

Guidelines

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration—a Korean translation

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): 국문판 설명문서

Jan P. Vandenbroucke¹, Erik von Elm^{2,3}, Douglas G. Altman⁴, Peter C. Gøtzsche⁵, Cynthia D. Mulrow⁶, Stuart J. Pocock⁷, Charles Poole⁸, James J. Schlesselman⁹, Matthias Egger^{2,10*} for the STROBE Initiative

¹ Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, ² Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland,

³ Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany,

⁴ Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, United Kingdom,

⁵ Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark, ⁶ University of Texas Health Science Center, San Antonio, United States of America,

⁷ Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom,

⁸ Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, United States of America,

⁹ Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, and

University of Pittsburgh Cancer Institute, Pittsburgh, United States of America,

¹⁰ Department of Social Medicine, University of Bristol, Bristol, United Kingdom

It is a Korean translation of the Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007 Oct 16;4(10):e297. doi: 10.1371/journal.pmed.0040297. PMID: 17941715; PMCID: PMC2020496. Translation was done under the

permission of the STROBE.

Translator: Soo Young Kim, Department of Family Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea

Scientific reviewer: Hyeon Chang Kim, Cardiovascular and Metabolic Diseases Etiology Research Center, Yonsei University College of Medicine, Seoul, Korea; Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Korea.

Reviewer for Korean: Yoon Joo Seo, Infolum, Seongnam, Korea

Copyright: 2007 Vandembroucke et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. In order to encourage dissemination of the STROBE Statement, this article is freely available on the Web site of PLoS Medicine, and will also be published and made freely available by Epidemiology and Annals of Internal Medicine. The authors jointly hold the copyright of this article. For details on further use, see STROBE Web site (<http://www.strobe-statement.org/>).

초록

의학 연구의 대부분은 관찰 연구이다. 관찰 연구의 보고는 종종 불충분한 품질을 보이기도 한다. 부실한 보고는 연구의 강점과 약점을 평가하고 연구 결과의 일반화 가능성을 평가하는 데 방해가 된다. 방법론 전문가, 연구자, 편집자 그룹은 경험적 근거와 이론적 고려 사항을 고려하여 관찰 연구 보고의 질을 개선하기 위한 역학 STROBE 권고안을 개발했다. STROBE statement은 논문의 제목, 초록, 서론, 방법, 결과 및 토론 부분에 대한 22개 항목의 체크리스트로 구성되어 있다. 18개 항목은 코호트연구, 환자

대조군 연구, 단면연구에 공통으로 적용되며, 4개 항목은 세 가지 연구설계 각각에 따라 다르다. STROBE statement는 저자에게 관찰 연구에 대한 보고를 개선하는 방법에 대한 지침을 제공하고 심사자, 편집자 및 독자가 연구를 비판적으로 평가하고 해석하는 데 도움이 된다. 이 설명 문서는 STROBE statement의 사용, 이해 및 보급을 향상 시키는 것이 목적이다. 각 체크리스트 항목의 의미와 근거가 제시되어 있다. 각 항목에 대해 하나 또는 여러 개의 출판된 예시 논문과 가능한 경우 관련 경험적 연구 및 방법론 문헌에 대한 내용이 참고사항으로 제공된다. 유용한 흐름도의 예도 포함되어 있다. 본 문서 및 관련 웹사이트(<http://www.strobe-statement.org/>)는 관찰 연구 보고를 개선하는 데 유용한 자료가 될 것이다.

서론

합리적인 의료 행위를 위해서는 질병의 원인과 발병 기전, 진단, 예후, 치료에 대한 지식이 필요하다.

무작위배정 임상시험은 치료법 및 기타 중재에 대한 귀중한 근거를 제공한다. 그러나 임상 또는 공중보건 지식의 상당수는 관찰 연구에서 나온 것이다[1]. 임상 전문 저널에 게재된 연구 논문 10편 중 약 9편이 관찰 연구이다[2,3].

STROBE statement

관찰 연구 보고는 조사의 강점과 약점을 평가할 수 있을 만큼 상세하지 않고, 명확하지 않은 경우가

많다[4,5]. 관찰 연구 보고를 개선하기 위해, 우리는 STROBE statement 체크리스트를 개발했다. 항목은

논문의 제목, 초록, 서론, 방법, 결과 및 토론 파트와 관련이 있다. STROBE statement는 최근 여러

저널에 발표되었다[6]. 우리의 목표는 관찰 연구에서 계획, 수행 및 발견한 내용을 명확하게 제시하는 것이다. 이 권고사항은 연구 기획 또는 실제 수행에 관한 내용은 아니며, 방법론을 지시하거나 발표 통일안을 강제하지 않는다.

STROBE는 노출과 건강 결과 사이의 연관성을 조사하는 관찰연구에 대한 일반적인 보고 권장사항을 제공한다. STROBE는 관찰 연구의 세 가지 주요 유형인 코호트, 환자 대조군, 단면연구를 다룬다. 저자는 이러한 연구 설계를 설명하기 위해 다양한 용어를 사용한다. 예를 들어, '추적 연구'와 '추적 연구'는 '코호트연구'의 동의어로, '유병률 연구'는 '단면연구'의 동의어로 사용되기도 한다. 하지만 일반적으로 코호트, 환자 대조군, 단면연구 용어를 가장 많이 사용하므로 이 용어를 선택했다. 안타깝게도 용어가 잘못 사용되거나[7] 부정확하게 사용되는 경우가 많다[8]. 박스 1에서는 세 가지 연구 설계의 특징을 설명한다.

박스 1. STROBE에서 다루는 주요 연구 디자인

코호트, 환자 대조군, 단면 설계는 특정 집단과 기간에 건강 관련 사건의 발생을 조사하는 다양한 접근 방식이다. 이러한 연구는 질병 또는 질병 관해, 장애, 합병증, 사망, 생존, 위험 요인 발생 등 다양한 유형의 건강 관련 사건을 다룰 수 있다.

코호트연구에서는 연구자가 사람들을 장기간 추적한다. 연구자는 기준 시점에 사람과 노출에 대한 정보를 수집하고 시간이 경과한 후 결과 발생을 평가한다. 연구자는 일반적으로 노출된 개인과 노출되지 않은 개인을 비교하거나 노출 범주가 다른 개인 그룹을 비교한다. 연구자는 여러 가지 다른 결과를 평가하고 추적 관찰 중 여러 시점에서 노출 및 결과 변수를 조사할 수 있다. 폐쇄형 코호트(예: 출생 코호트)는 연구 시작 시점에 정해진 수의 참가자를 등록하고 그 시점부터 정해진 종료 날짜까지 정해진 간격으로 추적한다.

개방형 코호트에서는 사람들이 서로 다른 시점(예: 한 마을의 주민)에 연구 집단에 들어오고 나가는 등 연구 집단이 역동적이다. 개방형 코호트는 사망, 출생, 이동으로 인해 변화하지만 연령, 성별과 같은 변수와 관련된 인구 구성은 특히 단기간에 걸쳐 거의 일정하게 유지될 수 있다. 폐쇄형 코호트에서는 누적 발생률(위험도)과 발생률을 추정할 수 있으며, 노출된 그룹과 노출되지 않은 그룹을 비교하면 위험비 또는 비율비를 추정할 수 있다. 공개 코호트는 발생률과 비율을 추정한다.

