りではない。Il'yasova ら⁹²は、連続変数とカテゴリー変数との双方を用いた推定を標準誤差と共に記載するよう推奨している。なぜなら、これらの情報は、メタ・アナリシスに貢献するだけでなく、用量反応関係に関わる本質的で重要な情報を提供するからである。このとき、それぞれの変数を用いた解析はお互いに情報を提供し、また、その2つの解析は、いずれも何らかの仮定をおいている。著者は、カテゴリーの順序を無視し、参照群と比較した各群において別個に推定値（およびP値）を検討してしまうことがよくある。このようなことを記述すること自体は良いのかもしれないが、グループ間でのリスクの傾向性を見誤るものでもある。仮に傾向性が確認されるようであれば、その傾きの信頼区間により、観察された傾向性の強さを検討することが可能である。

12.統計・分析方法 (statistical methods) :

12(a)交絡因子の調整に用いた方法を含め、すべての統計学的方法を示す。

記載例

2群間における年齢あるいは性による交絡の存在を考慮し、調整済み相対リスクはMantel-Haenszel法を用いて計算された。調整済み相対リスクの95%信頼区間はGreenlandおよびRobinsらの方法に基づいた分散を用いて計算した⁹³。

解説

一般に、統計解析方法には1つの正しい方法があるのではなく、むしろ同じ問題を解決しうる異なる仮定に基づくいくつかの方法がある。どのような場合でも、研究者は少なくとも研究計画書における当初の研究目的に関しては解析手法を決めておくべきである。予定していた解析の代わりに、あるいは、予定していた解析に加え、さらなる解析が必要となることがよくある。それらは収集したデータを検討する中で必要性が生じるのであろう。研究を発表する際には、著者はデータの検討によって特別な解析がなされたかどうかを読者に示すべきである。たとえ事前に予定していた解析とその後の探索的な解析との違いが明確なものでなくとも、著者は特別な解析を行った理由を明確に示すべきである。

もし比較される群同士が、いくつかの背景としての特徴において異なっている場合には、層別解析 (stratification) もしくは多変量回帰分析 (multivariable regression) によって可能性のある交絡の調整を行うべきである (Box 5を参照)⁹⁴。多くの場合、研究デザインによりどのようなタイプの回帰モデルが選択されるのかが決定される。例えば、Cox 比例ハザード回帰 (Cox proportional hazard regression) はコホート研究においてよく用いられるのに対し⁹⁵、ロジスティック回帰 (logistic regression) はケース・コントロール研究においてよく選択

される手法である^{96,97}。解析を行う者は、最終モデルの結果を示すだけでなく、変数選択の経緯を余すところなく記述すべきである^{98,99}。もし、モデルの比較によって、最終モデルで投入された交絡因子を絞り込んだのであれば、その過程を記述する必要がある。データの解析において、1つないし2つの共変量が明らかに大きな交絡要因として作用していたのであれば、そのことについて述べることは読者にとって有用である。他の統計手法、例えば欠損値の補完の方法 (imputation procedures)、データ変換の方法、寄与危険 (attributable risk) の計算の方法なども記述する必要がある。一般的ではない手法または新たな手法を用いた場合には文献の引用を行い、また、使用した統計ソフトを報告すべきである。基本的な方針として、「知識のある読者が元データにアクセスすることができた場合、その研究結果を検証することが可能となるように十分詳細に」統計解析手法を記述することが望ましいと考える¹⁰⁰。

実証的な研究において、169報の論文のうち、93報 (55%) しか、統計モデルへどのように連続変数やカテゴリ変数を投入し、交絡要因の調整を行なったのか明瞭に述べていなかった¹⁰¹。また別の研究では、その統計解析手法において交絡要因の調整を行ったと述べている67報の論文のほとんどにおいて交絡要因の選択の方法が不明確であった⁴。

Box 5 – 交絡

交絡とは、文字どおり効果が混乱 (confusion) している状態をさす。研究においては、曝露要因と疾患リスクとの間に関連があるように見えたり、ないように見えたりするであろう。しかし、関連があるように見えたり、ないように見えたりすることは、実際には、疾患の発生を予測する他の因子で、かつ、曝露にも関連している因子に影響され生じている場合もある。このような他の因子を交絡因子という。それゆえ、交絡は曝露要因と疾患リスクとの真の因果関係を誤った評価へと導いてしまう。例えば、高血圧の中年女性は経口避妊薬を内服していることが少ない傾向にあるとすれば、経口避妊薬を使用している中年女性と使用していない中年女性との間で心血管疾患の発生頻度の単純な比較は、経口避妊薬が心疾患を予防するという誤った印象を与えてしまうであろう。

研究者は、潜在的な交絡因子に関して事前に検討すべきである。事前に詳細な情報を集める必要のある交絡因子を特定することにより、研究デザインの際の参考となる情報をもたらし、また、適切なデータ収集が可能となる。対象者の限定やマッチングなどの手段を用いてもよい。上記の例では、高血圧という交絡因子を有していない中年女性に対象を限定してもよいかもしれない。必ずしも望ましいものであるとは限らないが、血圧によるマッチングによる解決も可能である (Box 2 を参照)。解析の段階においては、研究者は交絡の影響を減らすために層別解析、または、多変量解析を用いても良い。層別解析は交絡因子によってデータを層にわけ (血圧で層別化するなど)、それぞれの層で関連の推定を行い、すべての層において重み付け平均を行って統合された関連の推定を行うものである。多変量解析は、同様の結果を得ることが可能であり、しかも、多数の交絡変数を同時に考

慮することができる。多変量解析は、非常に柔軟性に富む手法であるが、曝露と疾患の関連において数学的な仮定がさらに必要となる。

交絡因子を考慮に入れることは、観察研究にとって不可欠なことであるが、読者は交絡因子を調整した解析が検討している関連の「因果関係」の有無を明らかにしていると考えべきではない。結果は残差交絡（不十分な交絡調整によっておこる交絡）¹⁰²、ランダム・サンプリングの誤差、あるいは、選択バイアスや情報バイアスによって歪んだままであるかもしれない（Box 3 を参照）。

12(b)サブグループと交互作用 (inter action) の検証に用いたすべての方法を示す。

記載例

3 つの生活習慣に関連するリスク因子に対する性別の影響は Rothman の方法に基づき生物学的な交互作用のテストによって検討した。すなわち、新たに、性別と 2 値の関心のあるリスク因子の有無の組み合わせによる 4 つのカテゴリー変数を作成した (a^-b^- 、 a^-b^+ 、 a^+b^- 、および a^+b^+ ; a^- および b^- は曝露要因がないことを示す)。相対リスクは年齢による調整を行った後、各カテゴリーにおいて算出した。交互作用の影響は元の絶対効果 (absolute effects) からの加算分として定義し、交互作用により生じた超過相対リスク (relative excess risk from interaction: RERI) は以下のように計算した。

$$RERI = RR(a^+b^+) - RR(a^-b^+) - RR(a^+b^-) - 1$$

$RR(a^+b^+)$ は両方の要因に曝露された人の相対リスクを指し、 $RR(a^-b^+)$ は参照カテゴリー ($RR = 1.0$) として用いられている。95%信頼区間は Hosmer と Lemeshow の方法によって計算した。RERI = 0 であれば交互作用による影響はないことを意味する¹⁰³。

解説

後述の 17 項において詳細な議論を行っている通り、研究対象集団のうちサブグループに限定して行った解析の意味や価値について多くの論議がなされている^{4,104}。それにも関わらずサブグループ解析はよく行われている⁴。読者は前もってどのようなサブグループ解析が計画されていたか、またデータの解析の際にどのサブグループ解析を追加したかを知る必要がある。また、集団間における影響や関連が異なっているかどうかを調べるために用いた手法を説明することが重要である（詳細は項目 17 を参照）。

交互作用は、ある因子が別の因子の効果を変化させてしまうような場合において関わってくるものである（それゆえ「効果作用の修飾 (effect modification)」ともいわれる）。2 つの因子の複合的な影響は 2 通りの方法で特徴付けられる。1 つ目は、加法性、すなわちリスク差の点、2 つ目は、乗法性、すなわち相対リスクの点である (Box 8 を参照)。多くの著者あるいは多くの読者は、交互作用の分析方法に関して独自の好みがある。しかし、曝露要

因の複合的な影響が、単独での影響とどの程度異なっているのかについては誰しも興味を持つであろう。コンセンサスとしては、加法性、つまり絶対リスクを用いることが公衆衛生および臨床決断分析には、より適切であるとされている¹⁰⁵。どのような視点からの分析を行うにしても、上記の記載例のように読者に明瞭に提示するべきである¹⁰³。それぞれの曝露要因の単独での影響に加え、複合的な影響を、曝露がなかった場合との比較で同時に並べて示すことがもっとも有用である。交互作用に関する例については後述の17項を、また異なった指標における計算法についてはBox 8を参照のこと。

12(c)欠損データ (missing data) をどのように扱ったかを説明する。

記載例

欠損データの解析方法には、ランダムな欠損であるという仮定 (missing at random: MAR) をおいた。STATA で連鎖方程式による多重補完法 (multivariate imputation by chained equations: MICE) を行った。それぞれ適切に欠損補完された 10 組のデータを多重ロジスティック回帰分析を用いて解析した。10 組それぞれのデータから得られた推定値を用いて、Rubin の方法に基づき 1 つの要約推定値とその標準誤差を求めた¹⁰⁶。

解説

欠損データは観察研究において非常によくある問題である。研究参加者によって投函された質問票は必ずしも完全に記載されているとは限らないし、参加者は全てのフォローアップのための診察に参加するわけではないし、また、日常のデータソースや臨床データベースもしばしば不完全なものである。この問題は非常にありふれていてかつ重要なものであるにもかかわらず、欠損データの問題について詳細な記述をしている論文はほとんどない^{5,107}。研究者は欠損データの問題を解決するためにいくつかの手法を用いることがあるであろう。Box 6 にさまざまな手法の強みと限界を記した。著者はそれぞれの対象とする変数 (曝露変数、アウトカム変数、交絡要因) における欠損の数と、解析の各段階における欠損の数を報告することが望ましい。著者は、可能であれば欠損の理由を示すべきであり、研究における対象者の流れ (flow) を示す際には、欠損データによって何人の対象者が除かれたのかを示すべきである (項目 13 も参照)。欠損データを考慮した解析では、著者はその解析手法 (多重補完法[multiple imputation]など) の仕組みおよびその前提となる仮定 (ランダムな欠損[missing at random: MAR]など : Box 6 を参照) について示すべきである。

Box 6 – 欠損データ : その問題と可能な解決法

欠損データを扱う通常的手法としては、特定の解析に必要な変数が完全にそろっている個人のみ限定するという方法がある。このような欠損のない「complete case」を対象とした解析は多くの状況においてはバイアスを生じさせないが、バイアスを生じさせる可能性もあり、また、常に効率が悪い (inefficient)¹⁰⁸。全体の対象者と欠損データを有する対象

者との特質が異なっている場合にはバイアスが生じる。また、サンプルサイズが減少することにより効率の悪い解析となる。

繰り返し測定するような状況下で最後に観察された値を使用する場合（LOCF 法）、もし疾患がおこりそうな患者が選択的に脱落するとすれば、傾向性をゆがめうるものとなる¹⁰⁹。一部の対象者に欠損がある交絡因子について、その交絡因子を欠損の有無を示すカテゴリー変数を作成して解析で用いた場合、残差交絡の問題を増加させうる¹⁰⁷。欠損値の補完、すなわち、それぞれの欠損値を仮定または推定された値で置き換えるものであるが、知りたい関連を強調または減弱したりするかもしれないし、以下に述べる洗練された方法を用いなければ標準誤差が過小評価される可能性がある。

Rubin はデータが欠損する確率をモデル化するという観点から、欠損データの問題を定型化した^{108,110}。もし、欠損が起こる確率が他のいかなる観察された変数に依存しない場合、その欠損データは完全にランダムな欠損データ（missing completely at random: MCAR）とされる。もし、欠損が起こる確率が欠損データの実際の値と独立である場合には、ランダムな欠損データ（MAR）とされる。例えば、乳幼児は肺活量のデータがより多く欠損する傾向にあるが、その欠損の確率は、年齢を考慮すれば、観察されていない肺機能検査の値とは関連しない。そのとき欠損となっている肺機能検査の値は、年齢を考慮に入れたモデルにおいては、MAR とされる。もし、利用可能な他のデータを考慮した後も、欠損した変数にその欠損の確率が依存する場合には、ランダムではない欠損データ（missing not at random: MNAR）とされる。データが MNAR であるとき、なぜ欠損データが生じたのかについて明らかな説明が必要である。