환자대조군연구에서 연구자는 특정 질병 결과가 발생한 사람(환자)과 해당 결과가 발생하지 않은 사람(대조군) 간의 노출을 비교한다. 연구자는 기본 코호트 또는 인구의 단면을 대표하는 환자와 대조군을 수집하는 것을 목표로 한다. 이러한 인구는 지리적으로 정의할 수도 있지만, 의료 시설과 같이 더 느슨하게 정의할 수도 있다. 환자 표본은 100% 또는 가능한 환자의 일부일 수 있지만, 대조군 표본은 일반적으로 관련 결과가 없는 사람들 중 일부에 불과하다. 대조군은 환자가 발생한 코호트 또는 사람들의 집단을 나타낸다. 조사자는 환자와 대조군 사이에서 질병의 추정 원인에 노출될 확률의 비율을 계산한다(박스 7 참조). 환자와 대조군에 대한 추출 전략과 연구 대상 인구의 특성에 따라 환자대조군연구에서 얻은 오즈비는 위험비, 비율비 또는 (유병률) 오즈비로 해석된다[16,17]. 발표된 대부분의 환자대조군연구는 공개 코호트를 표본으로 하므로 비율을 직접 추정할 수 있다.

단면연구에서 연구자는 노출, 위험 요인 또는 질병의 유병률을 조사하기 위해 종종 같은 시점에 표본의 모든 개인을 평가한다. 일부 단면연구는 분석적이며 노출과 질병 사이의 잠재적 인과 관계를 정량화하는 것을 목표로 한다. 이러한 연구는 노출 그룹 간의 질병 유병률을 비교하여 코호트연구처럼 분석할 수 있다. 또한 질병이 있는 그룹과 없는 그룹 간의 노출 확률을 비교하여 환자 대조군 연구처럼 분석할 수도 있다. 모든 설계에서 발생할 수 있지만 단면연구에서 특히 어려운 점은 노출과 결과의 시간 순서가 때때로 명확할 수

있지만 노출이 질병에 선행했음을 입증하는 것이다. 예를 들어 노출 변수가 선천적이거나 유전적인 연구에서는 두 가지를 동시에 측정하더라도 노출이 질병에 선행했다고 확신할 수 있다.

관찰 연구의 범위

관찰 연구는 질병의 잠재적 원인에 대한 첫 번째 힌트를 보고하는 것부터 이전에 보고된 연관성의 규모를 확인하는 것까지 다양한 목적으로 수행된다. 연구에 대한 아이디어는 임상 관찰이나 생물학적 통찰력에서 나올 수 있다. 또한 비공식적으로 자료를 살펴보다가 추가 탐색으로 이어지는 아이디어가 떠오를 수도 있다. 수천 명의 환자를 진료한 임상 의사가 눈에 띄는 환자 한 명을 메모하는 것처럼, 연구자는 자료에서 특별한 것을 발견할 수 있다. 자료를 여러 각도에서 살펴보는 것은 불가능하거나 바람직하지 않을 수 있지만[9], 초기 관찰을 확인하거나 반박하기 위한 추가 연구가 필요한 경우가 많다[10]. 기존 자료는 잠재적 인과 요인에 대한 새로운 아이디어를 조사하는 데 사용될 수 있으며, 그 자체만으로 거부 또는 확증에 충분할 수 있다. 하지만 이전 보고서의 잠재적 문제를 극복하기 위해 특별히 고안된 연구가 뒤따르는 경우도 있으며, 이 경우 새로운 자료를 수집하는 목적을 위해 계획된다. 이를 통해 다양한 견해 예를 들어 하위 그룹을 살펴보는 것의 장점이나 미리 정해진 표본 크기의 중요성과 같은 다양한 관점으로 이어지기도 한다. STROBE는 발견, 이에 대한 반박 또는 확증에 이르기까지 관찰 연구의 다양한 용도를 수용하려고 노력한다. 필요한 경우 특정 권장 사항이 적용되는 상황을 명시한다.

이 문서 사용 방법

이 문서는 여러 저널[6]에 체크리스트 항목을 소개한 짧은 STROBE 논문과 연결되어 있으며, STROBE의 필수적인 부분을 구성한다. 우리의 의도는 연구를 어떻게 해야 하는지가 아니라 연구를 잘 보고하는 방법을 설명하는 것이다. 각 체크리스트 항목에 대한 자세한 설명을 제공한다. 각 설명 앞에는 적절한 예시가 제시된다. 이는 예시를 든 연구가 일률적으로 잘 보고되었거나 잘 수행되었다는 의미도 아니며, 나중에 다른 사람들에 의해 확인되었다는 의미에서 해당 연구 결과가 신뢰할 만하다는 의미도 아니다. 설명과 예시 외에도 박스 1-8에는 보충 정보가 포함되어 있다. 이는 이론적 요점에 대한 기억을 되살리거나 기술적 배경에 대한 세부 정보를 빠르게 얻고자 하는 독자를 위한 것이다. 이러한 요점을 완전히 이해하려면 인용된 교과서나 방법론 논문을 공부해야 할 수도 있다.

STROBE는 유전적 연관성 연구, 감염병 모델링 또는 증례보고 및 환자군연구와 같은 연구설계에는 적용되지 않는다[11,12]. 그럼에도 불구하고 STROBE의 많은 핵심 요소가 이러한 설계에 적용되므로, 이러한 연구를 보고하는 저자는 현재의 권고안을 유용하게 활용할 수 있다. 진단 검사, 종양 표지자 및 유전적 연관성을 구체적으로 다루는 관찰 연구 저자의 경우, STARD [13], REMARK [14], STREGA [15] 권고사항이 유용할 수 있다.

STROBE 체크리스트의 항목

이제 STROBE 체크리스트의 22개 항목(표 1)에 대해 설명하고, 각 항목에 대해 공개된 예시를 제시한다.

표 1. STROBE Statement—checklist of items that should be included in reports of observational studies

	No	권고사항
--	----	------

제목, 초록	1	(a) 제목 또는 초록에 일반적으로 사용되는 용어를 사용하여 연구 설계를 제시한다.
		(b) 초록에 수행된 연구와 발견한 내용에 대한 유익하고 균형 잡힌 요약을 제공한다.
서론		
배경/정당성	2	연구의 과학적 배경과 근거를 제시한다.
목적	3	사전 설정한 가설을 포함해서 연구목적을 제시한다.
방법		
연구디자인	4	논문 초반부에 연구 설계의 핵심 요소를 제시한다.
세팅	5	연구 환경, 장소와 주요 일시(모집, 노출측정, 추적관찰, 자료 수집기간 등을 포함)를 제시한다.
참여자	6	(a) 코호트연구-포함기준과 대상자참여자 선정방법을 제시한다. 환자대조군연구-포함기준과 환자 확인과 대조군 선정 방법을 제시한다. 환자와 대조군 선택의 정당성을 제시한다. 단면연구- 포함기준과 참여자 선정방법을 제시한다
		(b) 코호트연구-짜지은 연구의 경우 짝짓기 기준과 노출군과 비노출군의 수를 제시한다 환자대조군연구- 짜지은 연구의 경우 짝짓기 기준과 환자군 당 대조군의 수를 제시한다.
변수	7	모든 결과, 노출, 예측인자, 잠재적 교란요인 및 효과변경인자를 명확히 정의한다. 해당되는 경우 진단 기준을 제시한다.
자료원/측정	8*	개별 변수에 대해, 자료원과 평가(측정) 방법을 제시한다. 두 군 이상이 있으면 평가방법의 비교성에 대해 기술한다.
바이어스	9	잠재적 바이어스를 해결하기 위한노력에 대해 기술한다.
연구규모	10	연구규모가 어떻게 결정되었지 설명한다.
양적 변수	11	분석에서 양적변수를 어떻게 다루었는지 설명한다. 그룹화한 경우, 그 이유와 방법을 기술한다.
통계적 방법	12	(a) 교란 통제 방법을 포함한 모든 통계분석 방법을 기술한다.
		(b)하위그룹과 상호작용을 평가하는 데 사용한 모든 방법을 기술한다.
		(c) 결측치를 어떻게 다루었는지 기술한다
		(d) 코호트연구-추적 실패를 어떻게 다루었는지 기술한다. 환자대조군연구-환자군과 대조군의 짝짓기 방법에 대해 설명한다.