MAR のデータを扱う手法は、以下の 3 つの方法に大別される^{108,111}。すなわち、尤度に基づいた手法（likelihood-based approaches）¹¹²、重みづけに基づいた手法（weighted estimation）¹¹³、そして、多重補完法である^{111,114}。これら 3 つの手法のうち、多重補完法がもっとも一般的に使用されており、とりわけ複数の変数が欠損値を有する場合において柔軟性に富む¹¹⁵。これらの手法のいかなるものを使用した場合でも、その結果を、特定の解析に必要な変数が完全にそろっている個人のみに限定した解析結果と比較し、主要な違いについて議論を行うべきである。欠損データの解析を行う上で前提となる仮定が妥当かどうかについては一般的には検証不可能である。特に、その欠損データが MNAR よりもむしろ MAR であることを証明することは不可能である。したがって、これらの解析は感度分析の考えでとらえるべきである（項目 12e、17 を参照）。

12(d)コホート研究：該当する場合は、脱落例（loss to follow-up）をどのように扱ったかを説明する。

記載例

積極的にフォローアップを実施した治療プログラムにおいては、脱落した症例および 1 年間追跡可能であった症例のベースラインでの CD4 細胞数の差はなかった（中央値 115 個

/ μL および 123 個/ μL)。一方、積極的にフォローアップを実施しなかった治療プログラムにおいては、脱落した症例は追跡可能であった症例と比較して CD4 が著しく少なかった（中央値 64 個/ μL および 123 個/ μL)。(中略) 積極的にフォローアップを実施しなかった治療プログラムの対象者は、それ以降の解析から除外した¹¹⁶。

解説

コホート研究の解析には、生命表法 (life table methods)、対象とする疾患の発生までの時間や追跡できた期間の観察人時 (person-time) に基づいた他の方法を用いる。各々の観察期間の終了時において疾患の発生が認められなかった対象者においては、追跡できた期間はアウトカムが発生する確率とは関連しないという仮定をおいている。例えば、追跡する期間の終了を、特定の年齢になった時点や、固定された日時とした場合があげられる。脱落は対象者がその日時の以前に研究から脱落することによって起こる。もし脱落が曝露群の対象者において、あるいは疾患発生のリスクが高い対象者において（「情報のある打ち切り [informative censoring]」として）選択的に起こるとすれば、研究の妥当性に大きく影響するであろう。上記の記載例においては、積極的なフォローアップを行わなかった治療プログラムにおける脱落例は、追跡可能例に比して CD4 ヘルパーT 細胞の数が少なかったため、より高い死亡リスクを有しているといえる¹¹⁶。

研究終了まで追跡可能であった対象者と脱落した対象者を区別することは重要である。残念ながら、統計ソフトは通常これら 2 つの状況の区別を行わない。すなわち、どちらの状況においても、各々の観察期間の終了時に、追跡期間で自動的に「打ち切り (censoring)」として処理される。それゆえ、研究者は脱落した対象者をどのように扱うかを、理想的には研究計画の段階で決めておくことが必要である。脱落した患者の数が少ないときには、研究者は不完全な追跡しかできなかった対象者を除外するか、あるいは、研究終了時もしくは脱落した日時のいずれかの時点において健康な状態で脱落したとして扱ってもよい。われわれは、著者が脱落した対象者の数、および、用いた打ち切りの基準を述べることを推奨する。

12(d)ケース・コントロール研究：該当する場合は、ケースとコントロールのマッチングをどのように行ったかを説明する。

記載例

われわれは、認知症の患者と、心血管系のリスク因子、脳梗塞の有無、頸動脈疾患の有無、静脈から動脈へのシャント (venous to arterial circulation shunt) の有無によってマッチングを行ったコントロールとを比較するために McNemar 検定、対応のある t 検定、および、条件付きロジスティック回帰分析を用いた¹¹⁷。

解説

個別マッチングを行ったケース・コントロール研究においては、マッチングを無視した

粗オッズ比 (crude analysis of the odds ratio) は、通常関連がない方向へのバイアスを含んだ推定値となることが多い (Box 2 を参照)。それゆえ、マッチングを考慮した解析は必要である。マッチングを考慮した解析は直感的には、層別解析として理解しうるものである。すなわち、それぞれのケースは対応するコントロールと共に一つの層として扱われる。この解析は、マッチングによりケースと類似したコントロール集団を作成したにもかかわらずケースがコントロールよりも曝露をより多く受けているかどうか、ということについて検討することを基礎としている。研究者は「マッチングされた」2×2 表において Mantel-Haenszel 法を用いた層別解析を行うことができる。もっとも単純な場合、オッズ比は、曝露の状態が一致しなかったペアの比となる。もし、マッチングが性や年齢のような普遍的な性質を持つ変数でなされるのであれば、その場合の解析は個人つまり人と人のマッチングを考慮する必要はなく、性や年齢のカテゴリーを用いた単純な解析で十分である⁵⁰。しかし、近隣集団、血縁関係、友人関係など他の変数をマッチングに用いた場合には、それぞれのマッチングされたケースとコントロールの組はそれぞれ独自の層として考慮すべきである。

個別マッチングの研究で、もっとも広く使用されている解析方法は条件付きロジスティック回帰分析 (conditional logistic regression) であり、この方法では各ケースとそのコントロールがまとめて扱われる。条件付きロジスティック回帰分析は他の変数での調整を要する場合に加え、対応するコントロールの数がそれぞれのケースの間で異なるときに必要となる。マッチングのデザインが解析において適切に考慮されているかを読者が判断することができるように、著者はデータの解析において用いた統計解析手法の詳細を記述することが望ましい。もし、マッチングを考慮しても推定値にわずかな影響しか与えないならば、マッチングを考慮しない解析を示すことを選択してもよい。

12(d)横断研究：該当する場合は、サンプリング方法を考慮した分析法について記述する。

記載例

標準誤差 (SE) は、複雑なサンプリングのデザインに基づく推定量のサンプル誤差を推定するためテーラー展開 (Taylor expansion method) を用いて算出した。(中略) 拡張期血圧に関するデザイン効果 (design effect) は男性で 1.9、女性で 1.8 であり、また収縮期血圧に関するデザイン効果は男性で 1.9、女性で 2.0 であった¹¹⁸。

解説

たいていの横断研究はもともとの母集団から対象者を選ぶためにあらかじめ設定したサンプリング方法を用いている。しかし、そのサンプリング方法は、単純なランダム・サンプリングよりも複雑な場合がある。例えば、いくつかの段階や対象者のクラスター (例えば地域や村など) を用いたサンプリング方法である。集団のサイズに比例した確率で層別サンプリングを行うと、そのサブグループのある特性は正しく代表性を持つといえるだろ

う。そのような方法をとらない層別サンプリングは、特定の興味のあるサブグループを重点的にサンプリングしたいときには有益であることもある。

複雑なサンプルから得られた関連の指標の推定値は、単純なランダム・サンプリングから得られたものより精度が良かったり悪かったりする。標準誤差や、信頼区間などの精度の指標はデザイン効果—これはランダム・サンプリングの代わりに複雑なサンプリングの手法を使用した場合における、精度の差を表す指標であるが—を用いて調整されるべきである¹¹⁹。たいていの場合において複雑なサンプリング方法を用いると精度は低下し、デザイン効果は1より大きい値となる。

著者は、選択したサンプリング方法が得られた推定値の精度にどのように影響をあたえるのかを読者が理解できるように、複雑なサンプリング方法を調整するために用いた方法を明瞭に述べるのが望ましい。例えば、クラスター・サンプリング (cluster sampling) では、デザイン効果が記載されていれば、データ収集の容易さと精度の低下との間にある隠れたトレード・オフの関係が明確なものとなる。上記の記載例において、男性における1.9というデザイン効果は、単純なランダム・サンプリングを用いた場合に比べ、同等の精度での推定が得られるには、実際に1.9倍のサンプルサイズを必要とすることを意味する。

12(e)あらゆる感度分析 (sensitivity analysis) の方法を示す。

記載例

生存している患者 (15/437、3.4%) に比べ、死亡した患者 (38/148、25.7%) のデータ欠損の割合がかなり高かったため、(中略) 結果にバイアスが生じている可能性がある。そのため、われわれは感度分析を実施した。研究において経口避妊薬を使用した女性の割合は全体に当てはまると仮定し (死亡者の19.1%、生存している患者の11.4%)、2つの極端な状況を想定した。すなわち、欠損となった曝露された患者全員が第二世代のピルを使用していたとする状況、またはその全員が第三世代のピルを使用していたとする状況である¹²⁰。

解説

感度分析は他の解析方法や仮定に基づいても主要な結果に変わりがないかどうかを調べるのに有益である¹²¹。検討すべき事項としては、解析対象集団、曝露およびアウトカムの定義¹²²、調整に値する交絡要因、欠損データの扱い^{120,123}、曝露、疾患、その他の変数の不正確な測定もしくは一致しない測定に由来するバイアスや選択バイアス、量的変数の扱いなどにおける特定の解析手法の選択などが含まれる (項目11を参照)。バイアスや仮定の影響を同時にモデリングするために、洗練された手法の利用が増加している¹²⁴⁻¹²⁶。

1959年にCornfieldらは、交絡要因が非喫煙者にくらべ喫煙者に最低9倍必要であるという理由から、喫煙の肺がんに対する9倍という相対リスクが、他のいかなる想定可能な交絡要因によってもたらされた可能性は限りなく低いという有名な結果を示した¹²⁷。この解析はそのような要因があるという可能性を除外したのではなく、必要とされる要因の割合を特定したのである。最近でも、小児白血病と電力線の近郊に居住することとの関連を説

明しうる適当な交絡要因を確認するために、同様の手法が用いられた¹²⁸。より一般的には、感度分析は、関連を歪めてしまうような交絡、選択バイアス、または、情報バイアスの程度を確認するために用いられうる。おそらくあまり認識されていないことであるが、感度分析は、曝露とアウトカムとの間に弱い関連しか示されないとき、もしくは関連が示されない場合に使用することも重要である。関連がないとする方向に導くような交絡要因や他のバイアスが存在する可能性があるためである。

結果 (RESULTS)

結果のセクションでは、研究参加者の募集や参加集団の記述に始まり、主要な結果と副次的な解析結果に至るまで、何が明らかになったかという事実を記述する。これらの記述においては、著者の視点や意見を反映した解釈や推論を述べるべきではない。

13.参加者 (participants) :

13(a)研究の各段階における人数を示す (例: 潜在的な適格 (potentially eligible) 者数、適格性が調査された数、適格と確認された数、研究に組入れられた数、フォローアップを完了した数、および、分析された数)。

記載例

105 の独立営業のバーと居酒屋のうち、13 店舗が廃業し、9 店舗はレストラン内にあり、83 店舗が営業を続けていた。22 店舗では、店舗のオーナーに 6 回以上電話でのコンタクトを試みたが電話に出なかった。36 店舗のオーナーが研究参加を拒否した。(中略) 研究に参加した 25 のバーと居酒屋では、124 名のバーテンダーを雇用しており、うち 67 名のバーテンダーが少なくとも 1 回は日中のシフトで勤務していた。日中のシフトのバーテンダーのうち 54 名 (81%) がベースライン時のインタビューと肺活量測定を行った。そのうち 53 名 (98%) がフォローアップを完了した¹²⁹。

解説

研究参加者の登録方法に関わる詳細な情報は、いくつかの点から重要である。研究には、その結果を適用する標的集団 (target population) とは異なった参加者がよく含まれている。このことにより、有病割合や発生の推定が標的集団を反映できない可能性がある。例えば、性的行動についての郵送調査への参加に同意した人は、拒否した人と比較して教会へ行くことが少なく、性に関する考え方が保守的ではなく、最初の性交年齢が若く、喫煙者や飲酒者が多いという傾向があった¹³⁰。これらの違いにより、郵送調査では性的なことがらに関するリベラリズムや活動を過大評価する可能性がある。このような応答バイアス (Box 3 を参照) は、研究の適格基準に合致する人と、実際に研究に含まれる人において曝露と疾患の関連が異なる場合に、結果を歪めてしまうことがある。他の例としては、若くして結婚することと子供の白血病発症との関連が複数のケース・コントロール研究でみられてい

るが^{131,132}、これは若年女性の参加率がケース群とコントロール群とで異なることによる。健康な子供をもつ若年女性は、健康でない子供をもつ若年女性よりも研究に参加しない傾向にあった¹³³。低い研究参加率は、必ずしも研究の妥当性を損なうものではないが、参加者の情報と参加を拒否した理由の開示は必要である。また、研究参加、参加の承諾、および、フォローアップ割合の定義に関して、普遍的に合意されたものはないため、読者は著者がそれらの割合をどのように算出したか知る必要がある¹³⁴。

理想的には、研究者はその研究の標的集団を決定するところから解析に含めるデータを決定するところまでの、各段階での対象者の人数を記述すべきである。研究のタイプによって異なるが、潜在的な適格者数 (potentially eligible)、適格性が調査された数 (the number assessed for eligibility)、適格と確認された数 (the number found to be eligible)、研究に組み入れられた数 (the number included in the study)、フォローアップを完了した数 (the number followed up)、および、分析された数 (the number included in the analysis) が含まれる。研究参加者のサンプリングが、上記の記載例のように2つ以上の段階でサンプリング (multistage sampling) された場合には、それぞれの段階における情報が必要である。ケース・コントロール研究では、ケース群とコントロール群の参加人数のフローを別々に作成することを勧める¹³⁵。コントロール群は、例えば、入院患者と地域住民など、複数のソース (source) から選ばれることがよくある。そのような場合には、コントロール群のタイプごとに人数を計算することを推奨する。Olsonらは、ランダム番号ダイアル法や他の方法によりサンプリングされたコントロール群をどのように報告するかについての有用なガイドラインを提案している¹³⁶。