		단면연구-해당되는 경우, 표본 추출 전략을 고려한 분석방법을 기술한다.
		(e) 모든 민감도 분석에 대해 기술한다

결과

참여자	13*	(a) 각 단계별 인원수를 보고한다. - 잠재적 적격자 수, 적격기준을 평가한 수, 적격 확인된 수, 연구 포함된 수, 추적 완료자 수, 분석자 수 (b) 각 단계에서 제외된 이유를 밝힌다. (c) 흐름도 사용을 고려한다.
자료기술	14*	(a) 연구참여자의 일반적 특성(인구학적, 임상적, 사회적), 노출변수 및 잠재적 교란변수에 대한 정보를 제시한다. (b) 관심 변수별 자료가 결측된 사람 수를 제시한다. (c) 코호트연구-추적관찰시간 정보를 요약한다(평균, 총관찰기간).
결과자료	15*	코호트연구-시간에 따른 결과 발생 수 혹은 요약 정보를 제시한다. 환자대조군연구-노출 범주별 수 혹은 요약 정보를 제시한다. 단면연구- 결과 발생 수 혹은 요약정보를 제시한다.
주요 결과	16	(a) 보정하지 않은 추정치를 제시한다. 해당되는 경우, 교란변수를 보정한 추정치와 그 정밀도(예, 95% 신뢰구간)를 제시한다. 보정한 교란변수들과 선택한 이유를 밝힌다. (b) 연속변수를 범주화 했으면 범주의 범위를 제시한다. (c) 적절하다면, 상대위험도를 의미 있는 기간 동안의 절대위험도로 변환하는 것을 고려한다.
기타 분석	17	하위그룹 및 상호작용 분석과 민감도 분석 등 기타 수행된 분석을 보고한다.

고찰

주요 결과	18	연구 목적에 비추어 주요 결과를 요약한다.
제한점	19	잠재적 바이어스나 오류의 가능성을 고려하여 연구의 제한점을 고찰한다. 잠재적 바이어스에 대해서는 그 방향성과 크기에 대해서도 고찰한다.
해석	20	연구목적, 제한점, 유사한 연구 결과와 다른 관련 근거들을 고려하여 결과에 대한 신중한 전반적 해석을 제시한다.
일반화가능성	21	연구 결과의 일반화 가능성(외적타당도)에 대해 고찰한다.

다른 정보

자금원	22	본 연구의 연구비 출처와 연구비 제공자의 역할을 보고한다. 해당되는 경우, 본 논문이 기반한 원래 연구에 대해서도 설명한다.
-----	----	---

일부 예시는 인용문을 삭제하거나 약어를 수정하여 편집했다. 18개 항목은 세 가지 연구 디자인에 모두 적용되며, 4개 항목은 디자인에 따라 다르다. 별표 제시된 항목(예: 항목 8*)은 환자대조군연구에서 환자군과 대조군, 코호트 및 단면연구에서 노출군과 비노출군에 대해 정보를 별도로 제공해야 함을 나타낸다. 저자는 모든 항목을 논문의 어딘가에 언급할 것을 권장하지만, 정확한 위치나 순서를 규정하지는 않는다. 예를 들어, 여러 개의 개별 항목으로 결과를 보고하는 것에 대해 논의하지만, 저자가 단일 텍스트 파트 또는 표에서 여러 항목을 다룰 수도 있다.

항목

제목과 초록

1 (a). 제목 또는 초록에 일반적으로 사용되는 용어를 사용하여 연구 설계를 제시한다.

예시

"신발 및 부츠 제조 산업 근로자의 백혈병 발생률: 환자 대조 연구"[18].

설명

독자가 제목이나 초록에서 사용된 디자인을 쉽게 식별할 수 있어야 한다. 연구 설계에 대해 명시적이고 일반적으로 사용되는 용어는 전자 데이터베이스에서 논문의 올바른 색인에 도움이 된다[19,20].

1 (b). 초록에 수행한 작업과 발견한 내용에 대한 유익하고 균형 잡힌 요약을 제공한다.

예시

"배경: HIV에 감염된 환자의 예상 생존율은 공중 보건의 주요 관심사이다.

목적: 일반 인구와 비교하여 HIV에 감염된 인구의 생존 기간과 연령별 사망률을 추정한다.

디자인: 인구 기반 코호트연구.

설정: 1995년부터 2005년까지 덴마크에서 치료를 받은 모든 HIV 감염인.

환자: 전국적인 덴마크 HIV 코호트연구에 참여한 각 구성원은 성별, 생년월일, 거주 지역에 따라 일반 인구에서 최대 99명과 짝짓기되었다.

측정: 저자들은 나이를 시간 척도로 하여 카플란-마이어 수명표를 계산하여 연령에 따른 생존율을 25년으로 추정했다. HIV 감염 환자와 이에 해당하는 일반 인구는 환자의 HIV 진단일로부터 사망, 이민 또는 2005년 5월 1일까지 관찰되었다.

결과: 3990명의 HIV 감염 환자와 37만 9,872명의 일반 인구가 연구에 참여하여 2만 2,744명(중앙값, 5.8세/인)과 268만 9,287명(중앙값, 8.4년/인)의 관찰 기간을 기록했다. 참가자의 3%가 추적 조사에서 손실되었다. 25세부터의 생존기간 중앙값은 HIV 감염 환자의 경우 19.9년(95% CI, 18.5 ~ 21.3), 일반 인구의 경우 51.1년(CI, 50.9 ~ 51.5)이었다. HIV 감염 환자의 경우 2000~2005년 기간 동안 생존 기간이

32.5년(CI, 29.4~34.7)으로 증가했다. C형 간염 동시 감염이 확인된 환자(16%)를 제외한 하위 그룹에서는 같은 기간 동안 생존 중앙값이 38.9년(CI, 35.4~40.1)으로 나타났다. 일반 인구에 비해 HIV 감염 환자의 상대 사망률은 연령이 증가할수록 감소한 반면, 초과 사망률은 연령이 증가할수록 증가했다.

제한점: 관찰된 사망률은 현재 최대 관찰 기간인 10년 이후에도 적용되는 것으로 가정한다.

결론: 고강도 항레트로바이러스 치료 후기에 HIV 감염 진단을 받은 젊은이의 예상 생존 중앙값은 35년 이상이다. 그러나 일반 인구에 비해 이러한 사람들의 사망률을 더욱 낮추기 위해서는 여전히 지속적인 노력이 필요하다 [21].

설명

초록은 독자가 연구를 이해하고 논문을 읽을지 여부를 결정할 수 있도록 핵심 정보를 제공한다. 일반적인 구성 요소에는 연구 질문, 방법 및 결과에 대한 간략한 설명, 결론이 포함된다[22]. 초록은 연구의 주요 세부 사항을 요약해야 하며 논문에서 제공되는 정보만 제시해야 한다. 주요 결과는 참여자 수, 연관성 추정치, 변동성 및 불확실성에 대한 적절한 측정치(예: 신뢰 구간이 있는 오즈비)를 포함하는 수치 형식으로 제시하는 것이 좋다. 노출이 결과와 유의미하게 연관되어 있거나 연관되어 있지 않다는 것만 명시하는 것은 불충분하다. 연구의 배경, 설계, 수행 및 분석과 관련된 일련의 제목은 독자가 필수 정보를 빠르게 습득하는 데 도움이 될 수 있다[23]. 많은 학술지에서는 이러한 구조화된 초록을 요구하며, 이는 구조화되지 않은 요약보다 더 높은 품질과 더 많은 정보를 제공하는 경향이 있다[24,25].