疫学、公衆衛生学および医学に関する10雑誌を調査したところ、研究参加に関する情報の記述があった論文は、ケース・コントロール研究では107報中47報 (44%)、同様にコホート研究では154報中49報 (32%)、横断研究では86報中51報 (59%) となっていた¹³⁷。疫学研究において、研究参加者の情報が不十分であることや欠如していることについては、他に2つの文献調査の論文で報告されている。最後に、ここ10年で疫学研究に参加する人が減少しているというエビデンスが報告されており^{137,138}、透明性の高い報告の必要性が強調されている¹³⁹。

13(b)研究の各段階での研究不参加の理由を記述する。

記載例

研究への不参加の最も大きな理由は、健康状態が悪かったこと、インタビューの前に死亡してしまったこと (ケース群 30%、コントロール群 <1%)、研究参加への返事がなかったこと (ケース群 2%、コントロール群 21%)、および、他の理由 (コンサルタントや医師による拒否、英語が話せない、精神的障害) (ケース群 7%、コントロール群 5%) であった¹⁴⁰。

解説

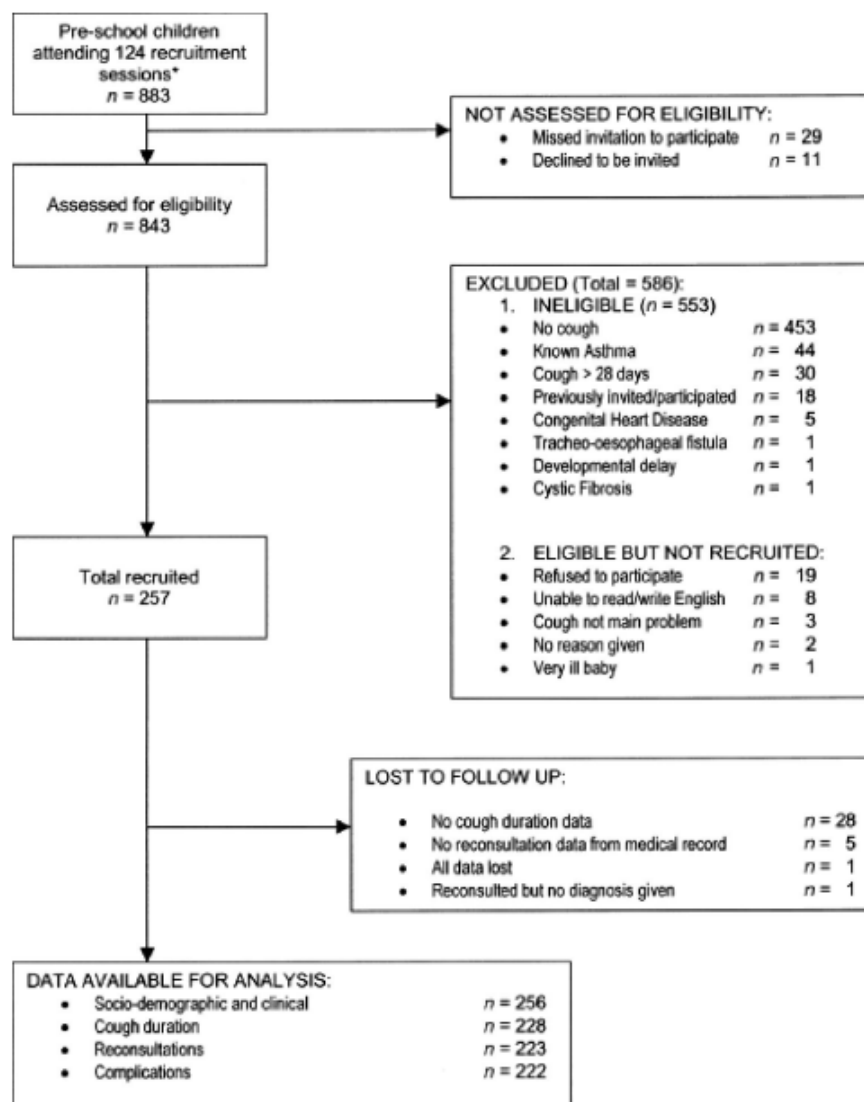
対象者が研究に参加することが出来なくなった理由や統計解析から除外された理由を説

明することは、研究参加者が研究の標的集団を代表しているか、および、バイアスが生じている可能性がないかということを読者が判断するときに役立つ。例えば、横断的な健康調査の場合、健康状態と関連していないと考えられる研究不参加の理由（例：住所が間違っており、研究参加依頼の案内が届かなかった）は、推定の精度に影響するかもしれないが、それによってバイアスが生じるとは考えにくい。反対に、多くの参加者が病気や主観的な健康状態を理由に選択的に研究から脱落した場合、ある集団の有病割合や健康な人の割合を過大推定または過小推定してしまう可能性がある。

13(c)フローチャートによる記載を考慮する。

記載例

FIGURE 1 参照¹⁴¹。



*Denominator data missing from one session at which at least 3 attended with cough, 2 recruited

FIGURE 1. Child recruitment and follow-up flow diagram. From Hay et al.¹⁴¹

解説

上記の記載例のようなよく出来たフローチャートを用いることにより、文章では長い記述となってしまう情報を迅速に明確に伝えることができる¹⁴²。フローチャートには主要アウトカムの発生数といった主要な結果を記載するとよい。複雑な観察研究では、とくにフローチャートによる記載を推奨するが、しかし、特定の決まったフローチャートのフォーマットを推奨するものではない。

14.記述的データ (descriptive data) :

14(a) 参加者の特徴 (例 : 人口統計学的、臨床的、社会学的特徴) ならびに曝露および潜在的な交絡因子の情報を示す。

記載例

TABLE 2 参照¹⁴³。

	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*	853 (59%)	272 (53%)	293 (58%)
Excessive†	355 (24%)	110 (22%)	91 (18%)

HCV, Hepatitis C virus.
*Males <60 g ethanol/day, females <30 g ethanol/day.
†Males >60 g ethanol/day, females >30 g ethanol/day.
Table adapted from Osella et al.¹⁴³

解説

結果の一般化可能性を読者が判断するためには、研究の対象集団と曝露の情報が必要になる。潜在的な交絡因子に関する情報—交絡因子を測定したかどうか、その測定法も含めて—は、研究の妥当性の判断に影響する。したがって、連続変数の要約については、それぞれの群別に、その平均値と標準偏差を用いることを推奨し、非対照的な分布をもつデータ (特にケース群でよくみられる) であれば中央値やパーセンタイル値 (25%、75%など) を用いることを推奨する。数の少ない順序カテゴリー (重症度 I から IV のように) 変数は、連続変数のように扱うべきではなく、各カテゴリー内の構成人数や割合を記述することが望ましい (Box 4 も参照)。上記の記載例のように、比較する群がある場合には、群ごとに研究対象者の特徴の記述や人数を記述するべきである。

標準誤差や信頼区間といった推定に関する指標 (inferential measures) は参加者の特徴の記述では用いるべきではない。また、記述的なデータの表では、検定結果を記載することは避けるべきである。また、わずかな差しかみられない交絡因子であってもアウトカムに

大きな影響を与える重要な因子となる可能性があるため、交絡因子として調整する因子を P 値により選択することは適切ではない^{144,145}。

コホート研究では、曝露がどのように他の特徴や潜在的な交絡因子と関連しているかを記述することは有用である。著者はこれらの情報を、参加者の特徴を示す表の中で曝露カテゴリー（2つ以上）ごとに欄を設けて示すことができる。これによりこれらのカテゴリー間での交絡因子の分布の違いを判断することができる。

ケース・コントロール研究における潜在的な交絡因子は、ケース群とコントロール群の比較では判断できない。コントロール群の参加者が源泉（ソース）集団を代表していても、多くの点でケース群の参加者とは異なっていると考えられる。例えば、経口避妊薬と心筋梗塞に関する研究では、梗塞症を持つ若年女性はコントロール群と比較して、高コレステロール血症、喫煙、および、家族歴などのリスク因子をたいてい多く持っている¹⁴⁶。しかし、このようなリスク因子のケース群とコントロール群での分布の違いは、リスク因子によって経口避妊薬の処方左右されることはないため（リスク因子はイベントの後に測定されたものである）、経口避妊薬との関連の評価には影響しない（Box 5 を参照）。ケース・コントロール研究では（コホート研究でも同様だが）、ケース群の源泉（ソース）集団を検討することによって、曝露群と非曝露群において潜在的な交絡因子の分布が同等かどうか検討できる。もしコントロール群の人数が十分で、源泉（ソース）集団を代表しているならば、コントロール群内の曝露群と非曝露群での潜在的な交絡因子の分布を比較することによりこれらの検討が可能となる^{121,147}。

14(b)それぞれの変数について、データが欠損した参加者数を記載する。

記載例

TABLE 3 参照¹⁴¹。

	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

解説

欠損データは、バイアスを生じさせたり、結果の一般化可能性に影響を与える可能性があるため、著者は、曝露、交絡因子、および、参加者の重要な特徴ごとの欠損データ数を記すべきである（項目 12c と Box 6 を参照）。コホート研究では、不完全なフォローアップ（incorporate follow-up）は結果にバイアスをもたらす可能性もあるため、著者は追跡できなかった人数およびその理由を示すべきである（項目 12d、13 を参照）¹⁴⁸。欠損データを示

す場合には、表や図を利用することを推奨する。

14(c)コホート研究:追跡期間を平均および合計で要約する。

記載例

4366人年の追跡期間（中央値 5.4年、最大 8.3年）中、265例が認知症と診断され、うち202例のアルツハイマー病を含んでいた¹⁴⁹。

解説

読者は、アウトカム・データを得るために要した追跡期間とその範囲を知る必要がある。著者は平均値と中央値のいずれか、あるいは、両者を用いて追跡期間を示すことができる。追跡期間の平均値から、読者は研究対象者の追跡期間（人年[person-years]）を算出できる（追跡期間の平均値に研究参加者数を乗じる）。さらに著者は、追跡期間の幅を読者に示すため、追跡期間の最小値、最大値またはパーセンタイル値を示すことができる。また追跡可能期間総人年に対する実際の追跡期間総人年の割合を示すことができる¹⁴⁸。こういった情報は全て曝露カテゴリー（2つ以上）ごとに示すべきである。がんに関する医学雑誌の132論文（ほとんどがコホート研究）のほぼ半数では、追跡期間の要約が一切示されていなかった³⁷。

15.アウトカム・データ（outcome data）：

コホート研究：経時的なアウトカム事象の発生数や要約指標（summary measures）を報告する。

記載例

TABLE 4 参照¹⁵⁰。

TABLE 4. Rates of HIV-1 Seroconversion by Selected Socio-Demographic 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

ケース・コントロール研究：各曝露カテゴリ内の対象者数または要約指標を示す。

記載例

TABLE 5 参照¹⁵¹。

TABLE 5. 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

横断研究：アウトカム事象の発生数または要約指標を示す。

記載例

TABLE 6 参照¹⁵²。

TABLE 6. Prevalence of Current Asthma and Diagnosed Hay Fever by Average *Alternaria Alternata* Antigen Level in the Household

Categorized <i>Alternaria</i> 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

解説

曝露（リスク因子）とアウトカムの関連性について言及する前に、著者はそれに関係する記述的データを報告しなければならない。記述的データと同じ表の中で関連の程度を示すことも可能かつ有用である（項目 14a を参照）。複数のイベントをアウトカムとして扱うコホート研究の場合、それぞれのアウトカムにおけるイベント数を報告する。追跡期間に

おける人年あたりのイベント発生率の報告も検討する。もしイベント発生リスクが追跡期間中に変化するようであれば、追跡期間の中で適切な期間ごとのイベント発生率を報告するか、Kaplan-Meier 法により計算した生命表またはプロット図 (Kaplan-Meier life table or plot) を作成する。特にイベントの発生率 (割合) が約 30% よりも低い場合には、0% から累積罹患を示すプロットの方が、100% から下がるタイプのものよりも好ましい¹⁵³。これらの情報は、関心のある曝露カテゴリーごとに示す。コホート研究で時間依存的なアウトカム (血圧のような量的疾病マーカー) を調査している場合には、表や図に適切な要約指標 (例: 平均値、標準偏差) を示す。

横断研究でも、アウトカム・イベントの割合や要約指標に関して同様の情報を記述することを勧める。ケース・コントロール研究では、中心となるのはケース群とコントロール群別の曝露の頻度または要約指標の報告である¹⁵⁴。全ての研究デザインにおいて、連続変数のアウトカム変数や曝露変数を、カテゴリー変数に変換して表に一覧することは、そのような形で分析するか否かに関わらず有用である。

16. おもな結果 (main results) :

16(a)調整前 (unadjust) の推定値、該当する場合は交絡因子での調整後の推定値、および、それらの精度 (例: 95%信頼区間) を記述する。どの交絡因子がなぜ調整されたかを明確にする。

記載例 1

Mantel-Heanszel 法を用いた層別解析によって、われわれは最初に潜在的な交絡因子として以下の変数を考えた: (中略)。われわれが最終的にロジスティック回帰モデルに含めた変数は、以下であった (中略)。Mantel-Heanszel 調整後に、オッズ比は 10% 変化した¹⁵⁵。

記載例 2

TABLE 7 参照¹⁵⁶。

TABLE 7. 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

多くの状況において、著者は、調整前の、あるいは、最低限の調整済みの解析結果、および、全体を調整した解析結果を提示する。われわれは、例えば、ケース群とコントロール群における曝露と非曝露の人数のような主要なデータとともに調整前の解析も記載することを推奨する。これらを提示することにより、読者は関連の測定背景にあるデータを理解することができる（項目 15 も参照）。調整による解析では、解析に用いた人数を報告すべきである。なぜならその数は共変量の欠測値によって異なるかもしれないからである（項目 12c も参照）。推定値は信頼区間とともに記載されるべきである。

読者は調整前の関連の指標と潜在的な交絡因子で調整後の結果を比較して、どのくらい、そしてどの方向に結果が変化したかを判断することができる。読者は「調整された」結果が、検討している関連の因果関係の有無を示したことに等しいと考えるかもしれないが、調整された結果には、ランダム・サンプリング・エラー、選択バイアス、情報バイアス、あるいは、残差交絡が含まれていないとは必ずしもいえない（Box 5 を参照）。このように、結果の妥当性は、重要な交絡因子、測定の正確さ、および、統計モデルの適切かつ詳細な知識に依存するので、調整された結果を解釈する際には慎重さが求められる（項目 20 を参照）^{157,158}。

著者らは潜在的な交絡因子で検討したものすべてと統計モデルへの変数の除外あるいは投入する基準を説明するべきである。変数の除外あるいは投入の決定は、因果関係に関わる背景の知識、あるいは、明白な前提により導かれるべきである。不適切な決定はバイアスを招くかもしれない。例えば、曝露と疾患の因果関係の中間にある変数で調整した場合、その解析が中間変数によって結果にどのくらいの影響があるのかを評価する目的ではない限り、バイアスを招くことになる。もし変数をモデルに投入する基準が推定値の変化に基づいているならば、どういった変化があれば十分にその変数の投入を正当化することができるかとみなしたのかを報告することは重要である。もし変数減少法（モデルに既に入っている変数を落とす[backward deletion]を逐次的に行う）か、変数増加法（新しい変数を取り入れる[forward inclusion]を逐次的に行なう）が交絡因子を選ぶ方法として用いられたのであれば、その過程や、交絡ではないという帰無仮説を棄却する有意水準を説明するべきである。もっとも、われわれは、交絡因子を選ぶのに統計学的有意の検定のみに基づいて行うことをアドバイスしているわけではないことを述べておく^{147,159,160}。

最近の疫学的研究の報告の質についての研究では、大部分の論文において信頼区間が報告されていることがわかった⁴。しかし、交絡因子の変数の選択理由について説明している著者はほとんどいない^{4,5}。

16(b)連続変数がカテゴリー化されているときはカテゴリーの境界値（category boundaries）を報告する。

記載例

TABLE 8 参照¹⁶¹。

TABLE 8. 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

解説

連続変数のデータをカテゴリー化することは、分析にいくつかの重要な影響をもたらす（Box 4 を参照）、結果の提示の仕方にも影響する。表においては、アウトカムはそれぞれの曝露のカテゴリーに対して与えられるべきである。適切に分けられた各群（ケース群とコントロール群など）別に、例えば、標的イベントが発生する可能性（at risk）がある人の数、あるいは、標的イベントが発生する可能性がある人時（person-time）などを示すべきである。カテゴリーを詳細に述べることは、研究の比較やメタ・アナリシスの助けになるであろう。もしデータが伝統的なカット・ポイント、例えば Body Mass Index（BMI）の基準値などでグループ分けされていれば¹⁶²、カテゴリーの最高値と最低値を除いては、グループの境界線（値の範囲）は容易に得ることができる。もし四分位によるカテゴリーを使用するならば、カテゴリーの境界はデータから推測することができない。少なくとも、著者はカテゴリーの境界値を報告するべきである。さらに、データの範囲、カテゴリー内の平均または中央値を報告するとよいであろう。

16(c)意味のある（relevant）場合は、相対リスク（relative risk）を意義ある期間の絶対リスク（absolute risk）に換算することを考慮する。

記載例

ホルモン置換療法（hormone replacement therapy: HRT）の「10 年間」の使用は、エストロゲン単独使用では 1000 人の使用者あたり 5 人（95%信頼区間: 3-7）の乳がんの発生を増加させ、エストロゲン・プロゲステロン併用療法では 1000 人の使用者あたり 19 人（95%信頼区間: 18-20）のがんの発生を増加させる¹⁶³。

解説

曝露と疾患の関連を検討した研究結果は、通常、リスク、率、あるいは、オッズの比のように相対的な指標として報告される（Box 8 を参照）。相対的な指標は、曝露と疾患の関連の強さをとらえる。もし相対リスクが 1 よりもかけ離れていれば、その関連が交絡によるものであるとは考えにくい^{164,165}。相対効果（relative effect）や相対関連（relative association）は、絶対的な評価尺度よりも個々の研究や対象集団に関わらず、結果が一貫する傾向にあるが、特定の状況においては相対的な指標を用いることが不適切な場合もある。例えば、北アイルランド、フランス、アメリカ、ドイツに住む男性において、冠動脈心疾患を発症

するリスクはこれらの国々において重大な差があるが^{166,167}、古典的な心血管系リスク因子の影響を相対リスクとして評価すると国間ではほぼ同じである。対照的に、心血管系疾患での死亡のリスク因子としての高血圧の影響を調べた研究データからは、発生率の差 (rate difference) が一定であるという仮定の方が、発生率の比 (rate ratio) が一定であるという仮定よりも適切であるという結果が得られている¹⁶⁸。

広く使われている統計モデルである、ロジスティック回帰¹⁶⁹ や比例ハザード (Cox) 回帰¹⁷⁰ は、比の指標に基づいている。これらのモデルでは、比の効果の指標 (ratio effect measures) が一定でない場合には容易に交互作用を識別可能である。それにもかかわらず、交互作用からの相対超過リスク (RERI) (項目 12c と Box 8 を参照) といった指標、つまり効果の指標として差の指標を用いた場合の交互作用を評価する指標は、比の指標に基づくモデルを用いても推測可能である。

多くの状況において、曝露に関連する絶対リスクは相対リスクよりも関心をもたれる。例えば、研究の目的がある薬剤の副作用を知ることであれば、ある単位時間 (日、週、年) の使用によって増加するケース数を知りたいと思うだろう。記載例では 10 年間のホルモン置換療法を行った 1000 人の女性あたりの乳がんの発症数を知ることができる¹⁶³。寄与危険 (attributable risk) や人口寄与割合 (population attributable fraction) のような指標は、もしその曝露を除去すればどのくらいの疾患を予防できるかを推測するのに役立つであろう。それらはできれば統計的な不確実性を示す指標 (この記載例のように例えば信頼区間) とともに示されるべきである。著者は、ここで前提とされているリスク因子と疾患の因果関係を含む、強い前提を意識しておくべきである (Box 7 を参照)¹⁷¹。専門用語の定義や意味が複数存在しかつこれらが複雑に絡み合っていることから、著者らは寄与危険度を計算するのに用いた方法を詳細に報告するべきである。理想的には用いた公式を示すべきである¹⁷²。

主要医学雑誌に掲載された 222 報の論文の抄録についての最近の研究ではランダム化試験の抄録の 62% で、各比較グループでのイベント発生割合が示されたうえで、比の指標が示されていたが、コホート研究の抄録ではそのような報告は 21% に過ぎなかった¹⁷³。Medline での 1966 年から 1997 年までの検索 (Free Text Search) では、タイトルか抄録に、寄与危険度を挙げたのは 619 本だったのに対し、相対リスクやオッズ比を用いたのは 18,955 報であった (1 対 31)¹⁷⁴。

Box 7 – 関連、効果、および、影響の指標

観察研究は、その対象集団におけるある健康問題の大きさと分布を記述する目的のためだけに行われることがあるかもしれない。そのために、特定の時点にある疾患を持つ人の数 (有病割合)、あるいは、ある期間内にある疾患を発症する人の数 (罹患数) を検討するのであろう。罹患数は、その疾患を発症していく人々の割合 (累積罹患)、あるいは追跡人年あたりの率 (罹患率) で示される。特定の発生率を示す専門用語には、死亡率 (mortality rate)、出生率 (birth rate)、発病率 (attack rate)、あるいは致命率 (case fatality rate) などがある。

同様に有病数を示す専門用語には、一時的な有病割合 (point prevalence) や、期間有病割合 (period prevalence)、年間有病割合 (annual prevalence)、生涯有病割合 (lifetime prevalence) などがある³⁰。

観察研究は因果関係を検討する場合もある。この場合、焦点が置かれるのは、その研究で調べたいリスク因子へ曝露している群と曝露していない群との間で関心のあるイベントの発生割合、発生率、あるいは、有病割合の比較である。これらの研究では「相対リスク (relative risk)」を推定することがよくあるが、これはリスク比 (累積罹患の比[cumulative incidence ratio]) を指すこともあれば、率比 (罹患率の比) を指すこともある。ケース・コントロール研究では、コントロール群での曝露群と非曝露群の人数の割合が示される。結果は、ケース群とコントロール群の曝露オッズの比で示される。このオッズ比は、リスク比や率比を推定するが、ケース群とコントロール群のサンプリング方法に依存している (Box 1 を参照)^{175,176}。横断研究における有病割合の比 (あるいは有病率比[prevalence ratio]) や有病オッズ比は、状況によっては有用であろう¹⁷⁷。

結果を相対的および絶対的な指標の両方で示すことは多くの場合有用である。例えば、男性英国人医師の研究で 50 年以上追跡した肺癌による死亡率は、喫煙者では 10 万人年当たり 249 人に対して、非喫煙者 10 万人年当たり 17 人で、率比は 14.6 (= 249 / 17) であった¹⁷⁸。冠動脈心疾患については、同じくそれぞれ対応する率が 10 万人年当たり 1001 人と 619 人であり、率比は 1.61 (= 1001 / 619) であった。喫煙の死亡に対する影響は、心疾患よりも肺癌に対して強かった。しかしながら、喫煙の影響を絶対指標でとらえた時、印象は異なる。肺癌の 10 万人年当たり発生率の差は 232 (= 249 - 17) であり、心疾患は 382 (= 1001 - 619) である。したがって、喫煙する医師においては、喫煙は肺癌より心疾患での死亡を多く引き起こすと考えられるのである。

ある対象集団において曝露を取り除くことによって予防できると考えられる疾患はどのくらいと見積もることができるだろうか？喫煙については全世界的な推定値が報告されており、ある研究では全ての肺癌の 91%、心疾患の 40%、2000 年男性の全ての死亡の 33% が喫煙によるものとされている¹⁷⁹。人口寄与割合はある特定の曝露が原因となったケースの割合、と一般的に定義される。しかし他にもいくつかの概念があり (確立した専門用語はない)、他の因子を調整するという間違っただけの方法がとられる場合もある^{172,180}。この指標を報告する意義とは何であろうか？相対的な指標は、ある関連の強さを強調し、病因論においてはもっとも有用である。もし曝露についての因果関係が述べられ、関連が効果 (effect) として解釈されるのであれば、相対リスクの推定値は、適切な絶対リスクの指標に変換すべきである。なぜなら公衆衛生政策において可能性のある影響を推定するために絶対リスクの方が適切であると考えられるからである。(項目 16c を参照)¹⁸¹。しかし、著者はこの状況においての強い前提に注意を払う必要がある¹⁷¹。つまり、ある特定の状況において、どのような概念と方法が適切かどうかを決定する必要がある。

17. 他の解析（other analyses）：その他に行われたすべての分析（例：サブグループと交互作用の解析や感度分析）の結果を報告する。

記載例 1

TABLE 9 参照^{182,183}。

TABLE 9. 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)

*Modified from Vandembroucke et al.¹⁸² by Botto et al.¹⁸³

記載例 2

TABLE 10 参照¹⁸⁴。

TABLE 10. 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

解説

観察研究では、主要な解析に加えて、その他の（副次的な）解析がよく行われる。その他の解析では、特定のサブグループ解析、リスク因子間の交互作用の可能性、寄与危険の計算、あるいは、感度分析として、変数の別の定義を用いた解析に取り組むものであろう。