서론

서론 파트에서는 연구를 수행한 이유와 연구 질문 및 가설을 설명해야 한다. 다른 사람들이 연구의 맥락을 이해하고 현재 지식에 대한 잠재적 기여도를 판단할 수 있도록 해야 한다.

2. 배경/근거: 현재 연구의 과학적 배경과 정당성을 제시한다.

예시

"아동과 청소년의 비만 유병률 증가에 대한 우려는 아동기 비만과 성인기 심혈관 질환 위험 및 사망률 증가 사이의 잘 문서화된 연관성에 초점을 맞추고 있다. 소아 비만은 아동기와 청소년기에 상당한 사회적, 심리적 영향을 미치지만 성인기의 사회적, 사회경제적, 심리적 영향에 대해서는 알려진 바가 거의 없다. 최근의 체계적 문헌고찰에 따르면 신체적 건강 결과 외에 아동기 비만의 결과에 대한 추적 연구는 없었으며, 청소년기 비만의 사회경제적 영향에 대한 추적 연구는 단 두 건에 불과했다. Gortmaker 등은 1981년 청소년기 후반에 비만이었던 미국 여성은 과체중이 아니었던 여성에 비해 7년 후 결혼할 확률이 낮고 소득이 낮은 반면, 과체중이었던 남성은 결혼할 확률이 낮다는 것을 발견했다. Sargent 등은 1974년 16세 때 비만이었던 영국 여성(남성은 제외)이 23세 때 비만이 아닌 또래보다 소득이 7.4% 적다는 사실을 발견했다. (... .) 우리는 1970년 영국 출생 코호트의 추적 자료를 사용하여 아동 비만의 성인 사회경제적, 교육적, 사회적, 심리적 결과를 조사했다"[26].

설명

이 연구의 과학적 배경은 독자들에게 중요한 맥락을 제공한다. 연구의 무대를 설정하고 연구의 초점을 설명한다. 주제에 대해 알려진 내용과 현재 지식의 격차에 대한 개요를 제공하고 연구에서 어떤 부분을 다루고 있는지 설명한다. 배경 자료에는 최근의 관련 연구와 관련 연구에 대한 체계적인 검토가 포함되어야 한다.

3. 목적: 미리 정해진 가설을 포함하여 구체적인 연구목표를 제시한다.

예시

"우리의 주요 목적은 1) 다양한 사회 경제적 배경을 가진 환자들에게 서비스를 제공하는 4개의 지역사회 기반 일차 진료 성인 의학 진료소에 내원하는 여성 환자들 사이에서 가정 폭력의 유병률을 확인하고, 2) 현재 학대받는 환자와 현재 학대받지 않는 환자 사이의 인구통계학적 및 임상적 차이를 확인하는 것이었다"[27].

설명

연구목적은 연구의 세부적인 목표이다. 잘 만들어진 목적은 모집단, 노출 및 결과, 추정할 매개변수를 명시한다. 목적은 특정 가설 또는 연구가 해결하도록 설계된 질문으로 공식화될 수 있다. 초기 발견 단계와 같이 일부 상황에서는 목적이 덜 구체적일 수 있다. 그럼에도 연구자의 의도를 명확하게 보고해야 한다. 예를 들어, 중요한 하위 그룹이나 추가 분석이 원래 연구의 목적이 아니었지만 자료 분석 중에 발생한 경우, 이에 대해 적절하게 기술해야 한다(항목 4, 17, 20 참조).

방법

방법 파트에서는 독자들이 연구의 본질적인 측면을 이해할 수 있도록 계획된 내용과 수행된 내용을 충분히 상세히 기술하고, 방법이 신뢰할 수 있고 유효한 답변을 얻는 데 적절한지 판단하고 신뢰할 수 있고 유효한 답변을 제공하는 데 적절한지 판단하고, 원래 계획에서 벗어난 부분이 합리적인지 평가할 수 있어야 한다.

4. 연구 설계: 각 논문의 초반부에 연구디자인의 핵심적인 요소를 제시한다.

예시

"우리는 환자-대조군 설계의 변형인 환자-교차 설계를 사용했는데, 이는 짧은 노출(운전자의 휴대폰 사용)이 드문 결과(충돌 사고)의 위험을 일시적으로 높힐 때 적절하다. 우리는 충돌이 발생한 것으로 추정되는 시간에 운전자가 휴대폰을 사용한 것과 다른 적절한 시간대에 같은 운전자가 휴대폰을 사용한 것을 비교했다. 운전자는 스스로 제어하기 때문에 충돌 위험에 영향을 미칠 수 있지만 단기간에 변경되지 않는 운전자의 특성을 설계에서 제어한다. 통제기간 동안의 위험과 충돌 위험이 유사해야 하므로, 위험 구간(충돌 직전 시간)의 휴대폰 활동과 전주 통제 구간(참가자가 운전 중이었지만 충돌하지 않은 동등한 시간) 동안의 휴대폰 활동을 비교했다[28].

설명

독자가 연구의 기본 사항을 이해할 수 있도록 방법 파트 초반(또는 서론 끝 부분)에 연구 설계의 핵심 요소를 제시하는 것이 좋다. 예를 들어, 저자는 연구가 일정 기간 동안 추적관찰한 코호트연구였다는 점을 명시하고 코호트별 노출 상태를 설명해야 한다. 마찬가지로, 환자-대조군 설계인 경우, 환자군과 대조군과 모집단을 설명해야 한다. 단면 조사인 경우, 인구집단과 횡단면 조사 시점을 언급해야 한다. 연구가 세 가지 주요 연구 유형의 변형인 경우 명확성이 추가로 필요하다. 예를 들어, 환자대조군연구 설계의 변형 중 하나인 환자 교차 연구의 경우 예시와 같이 간결한 설명이 필요할 수 있다[28].

이러한 연구 설계 용어는 정의가 명확하지 않으므로 저자는 연구를 단순히 '전향적' 또는 '후향적'이라고 부르는 것을 자제해야 한다[29]. 일부 문헌에서 코호트와 전향적 연구를 동의어로 간주하고 환자 대조군 연구를 후향연구라고 표현한다[30]. 다른 문헌에서는 연구 아이디어가 개발된 시점을 기준으로 전향적 코호트연구와 후향적 코호트연구를 구분한다[31]. 세 번째 용법은 환자를 선정할 때 관심 있는 노출에 대한 자료가 존재했는지 여부에 따라 전향적 및 후향적 환자 대조군 연구를 구분한다[32]. 일부에서는 이러한 용어를 사용하지 않거나[33], 코호트연구를 설명할 때 '동시적' 및 '역사적'이라는 대안을 채택할 것을 권고하기도 한다[34]. STROBE에서는 전향적 및 후향적이라는 단어와 동시적 및 역사적이라는 대체 용어를 사용하지 않는다. 저자가 이러한 단어를 사용할 때마다 그 의미를 정의할 것을 권장한다. 가장 중요한 것은 저자가 자료 수집이 언제 어떻게 이루어졌는지 정확히 설명하는 것이 좋다.