一般に、サブグループ解析の危険性と、解析における検定の多重性については議論のあるところである。われわれの見るところでは、全体の結果がわずかな効果しか示さなかったとき、あるいは全く効果を示さなかったとき、サブグループにおける特異的な関連あるいは効果指標の修飾（effect-measure modification）を探そうとする傾向が強く表れる。一方

で、特にある研究が各サブグループに十分なデータが得られるくらい十分大規模であるときには、全体の解析でみられた関連が、サブグループ間で一貫して（より望ましいのは、事前に特定したサブグループの間にも）みられるのかを検討することは有用である。副次的な解析に関わる第 2 の論点として、データ解析の途中で事後的に行われるサブグループ解析についての議論がある。それら事後的なサブグループは、重要な発見であるかもしれないが、偶然により生じたものかもしれない。全てのサブグループ解析が事前に設定した解析であったという情報を読者に与えることは不可能かつ不必要であると主張する人もいる⁹。それは、すでに先行研究で明らかになっていることが解析を行った時点でどの程度適切かを示すことにしかならないからである。われわれはどの解析が事前に計画されたものであり、どの解析が事前に計画されなかったものであるかを報告することを推奨する（項目 4、12b、20 を参照）。このことは読者に（検定の）多重性の可能性を判断することを可能にし、発見から検証（verification）または論駁（refutation）までのその研究の連続的な立場を考慮することを可能にする。

さらに、副次的な解析に関わる第 3 の論点として、リスク因子間の相乗効果や交互作用をどのように評価したかについての議論がある：効果が相加的（加法的）なのか、あるいは相乗的な効果の指標なのか、いずれが適切かは統計モデルへの当てはまりのよさから決められるべきなのか（項目 12b と Box 8 を参照）？最もよいアプローチは、各リスク因子単独での効果と、リスク因子が組み合わさった場合の同時効果とを同様に報告することである。—もし表で示すことが可能であれば上記の記載例 1 あるいは Martinelli¹⁸⁵らの研究のように—。このような表は、相加的または相乗的な交互作用について評価するための十分な情報を与えることができる（具体的な計算方法は Box 8 で示す）。単独の効果とリスク因子が組み合わさった場合の同時効果の信頼区間を示すことは、読者がデータの強さを判断する助けになるであろう。加えて、超過相対リスク（RERI）のような交互作用の指標に関する信頼区間は、交互作用の検定や効果の均一性の検定に関係している。よくある問題の一つは、著者が各サブグループにおける P 値の比較を用いることである。それは、効果の修飾に関して誤った主張を導くことである。例えば、あるカテゴリー（例：男性）において統計学的に有意な関連があり、他方のカテゴリー（例：女性）において関連がない場合、それ自体が効果の修飾のエビデンスをもたらすわけではない。同様に、それぞれのカテゴリーの信頼区間が重なるときには交互作用がないと推論するというように不適切に使用される。より妥当な推論は、サブグループにおいて関連の強さが異なるかどうか直接評価することにより達成される。

感度分析は統計解析の際になされる様々な選択の影響や、得られた結果が欠損値や可能性のあるバイアスに対して頑健かどうかを検討するために有用である（項目 12b を参照）。感度解析の結果をどこまで報告するかは判断が必要である。多くの感度分析が行われているときには、それら全てについて細かい結果を提示するのは現実的ではない。感度分析を行ったことと、その結果が主要な結果と一貫することを示せば、十分である場合もある。感度分析での検討がそもそも主要な関心事項であった場合や、あるいは、効果の推定が相

当に異なるのであれば、詳細に結果を報告するべきである^{59,186}。

Pocock らは、観察研究の論文 73 本のうち 43 本が、サブグループ解析を含んでいることを示した。その大部分でグループ間の差異を報告し主張したが、事前に計画した交互作用の検討結果を報告したのは 8 本のみであった（項目 12b を参照）⁴。

Box 8 – 交互作用（効果の修飾）：同時効果（joint effects）の解析

ある曝露の疾患のリスクとの関連が、その他の曝露の存在によって異なるとき、交互作用が存在する。交互作用を評価し報告するときの問題の一つは、ある曝露の効果を二つの方法で測定することができることである。すなわち、相対リスク（ここではリスク比[あるいはリスク比の意味でのオッズ比]）（あるいは率比）として、あるいは、リスク差（あるいは率差）としてである。相対リスクの使用は相乗モデルに、リスク差の使用は相加（加法）モデルに対応する^{187,188}。さらに、相乗モデルあるいは相加モデルから導き出される「統計的交互作用」と、相加モデルから導き出される「生物学的交互作用」とを区別する場合もある¹⁸⁹。しかし、相加モデルも相乗モデルも、ある特定の生物学的メカニズムを示すものではない。

モデルの選択に関らず、主な目的は、二つの曝露の同時効果が各曝露の単独の効果（他方の曝露が存在しない場合）とどのように異なるかを理解することである。The Human Genomic Epidemiology Network (HuGENet) は、異なるタイプの交互作用も評価できるように、単独の効果と同時効果の両方を明確に提示するよう提案している。この提案を説明するために、経口避妊薬と第 V 因子 Leiden 変異についての研究からのデータを使用する。この例は 17 項でも用いられている。経口避妊薬と第 V 因子 Leiden 変異はそれぞれ静脈血栓症のリスクを増加させる。それらの単独の効果と同時効果は 2×4 の表によって計算することができ（項目 17 の記載例 1 を参照）、そこではオッズ比 1 とは第 V 因子 Leiden 変異がなく経口避妊薬の内服もない女性のベースラインを意味する。

交互作用において難しい点は、ケース・コントロール研究などいくつかの研究デザインやロジスティック回帰モデルや Cox 回帰モデルなどいくつかの統計モデルでは、相対リスク（あるいは率比）を推定し、本質的には相乗モデルに導くことにある。これらの場合では、相対リスクは相加スケールに変換することができる。項 17 の記載例 1 では、単独のオッズ比は 3.7 と 6.9 であり、同時オッズ比は 34.7 である。これらのデータを相乗モデルで解析すると、同時オッズ比は 25.7 と予想される（ $= 3.7 \times 6.9$ ）。この観察された 34.7 という結合された効果は、相乗スケールで予想した値より 1.4 倍大きい（ $= 34.7 / 25.7$ ）。この量（1.4）を測るのが相乗交互作用のオッズ比である。これはロジスティック回帰モデルから推定した交互作用の回帰係数の対数値と等しい。相加モデルにおいては同時オッズ比は 9.6 と予想される（ $= 3.7 + 6.9 - 1$ ）。観察された同時効果は相加性から強く逸脱している：その差は 25.1 である（ $= 34.7 - 9.6$ ）。オッズ比が相対リスク（あるいは率比）と解釈されるとき、後者の量（25.1）は交互作用により生じた超過相対リスク（RERI）である¹⁹⁰。このことは、参照

値 (オッズ比 1 に等しい) がベースラインの静脈血栓症の発生率を示すとき、例えば、10,000 人年 (女性) あたり 1 人とすると、それが曝露単独あるいは複数の要因に同時に曝露することによって増加するというイメージしたときにより容易に理解できるだろう。

考察 (DISCUSSION)

考察では、研究の妥当性と意義という重要な点を述べる¹⁹¹。ある調査によると、考察では、不完全またはバイアスのかかった研究結果の評価、示唆、および、著者が発見したことへの誇張が多々みられることが指摘されている^{192,193}。著者が根拠のない推察や結果の過剰な解釈をすることによって読者を混乱させることは、考察を構造化することにより防ぐことができるであろう^{194,195}。例えば、*Annals of Internal Medicine* は¹⁹⁶、著者に以下のように考察を組み立てることを推奨している。1) 主要な発見の概要、2) 考えられるメカニズムの考察と説明、3) 関連のある先行研究の結果との比較、4) 研究の限界、5) 研究結果が実践や今後の研究に対して与える示唆をまとめる短い節。同じような推奨が他からもなされている^{191,194}。今後の研究への示唆と研究の限界は、相互に密接に関連づけられているべきである。研究者は、単に「さらなる研究が必要である」と漠然と書くのではなく、次の研究が彼らの研究を改善するための方向性を提示するべきである^{197,198}。われわれは著者に対し、考察を構造化し、適切な副見出しを用いることを推奨する。

18. 鍵となる結果 (key results) : 研究目的に関しての鍵となる結果を要約する。

記載例

われわれの仮説は、国内の少数民族であることは心血管疾患 (cardiovascular disease: CVD) の高いリスク因子と関連しているが、その関連は社会経済状態 (socio economic status: SES) によって十分に説明できる、ということであった。われわれの仮説を裏付けることはできなかった。年齢と SES で調整した後においても、白人 (white) 女性と黒人 (black) およびメキシコ系アメリカ人 (Mexican American) 女性の間には BMI、血圧、糖尿病、および、身体活動に高度に有意差が残った。さらに、SES によって CVD のリスク因子に大きな差があり、SES の低い白人女性では両方の国内の少数民族女性と同じように CVD のリスクが高かった¹⁹⁹。

解説

研究の主要な発見の概要で考察を始めている良い例である。概要は、読者に主要な発見を思いおこさせ、著者が述べる解釈と示唆が、研究結果によって支持されているかどうかを評価するために有用である。

19. 限界 (limitations) : 潜在的なバイアスまたは精度の問題を考慮して研究の限界を議論す

る。潜在的なバイアスの方向性と大きさを議論する。

記載例

カウンセリングの受診割合は肥満の増加とともに増加しているので、われわれの推定は真の受診割合を過大に見積もっているかもしれない。電話調査もまた、カウンセリングの真の受診割合を過大に見積もっている可能性がある。電話を持っていない人は電話を持っている人と肥満である程度は変わらないはずだが、電話を持っていない人は教育レベルが低い傾向にあり、それは低いカウンセリング受診割合と関連していた。研究において体重についての質問に回答しなかった人と同様に研究参加を拒否した人々による潜在的なバイアスもまた重要である。さらに、データは横断的に集められているため、カウンセリングの受診が体重を減少させようとする行動より先に起こっていることを前提に推論することはできない²⁰⁰。

解説

研究の限界を同定し考察することは、科学的な論文の本質的な部分である。結果に影響を与えるバイアスや交絡の原因を明らかにすることだけでなく、他のバイアスの相対的な重要性を、いかなる潜在的バイアスの方向性や大きさを含めて議論することが重要である（Box 3 と項目 9 を参照）。

著者は、結果の精度についても考察するべきである。研究結果の精度は、研究の規模（項目 10）、曝露・交絡・アウトカムの測定（項目 8）を含め、様々な研究の側面において生じる。真の曝露を正確に測定できないと、結果が関連性のない方向にバイアスがかかる可能性がある。すなわち、リスク因子の測定が不正確になるほどバイアスは大きくなる。これは、「希薄化 (attenuation)」^{201,202}、または最近では「因果関係希薄バイアス (regression dilution bias)」²⁰³ と呼ばれる。しかし、関連する複数のリスク因子の測定における不正確性がそれぞれ異なっていた場合、それらのリスク因子に関連する調整済みの相対リスクは、関連がない方向にバイアスがかかる場合もあるが、関連がある方向にバイアスがかかる場合もある²⁰⁴⁻²⁰⁶。

研究の限界を考察するとき、著者は自分の研究と他の研究とを妥当性、一般化可能性、および、精度という観点から比較するとよい。このようにすることによって、それぞれの研究は独立した推論や貢献をするのではなく、研究領域に貢献するものとしてとらえることが可能になる²⁰⁷。驚くべきことに、限界についての重要な考察は、時として報告を公表する際に省略されることがある。原著論文が *Lancet* に掲載された著者を対象とした調査により判ったことは、原著論文では報告されていなかった重要な研究の限界が、この調査の調査票においては報告されたことであった¹⁹²。

20. 解釈 (interpretation) : 目的、限界、解析の多重性 (multiplicity)、同様の研究で得られた結果、および、その他の関連するエビデンスを考慮し、慎重で総合的な結果の解釈を記載する。

記載例

心筋梗塞による死亡と第二世代の経口避妊薬の使用との関連についてはどのような説明であれ推測的でなければならない。両者の直接的な生物学的なメカニズムを示唆する論文は出版されておらず、他の疫学研究でもそれを示す結果は出版されていない。(中略) 絶対リスクの増加は非常に小さく、おそらく喫煙者に主に当てはまる。確証するエビデンスが欠如しており、また、相対的に少ない参加者に基づいた分析しか行われていないため、今後さらなるエビデンスの構築が必要である。これまでの結果で得られたエビデンスの強さに基づいて、実践を行ったとしてもそれによる改善を期待することはできない¹²⁰。

解説

考察の中心は研究結果の解釈である。過剰な解釈は、一般的に人間にありがちなことである。いくら客観的な評価を示そうとしたとしても、いくつかの点において査読者が解釈が行き過ぎていることを正当に指摘することがあるだろう。結果を解釈する場合、著者は研究が連続的な検証の繰り返しにおける発見であることや、バイアスを生じうる潜在的な源 (フォローアップの脱落や研究の不参加) について十分に考慮すべきである (項目 9、12、19 も参照)。考察では、交絡 (項目 16a)、適切な感度分析の結果、および、多重性の問題やサブグループ分析 (項目 17) について述べるべきである。著者は、測定されなかった変数あるいは交絡の不正確な測定による残差交絡を考慮すべきである。例えば、社会経済状態 (SES) は多くの健康アウトカムに関係しており、比較されているグループ間で異なることがよくある。SES (収入、教育あるいは職業) の測定のために用いられる変数は、他の定義されていない測定されていない因子の代替因子 (surrogates) であり、真の交絡要因はその代替因子の定義によっては測定誤差を避けられないであろう²⁰⁸。著者は、推定値の不確実性の実際の範囲に言及し、それが信頼区間に反映されている統計学的不確実性より大きいということを示すべきである。後者の統計学的不確実性は、研究デザイン、研究の実施、および、測定方法による不確実性を考慮に入れていない²⁰⁹。

因果関係に関する考察や結論の執筆には、1965 年に Bradford Hill によって提案された基準が参考になる¹⁶⁴。曝露との関連はどれくらい強いだろうか？それは疾病の発生に先行していただろうか？その関連は他の研究や他のセッティングにおける研究とも一貫しているだろうか？その関連をサポートするような実証的な研究 (実験的な研究や動物実験を含む) によるエビデンスはあるだろうか？曝露の推定上の効果はどれくらい特定のだろうか、また、用量反応関係があるだろうか？その関連は生物学上の妥当性があるだろうか？しかしながら、これらの基準は機械的に適用されるべきではない。例えば、2 または 3 より低い相対リスクは、無視されるべきであると主張する人もいる^{210,211}。しかし、これは Cornfield が主張する大きな相対危険度の強さに関する主張と対立する (項目 12b を参照)¹²⁷。確か

に相対リスクが 9 であった場合、因果関係がありそうだが、例えば 3 より低いからといって必ずしも因果関係がないとはいえない。例えば、子宮内への放射線投射後の小児白血病のリスク上昇は、小さいものであっても重要である。なぜならそれは医学的手続きの副作用に関連することであり、他に明らかな説明がないからである²¹²。さらに、放射線による発癌効果はよく確立されている。1 週間当たり 2~4 個の卵を食べることと卵巣がんのリスクが 2 倍になることとの関連は即時に信頼できるものではない。なぜなら食習慣は多くの生活習慣要因が SES と同様に関連しているからである²¹³。対照的に、異なるタイプの経口避妊薬間の血栓症のリスクの差に関する疫学的発見の真実性は、ランダム化クロスオーバー研究により発見された血液凝固の差によって非常に補強された²¹⁴。先行研究の異なるタイプの研究による外的なエビデンスからの考察は常に述べられるべきだが、特に小さなリスクの増加を報告する研究においては重要である。さらに著者は、著者自身の研究結果を同じような先行研究のコンテキストに位置づけ、新たな研究結果が先行研究のエビデンスにどのような影響を与えたのかを、系統的なレビューを引用しながら述べるべきである。

21. 一般化可能性：研究結果の一般化可能性（外的妥当性[external validity]）を議論する。

記載例

われわれの推定結果は、他の HIV-1-感染者へ対してどれくらい適用可能だろうか？ 予後モデルは、モデルを開発するために使用したデータ以外のデータに適用した場合、たいていその正確性が低下する傾向にあるため、この問題は重要である。われわれは、この問題を考えるうえで、結果の一般化可能性が高い簡略化したモデルを選択することによって対処した。予後モデル構築用に用いたデータベースには、異なるセッティングで治療を受けているヨーロッパと北アメリカの多くの国々の患者を含めた。幅広い背景をもつ患者から成り、男性と女性、10 代から中年まで含まれており、主要な曝露カテゴリーに対して代表性のある集団であった。ベースライン時の免疫不全の重症度は、全く測定できなかった場合から重症まで、ウイルスの濃度は検出不能から非常に高い範囲までを網羅していた²¹⁵。

解説

一般化可能性は、外的妥当性または適用可能性（applicability）とも呼ばれるが、研究の結果を他の状況に適用することができる程度を示すものである²¹⁶。外的妥当性という本質的なものがあるわけではない。外的妥当性は、明白に特定された状況に関してのみ意味をなす²¹⁷。年齢、性別、人種、疾病の重症度、および、合併症に関してその研究の対象者と異なる個人、グループあるいは集団に研究結果を適用することができるだろうか？ 曝露の性質やレベルは比較可能であろうか？ アウトカムの定義は他の状況や集団に対しても妥当であるだろうか？ 何年も前に縦断的な研究で収集されたデータは、今日においても意味があるだろうか？ ある国でのヘルス・サービス研究の結果は他の国々のヘルス・システムに対しても適用可能だろうか？

研究結果に外的妥当性があるかどうかという問題は、多くの場合、研究セッティング、

参加者の特性、検討した曝露、および、評価した結果に依存する判断に関する問題である。したがって、研究のセッティングと場所、適格基準、曝露とその測定方法、アウトカムの定義、および、参加者の募集期間とフォローアップ期間に関する十分な情報を著者により提供されることが重要である。さらに、研究不参加の程度や曝露していないがアウトカムが生じた対象者の割合を示すことも意味がある。絶対リスクや曝露の割合は、集団によって異なることがよくあるため、その情報を示すことは、他のセッティングおよび集団に結果を適用する場合有用な情報となる（Box 7 を参照）。

その他の情報（OTHER INFORMATION）

22. 研究の財源（funding）：研究の資金源、本研究における資金提供者の役割を示す。該当する場合には、現在の研究の元となる研究についても同様に示す。

解説

いくつかの雑誌では、資金や他の利害衝突が存在するか否かを明示することを著者に対して求めている^{100,218}。いくつかの調査では、資金源と研究の結論の間に強い関連性があることを示している²¹⁹⁻²²²。研究資金が営利組織によって提供されたランダム化試験において、試験薬（資金提供者と関係がある薬）が有効であるとする結論の試験は、効果量（effect size）で調節しても非常に多く（オッズ比 5.3）みられた²²³。他にも、タバコや電気通信産業が資金源となっている研究においても結論に影響がみられることが示されている²²⁴⁻²²⁷。スポンサーが政府あるいは非営利団体の場合においても、そのような影響の例がみられる。

著者と資金提供者は、研究デザイン²²⁸、曝露の選択^{228,229}、アウトカム²³⁰、統計解析方法²³¹、および、一部のアウトカム²³⁰や研究結果²³²の選択的な公表、といったことに対して影響を与えるような利害衝突（conflicts of interest）を有しているかもしれない。したがって、資金提供者の役割は詳細に記述されるべきである。その中には、資金提供者が直接の責任をとった部分など（例：デザイン、データ収集、分析、原稿の起草、出版の決定）が記述されるべきである¹⁰⁰。他にも、影響を与えるものとして、雇用者（学術研究者にとっての大学管理者[university administrators for academic researchers]、政府の研究者にとっての政府管理者、特に政治上の被指名人[government supervisors, especially political appointees, for government researchers]）、諮問委員会（advisory committees）、訴訟当事者（litigants）、および、特別利益団体（special interest groups）もあげられる。

結論

STROBE 声明は、疫学における観察研究を報告するために有用な推奨を提供することを目標としている。よい報告は、研究の長所および短所を明らかにし、研究結果の健全な解釈と応用を促進する。STROBE 声明はさらに観察研究の計画時にも活用され、論文の評価の際に査読者や編集者のガイドとなるだろう。

われわれがこの説明的な論文を作成した目的は、観察研究が透明性高く完全に報告されることの重要性について議論するため、チェックリストのそれぞれの項目の背景にある論理的根拠について説明するため、そして、出版された論文からわれわれがよい報告であると考えられる記載例を示すためである。われわれは、ここで示した資料が著者や編集者が STROBE を活用する際の一助となることを希望する。

われわれは、研究の報告における STROBE や他の推奨^{13,233,234}が、発展的なものであって、今後継続的な評価、補強、また必要ならば変更を必要とするべきものであることを強調しておく^{235,236}。例えば、平行群間のランダム化試験の報告のための CONSORT 声明は、1990年代中頃に最初に開発された²³⁷。その後、グループのメンバーは声明を改訂する必要性を調査するために定期的に会合を開いた。そして2001年に改訂版が出され²³³、更なる改訂が検討されている。同様に、この論文で示した原則や、STROBE のチェックリストは新しいエビデンスや批判的なコメントの蓄積とともに改訂するものである。STROBE ウェブサイト (<http://www.strobe-statement.org>) では、チェックリストの改善のための議論や提案のためのフォーラム、この説明的なドキュメント、および、疫学研究のよい報告のための情報を提供している。

いくつかの雑誌では、STROBE 声明に従って報告するよう著者に指示している（ウェブサイトの現時点でのリストを参照）。他の雑誌に対しても STROBE 声明を採用することを歓迎する。ウェブサイトを通じて連絡してほしい。STROBE の推奨について出版された論文も公表している。したがって、STROBE 声明は生物医学分野において広くアクセス可能である。

引用文献

1. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328:39–41.
2. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest*. 2001;51:8–11.
3. Scales CD Jr, Norris RD, Peterson BL, Preminger GM, Dahm P. Clinical research and statistical methods in the urology literature. *J Urol*. 2005;174:1374–9.
4. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. 2004;329:883.
5. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *American Journal of Epidemiology*. 2005;161:280–8.
6. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Epidemiology*. 2007;18:800–804.
7. Mihailovic A, Bell CM, Urbach DR. Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg*. 2005;48:148–51.
8. Rushton L. Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup Environ Med*. 2000;57:1–9.
9. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1:43–6.
10. Moonesinghe R, Khoury MJ, Janssens AC. Most published research findings are false-but a

- little replication goes a long way. *PLoS Medicine*. 2007;4:e28.
11. Jenicek M. *Clinical Case Reporting, Evidence-Based Medicine*. Oxford: Butterworth-Heinemann; 1999; p. 117.
 12. Vandembroucke JP. In defense of case reports and case series. *Annals of Internal Medicine*. 2001;134:330–4.
 13. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Annals of Internal Medicine*. 2003;138:40–4.
 14. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93:387–91.
 15. Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet*. 2006;38:3–5.
 16. Rodrigues L, Kirkwood BR. 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. *International Journal of Epidemiology*. 1990;19:205–13.
 17. Rothman KJ, Greenland S. Case-Control Studies. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 93–114.
 18. Forand SP. Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. *Environ Health*. 2004;3:7
 19. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *The New England Journal of Medicine*. 2000;342:1878–86.
 20. Gøtzsche PC, Harden A. Searching for non-randomised studies. Draft chapter 3. Cochrane Non-Randomised Studies Methods Group, 26 July 2002 (<http://www.cochrane.dk/nrsmg>).
 21. Lohse N, Hansen AB, Pedersen G, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Annals of Internal Medicine*. 2007;146:87–95.
 22. *American Journal of Epidemiology*. Instructions to authors. Available from: http://www.oxfordjournals.org/our_journals/aje/for_authors/general.html.
 23. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Annals of Internal Medicine*. 1990;113:69–76.
 24. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. 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*. *Canadian Medical Association Journal*. 1994;150:1611–5.
 25. Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *Journal of Information Science*. 1996;22:27–37.
 26. Viner RM, Cole TJ. Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *BMJ*. 2005;330:1354.
 27. McCauley J, Kern DE, Kolodner K, et al. The “battering syndrome”: prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Annals of Internal Medicine*. 1995;123:737–46.
 28. McEvoy SP, Stevenson MR, McCart AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ*. 2005;331:428.
 29. Vandembroucke JP. Prospective or retrospective: what's in a name? *BMJ*. 1991;302:249–50.
 30. Last JM. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press; 2000.
 31. Miettinen OS. *Theoretical Epidemiology: principles of occurrence research in medicine*. New York: Wiley; 1985.
 32. Rothman KJ, Greenland S. Types of Epidemiologic Studies. In: Rothman KJ, Greenland S, editors. *Modern Epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 74–5.
 33. MacMahon B, Trichopoulos D. *Epidemiology, principles and methods*. 2nd ed. Boston: Little, Brown and Co.; 1996. p. 81.
 34. Lilienfeld AM. *Foundations of Epidemiology*. New York: Oxford University Press; 1976.
 35. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, Miletich JP. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction,