방법 파트의 첫 번째 부분에는 해당 보고서가 여러 연구 중 하나인지 여부를 언급할 수도 있다. 새로운 보고서가 원래의 연구 목표와 일치하는 경우, 일반적으로 이전 출판물을 참조하고 연구의 주요 특징을 간략하게 다시 설명함으로써 이를 제시한다. 그러나 연구의 목적은 시간이 지남에 따라 변화할 수도 있다. 연구자들은 주로 관리 목적으로 수집된 공식적인 생체 통계, 원래는 완전성을 위해서만 포함된 설문지의

항목, 다른 목적으로 수집된 혈액 샘플 등 원래 의도하지 않은 용도로 자료를 사용하는 경우가 많다. 예를 들어, 아스피린과 카로틴에 대한 무작위 대조 시험인 의사 건강 연구는 나중에 응고인자 V 유전자의 점 돌연변이가 정맥 혈전증 위험 증가와 관련이 있지만 심근경색이나 뇌졸중과는 관련이 없음을 입증하는 데 사용되었다[35]. 기존 자료를 이차적으로 사용하는 것은 관찰 연구의 창의적인 부분이며 반드시 결과의 신뢰도가 떨어지거나 중요도가 낮아지는 것은 아니다. 그러나 원래의 연구 목적을 간략하게 다시 설명하는 것은 독자가 연구의 맥락과 자료의 가능한 한계를 이해하는 데 도움이 될 수 있다.

5. 세팅 : 모집기간, 노출, 추적관찰, 자료 수집을 포함한 세팅, 위치, 적절한 일시를 명시한다.

예시

"파시토스 코호트연구는 1998년 4월부터 2000년 10월까지 텍사스 엘패소 카운티의 소코로와 산 엘리자리오에 있는 여성, 유아 및 아동 클리닉과 멕시코 시우다드 후아레스에 있는 멕시코 사회보장연구소의 모자 클리닉에서 임산부를 모집했다. 등록된 코호트 아동이 출생하기 전인 기준 시점에 연구진은 어머니를 대상으로 가정 환경에 관한 인터뷰를 실시했다. 현재 진행 중인 이 코호트연구에서는 생후 6개월부터 6개월 간격으로 후속 검사를 실시하였다[36].

설명

독자는 연구 결과의 맥락과 일반화 가능성을 평가하기 위해 환경 요인과 장소에 대한 정보가 필요하다. 환경 요인 및 치료법과 같은 노출은 시간이 지남에 따라 변할 수 있다. 또한 연구 방법도 시간이 지남에

따라 발전할 수 있다. 연구가 언제 진행되었는지, 어떤 기간 동안 참가자를 모집하고 추적 관찰했는지 아는 것은 연구를 역사적 맥락에서 파악하고 결과를 해석하는 데 중요하다.

세팅 정보에는 모집 장소 또는 출처(예: 선거인 명부, 외래 진료소, 암 등록 센터 또는 3차 진료 센터)가 포함된다. 위치에 대한 정보는 조사가 이루어진 국가, 도시, 병원 등을 나타낼 수 있다. 기간만 설명하지 말고 날짜를 명시하는 것이 좋다. 노출, 질병 발생, 모집, 추적 관찰 시작과 종료, 자료 수집 날짜는 서로 다를 수 있다. 주목할 점은 생존 분석을 사용한 종양학 저널의 132개 보고서 중 약 80%가 환자 발생 시작일과 종료일을 포함했지만 추적 관찰이 종료된 날짜를 보고한 경우는 24%에 불과했다[37].

6. 참가자

6 (a). 코호트연구: 포함기준과 참여자 선정방법을 제시한다. 후속 조치 방법을 설명한다.

예시

"아이오와 여성 건강 연구의 참가자는 1985년 아이오와 주 자동차 운전 면허증 목록에서 추출한 55~69세의 모든 여성 중 무작위 표본으로, 해당 연령대의 아이오와 여성 중 약 94%를 차지했다. (...) 1987년 10월과 1989년 8월에 후속 설문지를 우편으로 발송하여 생체 상태와 주소 변경 사항을 평가했다.(...). 비흑색종 피부암을 제외한 발생 암은 아이오와 주 보건 등록부 에서 확인했다.(...) 아이오와 여성 건강 연구 코호트는 이름, 성, 처녀 이름, 우편 번호, 생년월일 및 사회 보장 번호의 조합으로 레지스트리와 일치했다 "[38].

6 (a). 환자대조군연구—포함기준과 환자 확인과 대조군 선정 방법을 제시한다. 환자와 대조군 선택의 정당성을 제시한다.

예시

"1999년과 2000년에 진단된 피부 흑색종 환자는 아이오와 암 등록부를 통해 확인되었다.(...) 아이오와 암 등록부를 통해 확인된 대조군은 같은 기간에 진단된 대장암 환자이다. 대장암은 흔하고 생존 기간이 비교적 길며 비소 노출이 대장암 발생과 결정적으로 연관되지 않았기 때문에 대장암 대조군을 선택했다[39].

6 (a). 단면연구: 포함기준과 참여자 선정방법을 제시한다.

예시

"국제질병분류 9차 개정판에 따른 심근경색(코드 410)을 주 진단으로 하는 환자를 퇴원 진단 코드에서 후향적으로 식별했으며, 다섯 번째 자리가 2인 코드는 후속 치료 에피소드를 지정하였다(...) 1994년 2월부터 1995년 7월까지 심근경색이 발생한 전체 메디케어 코호트에서 무작위 표본이 선정되었다 (...) 적격 환자는 최소 30분 이상 12시간 미만의 흉통이 발생한 후 병원에 내원해야 했으며, 초기 심전도에서 2개의 연속된 리드에서 최소 1mm의 ST 분절 상승이 있어야 했다[40].

설명

연구 참여자에 대한 자세한 설명은 독자가 연구 결과의 적용 가능성을 이해하는 데 도움이 된다. 연구자는 일반적으로 적격 참여자의 임상적, 인구통계학적 및 기타 특성을 정의하여 연구 모집단을 제한한다. 일반적인 자격 기준은 연령, 성별, 진단 및 동반 질환과 관련이 있다. 이러한 중요성에도 불구하고 자격

기준이 적절하게 보고되지 않는 경우가 많다. 뇌졸중 관찰 연구에 대한 조사에서 49건의 보고서 중 17건(35%)이 자격 기준을 명시하지 않았다[5].

적격성 기준은 포함 및 제외 기준으로 제시될 수 있지만, 이러한 구분이 항상 필요하거나 유용한 것은 아니다. 어쨌든 저자는 모든 자격 기준을 보고하고 연구 모집단을 선정한 그룹(예: 지역 또는 국가의 일반 인구) 및 모집 방법(예: 광고를 통한 추천 또는 자체 선정)을 설명하는 것이 좋다.

후속 조사 절차가 무응답과 후속 조사 손실을 최소화했는지, 모든 참가자에게 유사한 절차를 적용했는지 등 후속 조사 절차에 대한 세부 정보를 파악하면 결과의 타당성을 판단하는 데 도움이 된다. 예를 들어, 급성 감염을 감지하기 위해 IgM 항체를 사용한 연구에서 독자들은 혈액 검사 간격이 너무 길어 일부 감염을 놓쳤을 가능성이 있는지 판단할 수 있도록 IgM 항체 검사 간격을 알아야 했다[41]. 노출된 그룹과 노출되지 않은 그룹 간에 추적관찰 절차가 달랐던 다른 연구들에서, 독자들은 사건의 불균등한 확인 또는 추적관찰에 대한 무응답 또는 손실의 차이로 인해 상당한 바이어스를 인지할 수 있다[42]. 따라서 연구자는 참가자를 추적하는 데 사용된 방법과 그 방법이 모든 참가자에게 동일하게 적용되었는지, 변수 확인의 완전성을 설명하는 것이 좋다(14번 항목 참조).

환자대조군연구에서 환자와 대조군의 선택은 결과를 해석하는 데 매우 중요하며, 그 선택 방법은 연구의 타당성에 큰 영향을 미친다. 일반적으로 대조군은 환자가 발생한 집단에서 나와야 한다. 일반 인구에서 발생한 환자의 경우 인구 명단 추출, 무작위 전화 걸기, 이웃 또는 친구 대조군 등 장단점이 있는 다양한 방법이 대조군 표본 추출에 사용된다. 이웃 또는 친구 대조군은 노출에 내재적 짝짓기가 있을 수 있다[17]. 다른 질병을 가진 대조군은 인구 기반 대조군, 특히 병원 기반 환자의 경우 병원의 모집단을 더 잘 반영하고 리콜의 비교가능성이 높으며 모집이 용이하기 때문에 이점이 있을 수 있다. 그러나 관심 노출이

대조군 질환의 발병 또는 입원 위험에 영향을 미치는 경우 문제가 발생할 수 있다[43,44]. 이 문제를 해결하기 위해 종종 최상의 방어 가능한 통제 질병이 사용된다 [45].

6 (b). 코호트연구—짜지은 연구의 경우 짝짓기 기준과 노출군과 비노출군의 수를 제시한다.

예시

"처음에 스타틴을 투여받은 각 환자에 대해 성향 기반 짝짓기를 사용하여 다음 프로토콜에 따라 스타틴을 투여받지 않은 대조군 1명을 식별했다. 첫째, 스타틴 사용 또는 패혈증 위험과 잠재적으로 관련된 광범위한 요인 목록을 기반으로 전체 코호트의 각 환자에 대한 성향 점수를 계산했다. 둘째, 각 스타틴 사용자는 성별, 연령(1년 이상 또는 미만), 기준일(3개월 이상 또는 미만)에 따라 더 작은 규모의 비스타틴 사용자 풀에 짝짓기되었다. 셋째, 각 스타틴 사용자와 가장 가까운 성향 점수(0.2 SD 이내)를 가진 대조군을 1:1 방식으로 선택하고 나머지 대조군은 폐기했다. [46].

6 (b). 환자 대조군 연구: 짜지은 연구의 경우 짝짓기 기준과 환자 당 대조군의 수를 제시한다.

예시

"우리는 일반 진료 기록에 자폐증 또는 기타 만연성 발달 장애(other pervasive developmental disorders, PDD) 진단이 기록되어 있지 않고, 환자의 PDD 진단일에 참여 진료소에 등록되어 있고

생존해 있는 연구 집단 내 개인 중에서 모든 환자에 대해 5명의 대조군을 선정하는 것을 목표로 삼았다.

대조군은 생년월일(최대 1세 이하), 성별, 일반 진료에 따라 환자와 개별적으로 짝짓기되었다. 300건의 환자 각각에 대해 모든 짝짓기 기준을 충족하는 대조군 5명을 식별할 수 있었다. 나머지 994건의 경우 하나 이상의 대조군이 제외되었다..." [47].

설명

짝짓기는 환자대조군연구에서 훨씬 더 일반적이지만, 때때로 연구자가 코호트연구에서 짝짓기를 사용하는 경우도 있는데 그 이유는 추적 관찰을 시작할 때 그룹을 비교 가능하게 해주기 때문이다. 코호트연구에서 짝짓기를 사용하면 잠재적 교란 요인에 대해 그룹을 직접 비교할 수 있으며 환자 대조군 연구보다 복잡한 문제가 덜 발생한다. 예를 들어, 상대 위험도 추정을 위해 짝짓기를 고려할 필요가 없다[48].

코호트연구에서의 짝짓기는 통계적 정밀도를 높일 수 있으므로 연구자는 분석에서 짝짓기를 허용하여 더 좁은 신뢰 구간을 얻을 수 있다.

환자대조군연구에서 짝짓기는 환자와 대조군 간의 변수 분포, 특히 잠재적 교란 변수의 분포의 유사성을 보장하여 연구의 효율성을 높이기 위해 수행된다[48,49]. 짝짓기는 환자당 하나 이상의 대조군을 사용하여 다양한 방법으로 수행할 수 있으므로, 짝짓기 변수 선택의 근거와 사용된 방법의 세부 사항을 설명해야 한다. 일반적으로 사용되는 짝짓기 형태는 빈도 짝짓기(그룹 짝짓기이라고도 함)과 개별 짝짓기이다. 빈도 짝짓기에서는 조사자가 대조군을 선택하여 짝짓기 변수의 분포가 환자의 분포와 동일하거나 유사하도록 한다. 개별 짝짓기는 각 환자에 하나 또는 여러 개의 대조군을 짝짓기하는 것이다. 직관적으로 매력적이고 때로는 유용하지만, 환자대조군연구에서의 짝짓기에는 여러 가지 단점이 있으며 항상 적절한 것은

아니므로 분석에서 이를 고려해야 한다(박스 2 참조).

간단한 짝짓기 절차조차 제대로 보고되지 않을 수 있다. 예를 들어, 저자는 대조군을 '5년 이내' 또는 '5년 연령대'를 사용하여 환자와 일치시켰다고 명시할 수 있다. 즉, 환자가 54세인 경우 해당 대조군은 50세에서 54세 사이 또는 54세에서 5년 이내인 49세에서 59세 사이에 있어야 한다는 의미인가? 넓은 연령대(예: 10년)를 선택하면 대조군이 평균적으로 환자보다 젊을 수 있기 때문에 연령에 따른 잔존 교란의 위험이 있다(박스 4 참조).

7. 변수: 적용가능 하다면, 모든 결과, 노출, 예측인자, 잠재적 교란인자, 효과변경자를 명확히 정의한다.

예시

"주요 선천성 기형만 분석에 포함되었다. 경미한 기형은 유럽 선천성 기형 등록(EUROCAT)의 제외 목록에 따라 제외되었다. 한 아이에게 한 장기 계통의 주요 선천성 기형이 두 개 이상 있는 경우, 해당 기형은 장기 계통별 분석에서 하나의 결과로 처리되었다 (...) 통계 분석에서 잠재적 교란 요인으로 고려된 요인은 분만 시 산모의 연령과 이전 동등성 수였다. 잠재적 효과 변경 요인으로 고려된 요인은 항간질제 환급 시 산모의 나이와 분만 시 산모의 나이였다[55].

설명

저자는 결과, 노출, 예측인자, 잠재적 교란변수, 잠재적 효과 변경변수 등 분석에 고려되고 분석에 포함된

모든 변수를 정의해야 한다. 질병 결과에는 진단 기준에 대한 적절하고 상세한 설명이 필요하다. 이는 환자대조군연구의 환자군, 코호트연구의 추적 관찰 중 질병 발생, 단면연구의 유행성 질병에 대한 기준에 적용된다. 명확한 정의와 이를 준수하기 위해 취한 조치는 연구에서 주요 관심사인 질병 상태에 대해 특히 중요하다.

일부 연구의 경우 노출 변수는 '결정요인' 또는 '예측요인'으로, 결과는 '평가변수'로 부르는 것이 적절할 수 있다. 다변량 모델에서 저자는 결과를 '종속 변수'로, 노출 및 교란 변수에 대해 '독립 변수' 또는 '설명 변수'라는 용어를 사용하는 경우가 있다. 후자는 노출과 교란 변수를 구분하지 않기 때문에 정확하지 않다. 초기 발견 단계에서 많은 변수를 측정하여 탐색 분석에 포함시킨 경우, 부록, 추가 표 또는 별도의 출판물에 각 변수에 대한 세부 정보가 포함된 목록을 제공하는 것이 좋다. 주목할 만한 점은 최근 국제 역학 저널에서 특정 연구에서 여러 시점에 측정된 항목에 대한 자세한 정보를 포함하는 '코호트 프로필' 파트를 신설했다는 점이다[56,57]. 마지막으로, 저자는 최종 모델에 포함된 변수만 선택적으로 보고하지 말고 통계 분석을 위해 고려한 모든 '후보 변수'를 보고할 것을 권장한다(항목 16a 참조)[58,59].