- stroke, and venous thrombosis in apparently healthy men. *The New England Journal of Medicine*. 1995;332:912–7.
36. Goodman KJ, O'Rourke K, Day RS, et al. Dynamics of *Helicobacter pylori* infection in a US-Mexico cohort during the first two years of life. *International Journal of Epidemiology*. 2005;34:1348–55.
 37. Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511–8.
 38. Cerhan JR, Wallace RB, Folsom AR, Potter JD, Munger RG, Prineas RJ. Transfusion history and cancer risk in older women. *Annals of Internal Medicine*. 1993;119:8–15.
 39. Beane Freeman LE, Dennis LK, Lynch CF, Thorne PS, Just CL. Toenail arsenic content and cutaneous melanoma in Iowa. *American Journal of Epidemiology*. 2004;160:679–87.
 40. Canto JG, Allison JJ, Kiefe CI, et al. Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *The New England Journal of Medicine*. 2000;342:1094–100.
 41. Metzker-Cotter E, Kletter Y, Avidor B, et al. Long-term serological analysis and clinical follow-up of patients with cat scratch disease. *Clin Infect Dis*. 2003;37:1149–54.
 42. Johnson ES. 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*. 1990;32:250–4.
 43. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin* 1946;2:53.
 44. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chronic Dis*. 1986;39:495–504.
 45. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *American Journal of Epidemiology*. 1978;107:1–7.
 46. Hackam DG, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet*. 2006;367:413–8.
 47. Smeeth L, Cook C, Fombonne E, et al. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet*. 2004;364:963–9.
 48. Costanza MC. Matching. *Prev Med*. 1995;24:425–33.
 49. Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *American Journal of Epidemiology*. 2002;155:593–602.
 50. Rothman KJ, Greenland S. Matching. In: Rothman KJ, Greenland S, editors. *Modern Epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 147–61.
 51. Szklo MF, Nieto J. *Epidemiology, Beyond the Basics*. Sudbury, MA: Jones and Bartlett; 2000. p. 40–51.
 52. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med*. 1971;25:242–4.
 53. Gissler M, Hemminki E. The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. *Eur J Obstet Gynecol Reprod Biol*. 1996;69:73–5.
 54. Gefeller O, Pfahlberg A, Brenner H, Windeler J. An empirical investigation on matching in published case-control studies. *Eur J Epidemiol*. 1998;14:321–5.
 55. Artama M, Ritvanen A, Gissler M, Isojarvi J, Auvinen A. Congenital structural anomalies in offspring of women with epilepsy—a population-based cohort study in Finland. *International Journal of Epidemiology*. 2006;35:280–7.
 56. Ebrahim S. Cohorts, infants and children. *International Journal of Epidemiology*. 2004;33:1165–66.
 57. Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975–2004. *International Journal of Epidemiology*. 2004;33:1185–92.
 58. Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol*. 2005;58:560–7.
 59. Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L. Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology*.

- 2005;16:155–63.
60. Winkelmayr WC, Stampfer MJ, Willett WC, Curhan GC. Habitual caffeine intake and the risk of hypertension in women. *JAMA*. 2005;294:2330–5.
 61. Lukanova A, Soderberg S, Kaaks R, Jellum E, Stattin P. Serum adiponectin is not associated with risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2006;15:401–2.
 62. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med*. 1992;11:1747–58.
 63. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology*. 1997;8:429–34.
 64. Phillips MR, Yang G, Zhang Y, Wang L, Ji H, Zhou M. Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet*. 2002;360:1728–36.
 65. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, Hankinson SE. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology*. 2006;113:1081–6.
 66. Craig SL, Feinstein AR. Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol*. 1999;22:51–6.
 67. Rogler LH, Mroczek DK, Fellows M, Loftus ST. The neglect of response bias in mental health research. *J Nerv Ment Dis*. 2001;189:182–7.
 68. Murphy EA. *The logic of medicine*. Baltimore: Johns Hopkins University Press; 1976.
 69. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51–63.
 70. Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. *American Journal of Epidemiology*. 1997;146:429–38.
 71. Bloemenkamp KW, Rosendaal FR, Buller HR, Helmerhorst FM, Colly LP, Vandembroucke JP. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med*. 1999;159:65–70.
 72. Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: W.B. Saunders Company; 1985.
 73. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA. Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *The American Journal of Tropical Medicine and Hygiene*. 2003;68:519–26.
 74. Anoop S, Saravanan B, Joseph A, Cherian A, Jacob KS. 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*. 2004;89:325–9.
 75. Carlin JB, Doyle LW. Sample size. *J Paediatr Child Health*. 2002;38:300–4.
 76. Rigby AS, Vail A. Statistical methods in epidemiology. II: A commonsense approach to sample size estimation. *Disabil Rehabil*. 1998;20:405–10.
 77. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365:1348–53.
 78. Drescher K, Timm J, Jockel KH. The design of case-control studies: the effect of confounding on sample size requirements. *Stat Med*. 1990;9:765–76.
 79. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med*. 1998;17:1375–89.
 80. Linn S, Levi L, Grunau PD, Zaidise I, Zarka S. Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. *Annals of epidemiology*. 2007;17:142–7.
 81. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86:829–35.
 82. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25:127–41.
 83. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology*. 1995;6:450–4.
 84. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*. 1999;28:964–74.
 85. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of

- quantitative variables. *Psychol Methods*. 2002;7:19–40.
86. Altman DG. Categorizing continuous variables. In: Armitage P, Colton T, editors. *Encyclopedia of biostatistics*. 2nd ed. Chichester: John Wiley; 2005. p. 708–11.
 87. Cohen J. The cost of dichotomization. *Applied Psychological Measurement*. 1983;7:249–53.
 88. Zhao LP, Kolonel LN. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *American Journal of Epidemiology*. 1992;136:464–74.
 89. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 1968;24:295–313.
 90. Clayton D, Hills M. Models for dose-response (Chapter 25). *Statistical Models in Epidemiology*. Oxford: Oxford University Press; 1993. p. 249–60.
 91. Cox DR. Note on grouping. *J Am Stat Assoc*. 1957;52:543–7.
 92. Il'yasova D, Hertz-Picciotto I, Peters U, Berlin JA, Poole C. Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes Control*. 2005;16:383–8.
 93. Berglund A, Alfredsson L, Cassidy JD, Jensen I, Nygren A. The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. *J Clin Epidemiol*. 2000;53:1089–94.
 94. Slama R, Werwatz A. Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique* 2005;53(Spec No 2):2S65–80.
 95. Greenland S. Introduction to regression modelling (Chapter 21). In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 401–32.
 96. Thompson WD. Statistical analysis of case-control studies. *Epidemiol Rev*. 1994;16:33–50.
 97. Schlesselman JJ. Logistic regression for case-control studies (Chapter 8. 2). *Case-control studies Design, conduct, analysis*. New York, Oxford: Oxford University Press; 1982. p. 235–41.
 98. Clayton D, Hills M. Choice and interpretation of models (Chapter 27). *Statistical Models in Epidemiology*. Oxford: Oxford University Press; 1993. p. 271–81.
 99. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J*. 1983;286:1489–93.
 100. International Committee of Medical Journal Editors (ICMJE). Uniform requirements for manuscripts submitted to biomedical journals. Electronic version updated February 2006, available at <http://www.icmje.org>. *The New England Journal of Medicine* 1997;336:309–15.
 101. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *AnnInternMed*. 2002;136:122–6.
 102. Olsen J, Basso O. Re: Residual confounding. *American Journal of Epidemiology*. 1999;149:290.
 103. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarod K, Holmen J. Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? *Am J Kidney Dis*. 2006;47:396–405.
 104. Gotzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*. 2006;333:231–4.
 105. Szklo MF, Nieto J. Communicating Results of Epidemiologic Studies (Chapter 9). *Epidemiology, Beyond the Basics*. Sudbury, MA: Jones and Bartlett; 2000. p. 408–30.
 106. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. *BMJ*. 2006;332:521–5.
 107. Vach W, Blettner M. 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. *American Journal of Epidemiology*. 1991;134:895–907.
 108. Little RJ, Rubin DB. A taxonomy of missing-data methods (Chapter 1. 4.) (need to check). *Statistical Analysis with Missing Data*. New York: Wiley; 2002. p. 19–23.
 109. Ware JH. Interpreting incomplete data in studies of diet and weight loss. *The New England Journal of Medicine*. 2003;348:2136–7.
 110. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–92.
 111. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.

112. Lipsitz SR, Ibrahim JG, Chen MH, Peterson H. Non-ignorable missing covariates in generalized linear models. *Stat Med.* 1999;18:2435–48.
113. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med.* 1997;16:81–102.
114. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley; 1987.
115. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res.* 1999;8:17–36.
116. Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet.* 2006;367:817–24.
117. Purandare N, Burns A, Daly KJ, et al. Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. *BMJ.* 2006;332:1119–24.
118. Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J. Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens.* 2001;19:1717–25.
119. Lohr SL. *Design Effects (Chapter 7.5). Sampling: Design and Analysis.* Pacific Grove: Duxbury Press.; 1999.
120. Dunn NR, Arscott A, Thorogood M. 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.* 2001;63:65–9.
121. Rothman KJ, Greenland S. Basic Methods for Sensitivity Analysis and External Adjustment. In: Rothman KJ, Greenland S, editors. *Modern epidemiology.* 2nd ed: Lippincott Raven; 1998. p. 343–57.
122. Custer B, Longstreth WT, Jr., Phillips LE, Koepsell TD, Van Belle G. Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. *BMC Cancer.* 2006;6:152.
123. Wakefield MA, Chaloupka FJ, Kaufman NJ, Orleans CT, Barker DC, Ruel EE. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ.* 2000;321:333–7.
124. Greenland S. 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. *Journal of the American Statistical Association.* 2003;98:47–54.
125. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology.* 2003;14:451–8.
126. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology.* 2003;14:459–66.
127. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst.* 1959;22:173–203.
128. Langholz B. Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? *Bioelectromagnetics* 2001;Suppl 5:S19–31.
129. Eisner MD, Smith AK, Blanc PD. Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA.* 1998;280:1909–14.
130. Dunne MP, Martin NG, Bailey JM, et al. Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *International Journal of Epidemiology.* 1997;26:844–54.
131. Schuz J, Kaatsch P, Kaletsch U, Meinert R, Michaelis J. Association of childhood cancer with factors related to pregnancy and birth. *International Journal of Epidemiology.* 1999;28:631–9.
132. Cnattingius S, Zack M, Ekblom A, Gunnarskog J, Linet M, Adami HO. Prenatal and neonatal risk factors for childhood myeloid leukemia. *Cancer Epidemiol Biomarkers Prev.* 1995;4:441–5.
133. Schuz J. 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.* 2003;17:106–12.

134. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD. Response rates among control subjects in case-control studies. *Annals of epidemiology*. 1995;5:245–9.
135. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet*. 2002;359:431–4.
136. Olson SH, Voigt LF, Begg CB, Weiss NS. Reporting participation in case-control studies. *Epidemiology*. 2002;13:123–6.
137. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *American Journal of Epidemiology*. 2006;163:197–203.
138. Olson SH. Reported participation in case-control studies: changes over time. *American Journal of Epidemiology*. 2001;154:574–81.
139. Sandler DP. On revealing what we'd rather hide: the problem of describing study participation. *Epidemiology*. 2002;13:117
140. Hepworth SJ, Schoemaker MJ, Muir KR, Swerdlow AJ, van Tongeren MJ, McKinney PA. Mobile phone use and risk of glioma in adults: case-control study. *BMJ*. 2006;332:883–7.
141. Hay AD, Wilson A, Fahey T, Peters TJ. The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. *Fam Pract*. 2003;20:696–705.
142. Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA*. 2001;285:1996–9.
143. Osella AR, Misciagna G, Guerra VM, et al. 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. *International Journal of Epidemiology*. 2000;29:922–7.
144. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. *International Journal of Epidemiology*. 1978;7:373–5.
145. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *American Journal of Epidemiology*. 1993;138:923–36.
146. Tanis BC, van den Bosch MA, Kemmeren JM, et al. Oral contraceptives and the risk of myocardial infarction. *The New England Journal of Medicine*. 2001;345:1787–93.
147. Rothman KJ, Greenland S. Precision and Validity in Epidemiologic Studies. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 120–5.
148. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet*. 2002;359:1309–10.
149. Qiu C, Fratiglioni L, Karp A, Winblad B, Bellander T. Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. *Epidemiology*. 2004;15:687–94.
150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, Ruberantwari A, Wagner HU, Mulder DW. Incidence of HIV-1 infection in adults and socio-demographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. *International Journal of Epidemiology*. 1996;25:1077–82.
151. Mastrangelo G, Fedeli U, Fadda E, et al. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environmental health perspectives*. 2004;112:1188–92.
152. Salo PM, Arbes SJ Jr., Sever M, et al. Exposure to *Alternaria alternata* in US homes is associated with asthma symptoms. *The Journal of allergy and clinical immunology*. 2006;118:892–8.
153. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet*. 2002;359:1686–9.
154. Sasieni P. A note on the presentation of matched case-control data. *Stat Med*. 1992;11:617–20.
155. Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA. A nested case-control study of residential and personal magnetic field measures and miscarriages. *Epidemiology*. 2002;13:21–31.
156. Tiihonen J, Wallbeck K, Lonnqvist J, et al. 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*. 2006;333:224.
157. Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med*. 2004;66:868–75.
158. Smith GD, Phillips A. Declaring independence: why we should be cautious. [published