박스 2. 환자대조군연구에서의 짝짓기

환자대조군연구에서는 환자와 대조군의 짝짓기를 사용할지, 사용한다면 어떤 변수를 짝짓기할지, 정확한 짝짓기 방법과 적절한 통계 분석 방법을 현명하게 선택해야 한다. 전혀 일치시키지 않으면 일부 주요 잠재적 교란 변수(예: 연령, 성별)의 분포가 환자와 대조군 간에 근본적으로 다르다는 것을 의미할 수 있다. 분석에서 이를 보정할 수 있지만 통계적 효율성에 큰 손실이 있을 수 있다.

환자대조군연구에서의 짝짓기 사용과 그 해석은 특히 여러 위험 요인에 대해 짝짓기를 시도하는 경우, 그 중

일부가 주요 관심사 노출과 관련이 있을 수 있는 경우 어려움이 따른다[50,51]. 예를 들어, 잠재적 대조군으로 사용할 수 있는 수천 명의 여성에 대한 정보가 있는 대규모 약물 역학 자료 베이스에 중첩된 심근경색 및 경구 피임약에 대한 환자 대조군 연구에서 연구자는 각 심근경색 환자와 유사한 수준의 위험 요소를 가진 일치하는 대조군을 선택하고 싶을 수 있다. 한 가지 목적은 경구 피임약 처방에 영향을 미칠 수 있는 요인을 보정하여 적응증에 따른 교란을 통제하는 것이다. 그러나 심근경색 환자는 나이가 많은 경향이 있기 때문에 대조군은 더 이상 경구 피임약 복용을 대표하지 않는 대조군이 될 것이다. 이는 몇 가지 시사점이 있다.

자료를 조잡하게 분석하면 짝짓기 요인이 노출과 연관된 경우 일반적으로 단일성에 편향된 오즈비가 산출된다. 이에 대한 해결책은 짝짓기 또는 계층화 분석을 수행하는 것이다(항목 12d 참조). 또한, 짝짓기된 대조군은 전체 모집단을 대표하지 않기 때문에 대조군 간의 노출 분포를 더 이상 모집단 기여 비율을 추정하는 데 사용할 수 없다(박스 7 참조)[52]. 또한 짝짓기 요인의 효과를 더 이상 연구할 수 없으며, 잘 짝짓기된 대조군을 찾는 것이 번거로울 수 있으므로 짝짓기되지 않은 대조군을 구하기가 더 쉽고 대조군 규모가 더 클 수 있으므로 짝짓기되지 않은 대조군을 사용한 설계가 더 바람직한다. 과도한 짝짓기는 또 다른 문제로, 일치하는 환자대조군연구의 효율성을 떨어뜨리고, 바이어스를 유발할 수도 있다. 짝짓기 변수가 노출과 밀접한 관련이 있는 경우 정보가 손실되고 연구의 힘이 감소한다. 그러면 동일한 짝짓기 세트의 많은 개인이 동일하거나 유사한 수준의 노출을 갖는 경향이 있으므로 관련 정보를 제공하지 못한다. 짝짓기 변수가 교란 변수가 아니라 노출과 질병 사이의 인과 경로에 있는 경우 짝짓기는 수정할 수 없는 바이어스를 유발할 수 있다. 예를 들어, 체외 수정은 다태아 및 저체중아 출산 증가로 인한 주산기 사망 위험 증가와 관련이 있다[53]. 다태아 또는 출생 체중을 기준으로 짝짓기를 수행하면 결과가 null로 편향되며, 이는 분석에서 해결할 수 없다.

짜짓기는 직관적으로 매력적이지만 관련된 복잡성으로 인해 방법론가들은 환자 대조군 연구에서 일상적인 짜짓기를 사용하지 말 것을 권고하고 있다. 대신 각 잠재적 짜짓기 요인을 신중하게 고려하고, 짜짓기를 하지 않고 보정 변수로 측정하여 사용할 수 있다는 점을 인식할 것을 권장한다. 이에 따라 사용되는 짜짓기 요인의 수가 줄어들고, 위에서 논의한 몇 가지 문제를 피할 수 있는 빈도 짜짓기의 사용이 증가하고 있으며, 짜짓기를 전혀 사용하지 않는 환자 대조군 연구도 증가하고 있다[54]. 짜짓기는 교란 요인(예: 연령)의 분포가 일치하지 않는 비교 그룹 간에 근본적으로 다를 수 있는 경우에 가장 바람직하거나 심지어 필수적이다[48,49].

8. 자료원/측정: 개별 변수에 대해, 자료원과 평가(측정) 방법을 제시한다. 두 군 이상이면 평가방법의 비교성에 대해 기술한다.

예시 1

"총 카페인 섭취량은 주로 미국 농무부 식품 성분 자료를 사용하여 계산했다. 카페인 함량은 커피 한 잔당 137mg, 차 한 잔당 47mg, 콜라 음료 한 캔 또는 한 병당 46mg, 초콜릿 캔디 1개 7mg이라고 가정했다. 이 (카페인) 섭취량 측정 방법은 NHS I 코호트와 남성 건강 전문가를 대상으로 한 유사한 코호트연구 모두에서 유효한 것으로 나타났다 (...) 자가 보고된 고혈압 진단은 NHS I 코호트에서 신뢰할 수 있는 것으로 밝혀졌다[60].

예시 2

"짜지은 환자와 대조군 샘플은 항상 동일한 배치에서 함께 분석되었으며 실험실 직원은 환자와 대조군을 구분할 수 없었다"[61].

설명

노출, 교란 요인 및 결과를 측정하는 방식은 연구의 신뢰성과 타당성에 영향을 미친다. 측정 오류와 노출 또는 결과의 잘못된 분류는 인과 관계를 감지하기 어렵게 만들거나 가짜 관계를 생성할 수 있다. 잠재적 교란요인에 대한 측정 오류는 잔류 교란의 위험을 증가시킬 수 있다[62,63]. 따라서 평가 또는 측정의 유효성 또는 신뢰성에 대한 연구 결과를 보고할 때 사용된 참조 표준에 대한 세부 정보를 포함하면 도움이 된다. 첫 번째 예에서처럼 단순히 타당도 연구를 인용하기보다는 저자가 측정 오차 보정 또는 민감도 분석에 사용할 수 있는 추정 타당도 또는 신뢰도를 제시하는 것이 좋다(12e 및 17번 항목 참조).

또한 비교 대상 그룹이 자료 수집 방식과 관련하여 차이가 있는지 파악하는 것도 중요하다. 이는 실험실 검사(두 번째 예시에서와 같이) 등에서 중요할 수 있다. 예를 들어, 면접관이 먼저 모든 환자에 대해 질문한 다음 대조군에 대해 질문하거나 그 반대의 경우 학습 곡선으로 인해 바이어스가 발생할 수 있으며, 면접 순서를 무작위로 지정하는 등의 방법으로 이 문제를 방지할 수 있다. 비교 그룹에 동일한 진단 검사를 제공하지 않거나 한 그룹이 다른 그룹보다 같은 종류의 검사를 더 많이 받는 경우에도 정보 바이어스가 발생할 수 있다(9번 항목 참조).