- erratum appears in *J Epidemiol Community Health* 1991 Mar;45(1):88]. *J Epidemiol Community Health*. 1990;44:257–8.
159. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *International Journal of Epidemiology*. 1980;9:361–7.
 160. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology*. 2001;12:313–20.
 161. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA. Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology*. 2007;18:120–9.
 162. World Health Organization. Body Mass Index (BMI). Accessed at: http://www.euro.who.int/nutrition/20030507_1.
 163. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet*. 2003;362:419–27.
 164. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58:295–300.
 165. Vineis P. Causality in epidemiology. *Soz Praventivmed*. 2003;48:80–7.
 166. Empana JP, Ducimetiere P, Arveiler D, et al. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur Heart J*. 2003;24:1903–11.
 167. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, Amouyel P. 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*. 1999;353:1547–57.
 168. Cambien F, Chretien JM, Ducimetiere P, Guize L, Richard JL. Is the relationship between blood pressure and cardiovascular risk dependent on body mass index? *American Journal of Epidemiology*. 1985;122:434–42.
 169. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *American journal of public health*. 1991;81:1630–5.
 170. Tibshirani R. A plain man's guide to the proportional hazards model. *Clin Invest Med*. 1982;5:63–8.
 171. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *American journal of public health*. 1998;88:15–9.
 172. Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res*. 2001;10:231–7.
 173. Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ*. 2006;333:1428
 174. Nakayama T, Zaman MM, Tanaka H. Reporting of attributable and relative risks, 1966–97. *Lancet*. 1998;351:1179.
 175. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst*. 1951;11:1269–75.
 176. Pearce N. What does the odds ratio estimate in a case-control study? *International Journal of Epidemiology*. 1993;22:1189–92.
 177. Rothman KJ, Greenland S. Measures of Disease Frequency. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 44–5.
 178. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. *BMJ*. 328:1529–33, 2004; discussion 33.
 179. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet*. 2003;362:847–52.
 180. Greenland S. Applications of Stratified Analysis Methods. In: Rothman KJ, Greenland S, editors. *Modern Epidemiology*. 2nd ed: Lippincott Raven; 1998. p. 295–7.
 181. Rose G. Sick individuals and sick populations. *International Journal of Epidemiology*. 2001;30:427–32 discussion 33–4.
 182. Vandenbroucke JP, Koster T, Briet E, Reitsma PH, Bertina RM, Rosendaal FR. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet*. 1994;344:1453–7.
 183. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction:

- the two-by-four table and beyond. *American Journal of Epidemiology*. 2001;153:1016–20.
184. Wei L, MacDonald TM, Walker BR. Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Annals of Internal Medicine*. 2004;141:764–70.
 185. Martinelli I, Taioli E, Battaglioli T, et al. Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med*. 2003;163:2771–4.
 186. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst*. 2005;97:1043–55.
 187. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *American Journal of Epidemiology*. 1980;112:467–70.
 188. Saracci R. Interaction and synergism. *American Journal of Epidemiology*. 1980;112:465–6.
 189. Rothman KJ. *Epidemiology. An introduction*. Oxford: Oxford University Press; 2002. p. 168–80.
 190. Rothman KJ. *Interactions Between Causes. Modern epidemiology*. Boston: Little Brown,; 1986. p. 311–26.
 191. Hess DR. How to write an effective discussion. *Respir Care*. 2004;49:1238–41.
 192. Horton R. The hidden research paper. *JAMA*. 2002;287:2775–8.
 193. Horton R. The rhetoric of research. *BMJ*. 1995;310:985–7.
 194. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ*. 1999;318:1224–5.
 195. Perneger TV, Hudelson PM. Writing a research article: advice to beginners. *Int J Qual Health Care*. 2004;16:191–2.
 196. *Annals of Internal Medicine*. Information for authors. [cited; Available from: http://www.annals.org/shared/author_info.shtml#manuscriptpreparation
 197. Maldonado G, Poole C. More research is needed. *Annals of epidemiology*. 1999;9:17–8.
 198. Phillips CV. The economics of ‘more research is needed’. *International Journal of Epidemiology*. 2001;30:771–6.
 199. Winkleby MA, Kraemer HC, Ahn DK, Varady AN. Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988–1994. *JAMA*. 1998;280:356–62.
 200. Galuska DA, Will JC, Serdula MK, Ford ES. Are health care professionals advising obese patients to lose weight? *JAMA*. 1999;282:1576–8.
 201. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15:72–101.
 202. Fuller WA, Hidiroglou MA. Regression estimates after correcting for attenuation. *J Am Stat Assoc*. 1978;73:99–104.
 203. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*. 1990;335:765–74.
 204. Phillips AN, Smith GD. How independent are “independent” effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol*. 1991;44:1223–31.
 205. Phillips AN, Smith GD. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med*. 1992;11:953–61.
 206. Greenland S. The effect of misclassification in the presence of covariates. *American Journal of Epidemiology*. 1980;112:564–9.
 207. Poole C, Peters U, Il'yasova D, Arab L. Commentary: This study failed? *International Journal of Epidemiology*. 2003;32:534–5.
 208. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology*. 1997;8:621–8.
 209. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990;1:421–9.
 210. Taubes G. Epidemiology faces its limits. *Science*. 1995;269:164–9.
 211. Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA*. 1999;281:841–4.
 212. Greenberg RS, Shuster JL Jr. Epidemiology of cancer in children. *Epidemiol Rev*.

- 1985;7:22–48.
213. Kushi LH, Mink PJ, Folsom AR, et al. Prospective study of diet and ovarian cancer. *American Journal of Epidemiology*. 1999;149:21–31.
 214. Kemmeren JM, Algra A, Meijers JC, et al. Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor VLeiden mutation: a randomized trial. *Blood*. 2004;103:927–33.
 215. Egger M, May M, Chene G, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet*. 2002;360:119–29.
 216. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull*. 1957;54:297–312.
 217. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Annals of Internal Medicine*. 1999;130:515–24.
 218. Krinsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics* 2001;7:205–18.
 219. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA*. 2003;289:454–65.
 220. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med*. 1986;1:155–8.
 221. Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *The New England Journal of Medicine*. 1998;338:101–6.
 222. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326:1167–70.
 223. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003;290:921–8.
 224. Barnes DE, Bero LA. 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*. 1996;21:515–42.
 225. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA*. 1998;279:1566–70.
 226. Glantz SA, Barnes DE, Bero L, Hanauer P, Slade J. Looking through a keyhole at the tobacco industry. The Brown and Williamson documents. *JAMA*. 1995;274:219–24.
 227. Huss A, Egger M, Hug K, Huwiler-Muntener K, Roosli M. Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environmental health perspectives*. 2007;115:1–4.
 228. Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *J Nerv Ment Dis*. 2002;190:583–92.
 229. Aspinall RL, Goodman NW. 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*. 1995;311:844–6.
 230. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291:2457–65.
 231. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326:1171–3.
 232. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews*. 2007;2. No. MR000005.
 233. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191–4.
 234. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008–12.
 235. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting

- randomized trials: explanation and elaboration. *Annals of Internal Medicine*. 2001;134:663–94.
236. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. *Consolidated Standards of Reporting Trials*. *JAMA*. 1998;279:1489–91.
237. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637–9.

日本語訳付録

表：STROBE 声明：観察研究の報告において記載すべきチェックリスト（上岡広晴,津谷喜一郎訳.「疫学における観察研究の報告の強化（STROBE 声明）：観察研究の報告に関するガイドライン」.中山健夫,津谷喜一郎編.2008.『臨床研究と疫学研究のための国際ルール集』ライフ・サイエンス出版. pp202-209.）から転載。

表 STROBE 声明：観察研究の報告において記載すべき項目のチェックリスト

	no	推奨	報告頁
タイトル・抄録 [title and abstract]	1	(a) タイトルまたは抄録のなかで、試験デザインを一般に用いられる用語で明示する。 (b) 抄録では、研究で行われたことと明らかにされたことについて、十分な情報を含み、かつバランスのよい要約を記載する。	
はじめに [introduction]			
背景 [background]/ 論拠 [rationale]	2	研究の科学的な背景と論拠を説明する。	
目的 [objective]	3	特定の仮説を含む目的を明記する。	
方法 [methods]			
研究デザイン [study design]	4	研究デザインの重要な要素を論文のはじめの [early] 部分で示す。	
セッティング [setting]	5	セッティング、実施場所のほか、基準となる日付については、登録、曝露 [exposure]、追跡、データ収集の期間を含めて明記する。	
参加者 [participant]	6	(a) ・コホート研究 [cohort study]：適格基準 [eligibility criteria]、参加者の母集団 [sources]、選定方法を明記する。追跡の方法についても記述する。 ・ケース・コントロール研究 [case-control study]：適格基準、参加者の母集団、ケース [case] の確定方法とコントロール [control] の選択方法を示す。ケースとコントロールの選択における論拠を示す。 ・横断研究 [cross-sectional study]：適格基準、参加者の母集団、選択方法を示す。 (b) ・コホート研究：マッチング研究 [matched study] の場合、マッチングの基準、曝露群 [exposed] と非曝露群 [unexposed] の各人数を記載する。 ・ケース・コントロール研究：マッチング研究 [matched study] の場合、マッチングの基準、ケースあたりのコントロールの人数を記載する。	
変数 [variable]	7	すべてのアウトカム、曝露、予測因子 [predictor]、潜在的交絡因子 [potential confounder]、潜在的な効果修飾因子 [effect modifier] を明確に定義する。該当する場合は、診断方法を示す。	
データ源 [data source]/ 測定方法	8*	関連する各因子に対して、データ源、測定・評価方法の詳細を示す。二つ以上の群がある場合は、測定方法の比較可能性 [comparability] を明記する。	
バイアス [bias]	9	潜在的なバイアス源に対応するためにとられた措置があればすべて示す。	
研究サイズ [study size]	10	研究サイズ [訳者注：観察対象者数] がどのように算出されたかを説明する。	
量的変数 [quantitative variable]	11	(a) 量的変数の分析方法を説明する。該当する場合は、どのグルーピング [grouping] がなぜ選ばれたかを記載する。	
統計・分析方法 [statistical method]	12	(a) 交絡因子の調整に用いた方法を含め、すべての統計学的方法を示す。 (b) サブグループと相互作用 [interaction] の検証に用いたすべての方法を示す。 (c) 欠損データ [missing data] をどのように扱ったかを説明する。 (d) ・コホート研究：該当する場合は、脱落例 [loss to follow-up] をどのように扱ったかを説明する。 ・ケース・コントロール研究：該当する場合は、ケースとコントロールのマッチングをどのように行ったかを説明する。 ・横断研究：該当する場合は、サンプリング方式 [sampling strategy] を考慮した分析法について記述する。 (e) あらゆる感度分析 [sensitivity analysis] の方法を示す。	

結果[result]	
参加者[participant]	13* (a) 研究の各段階における人数を示す(例:潜在的な適格[eligible]者数, 適格性が調査された数, 適格と確認された数, 研究に組入れられた数, フォローアップを完了した数, 分析された数)。 (b) 各段階での非参加者の理由を示す。 (c) フローチャートによる記載を考慮する。
記述的データ [descriptive data]	14* (a) 参加者の特徴(例:人口統計学的, 臨床的, 社会的特徴)と曝露や潜在的交絡因子の情報を示す。 (b) それぞれの変数について, データが欠損した参加者数を記載する。 (c) コホート研究: 追跡期間を平均および合計で要約する。
アウトカムデータ [Outcome data]	15* ・コホート研究: アウトカム事象の発生数や集約尺度[summary measure]の数値を経時的に示す。 ・ケース・コントロール研究: 各曝露カテゴリーの数, または曝露の集約尺度を示す。 ・横断研究: アウトカム事象の発生数または集約尺度を示す。
おもな結果[main result]	16 (a) 調整前[unadjust]の推定値と, 該当する場合は交絡因子での調整後の推定値, そしてそれらの精度(例: 95%信頼区間)を記述する。どの交絡因子が, なぜ調整されたかを明確にする。 (b) 連続変数[continuous variable]がカテゴリー化されているときは, カテゴリー境界[category boundary]を報告する。 (c) 意味のある[relevant]場合は, 相対リスク[relative risk]を, 意味をもつ期間の絶対リスク[absolute risk]に換算することを考慮する。
他の解析[other analysis]	17 その他に行われたすべての分析(例: サブグループと相互作用の解析や感度分析)の結果を報告する。
考察[discussion]	
鍵となる結果[key result]	18 研究目的についての鍵となる結果を要約する。
限界[limitation]	19 潜在的なバイアスや精度の問題を考慮して, 研究の限界を議論する。潜在的バイアスの方向性と大きさを議論する。
解釈[interpretation]	20 目的, 限界, 解析の多重性[multiplicity], 同様の研究で得られた結果やその他の関連するエビデンスを考慮し, 慎重で総合的な結果の解釈を記載する。
一般化可能性 [generalisability]	21 研究結果の一般化可能性(外的妥当性[external validity])を議論する。
その他の情報[other information]	
研究の財源[funding]	22 研究の資金源, 本研究における資金提供者[funder]の役割を示す。該当する場合には, 現在の研究の元となる研究[original study]についても同様に示す。

*ケース・コントロール研究では, ケースとコントロールに分けて記述する。コホート研究と横断研究において該当する場合には, 曝露群と非曝露群に分けて記述する。