9. 바이어스: 잠재적 바이어스를 다룬 노력에 대해 기술한다.

예시 1

"자살에 대한 대부분의 환자 대조 연구에서는 대조군이 살아있는 개인으로 구성되지만, 우리는 다른 원인으로 사망한 사람들로 구성된 대조군을 갖기로 했다 (...). 사망한 개인으로 구성된 대조군의 경우, 위험 요인을 평가하는 데 사용되는 정보 출처는 최근에 가족이나 가까운 동료의 죽음을 경험한 정보원이므로 살아있는 대조군을 사용할 때보다 자살 그룹의 정보 출처와 더 비슷하다"[64].

예시 2

"제2형 당뇨병이 있는 여성이 당뇨병이 없는 여성보다 더 면밀한 안과 감시를 받는 경우, 감지(detection) 바이어스가 제2형 당뇨병(T2DM)과 원발성 개방각 녹내장(POAG) 사이의 연관성에 영향을 미칠 수 있다. 당뇨병이 있는 여성과 없는 여성이 보고한 평균 안과 검진 횟수를 비교했다. 또한 보다 면밀한 안과 감시와 관련된 공변량(백내장, 황반변성, 안과 검사 횟수, 신체 검사 횟수에 대한 자가 보고)을 추가로 통제하여 POAG의 상대적 위험도를 다시 계산했다. [65].

설명

편향된 연구는 사실과 체계적으로 다른 결과를 초래한다(박스 3 참조). 독자는 바이어스의 가능성을 줄이기 위해 연구를 수행하는 동안 어떤 조치가 취해졌는지 아는 것이 중요하다. 이상적으로 연구자는 연구를 계획할 때 잠재적인 바이어스의 원인을 신중하게 고려한다. 보고 단계에서 저자는 항상 관련 바이어스의 가능성을 평가할 것을 권장한다. 특히 바이어스의 방향과 규모를 논의하고 가능하면 추정해야 한다.

박스 3. 바이어스

바이어스는 연구 결과가 실제 값과 체계적으로 벗어나는 것을 말한다. 일반적으로 바이어스는 연구 설계 또는 실행 중에 발생하며 나중에 수정할 수 없다. 바이어스와 교란은 동의어가 아니다. 바이어스는 잘못된 정보 또는 참여자 선택으로 인해 잘못된 연관성이 발견될 때 발생한다. 교란은 사실적으로는 맞지만 설명되지 않은 근본적인 요인이 노출과 결과 모두와 연관되어 있기 때문에 인과적으로 해석할 수 없는 관계를 만든다(박스 5 참조). 또한 바이어스는 측정된 자료의 통계적 변동(어느 방향이든)으로 인해 발생하는 실제 값과의 편차, 즉 무작위 오류와 구별해야 한다. 가능한 많은 바이어스의 원인이 설명되어 있으며 다양한 용어가 사용된다[68,69]. 정보 바이어스와 선택 바이어스라는 두 가지 간단한 범주가 도움이 될 수 있다.

정보 바이어스는 자료의 완전성 또는 정확성의 체계적 차이로 인해 노출 또는 결과와 관련하여 개인을 차별적으로 잘못 분류할 때 발생한다. 예를 들어, 당뇨병 여성이 더 정기적이고 철저한 안과 검사를 받는다면 당뇨병이 없는 여성보다 녹내장 확인이 더 많이 일어날 것이다(항목 9 참조)[65]. 비특이적 위장 불편을 유발하는 약물을 복용하는 환자는 약물이 더 많은 궤양을 유발하지 않더라도 약물을 복용하지 않는 환자보다 위 내시경 검사를 더 자주 받고 더 많은 궤양을 발견 할 수 있다. 이러한 유형의 정보 바이어스는 '탐지 바이어스' 또는 '의료 감시 바이어스'라고도 한다. 그 영향을 평가하는 한 가지 방법은 여러 연구 그룹에서 의료 감시의 강도를 측정하고 통계 분석에서 이를 보정하는 것이다. 환자대조군연구에서 정보 바이어스는 환자군에서 자가 해당 질병이 없는 대조군보다 과거 노출을 어느 정도 정확하게 회상하거나 보고할 의향이 있는 경우 발생한다('회상 바이어스'라고도 함). '면접관 바이어스'는 면접관이 연구 가설을 인지하고 무의식적 또는 의식적으로 자료를 선택적으로 수집하는 경우 발생할 수 있다[70]. 따라서 연구 참여자와 연구자를 어떤 형태로든 눈가림 처리하는 것이 유용할 수 있다.

환자대조군연구에서 환자 또는 대조군을 포함할 확률이 노출과 관련이 있는 경우 선택 바이어스가 발생할 수 있다. 예를 들어, 심부정맥 혈전증 연구를 위해 참가자를 모집하는 의사가 다리에 불편함이 있고 경구 피임약을 복용하는 여성에게서 이 질환을 진단할 수 있다. 그러나 피임약을 복용하지 않는 비슷한 증상을 가진 여성에게는 심부정맥 혈전증을 진단하지 않을 수도 있다. 이러한 바이어스는 진단 서비스에 동일한 방식으로 의뢰된 환자와 대조군을 사용하여 대응할 수 있다[71]. 마찬가지로, 질병 등록부를 사용하면 노출과 질병 사이의 가능한 관계가 알려진 경우, 의심되는 원인 물질에 노출된 환자가 등록부에 제출될 가능성이 더 높을 수 있다[72]. '응답 바이어스'는 연구에 응답한 사람과 참여를 거부한 사람 사이의 특성 차이가 유병률, 발병률, 일부 상황에서는 연관성 추정치에 영향을 미치는 경우 발생하는 또 다른 유형의 선택 바이어스이다. 일반적으로 선택 바이어스는 연구의 내부 타당도에 영향을 미친다. 이는 일반적으로 연구 참여자를 선정할 때 발생할 수 있는 문제와는 다르며, 이는 연구의 내적 타당성보다는 외적 타당성에 영향을 미친다(21번 항목 참조).

예를 들어, 환자대조군연구에서 정보 바이어스가 발생할 수 있지만, 첫 번째 예에서와 같이 적절한

대조군을 선택하면 바이어스를 줄일 수 있다[64]. 두 번째 예에서는 참가자 의료 감시의 차이가 문제가

되었다 [65]. 따라서 이 문제를 해결하기 위해 수집한 추가 자료에 대해 더 자세히 설명해야 한다.

연구자가 추적 연구에서 변수 측정에서 발생할 수 있는 '드리프트'(drift)에 대응하거나 관찰자가 다수인

경우 변동성을 줄이기 위한 자료 수집 품질 관리 프로그램을 설정한 경우, 이를 설명해야 한다.

안타깝게도 결과를 보고할 때 중요한 바이어스를 언급하지 않는 경우가 많다. 암 병력이 있는 환자의 두

번째 암 발생 위험을 조사한 1990년부터 1994년까지 발표된 43건의 환자 대조군 및 코호트연구 중 의료 감시 바이어스가 언급된 논문은 5건에 불과했다[66]. 1998년에 3개의 정신의학 저널에 발표된 정신 건강 연구 보고서를 조사한 결과, 392개 논문 중 13%만이 응답 바이어스에 대해 언급했다[67]. 뇌졸중 연구의 코호트연구 조사에 따르면 1999년부터 2003년까지 발표된 49편의 논문 중 14편(28%)이 연구 참여자 모집 시 잠재적인 선택 바이어스에 대해 언급했으며 35편(71%)은 모든 유형의 바이어스가 결과에 영향을 미칠 수 있는 가능성을 언급했다[5].

10. 연구 규모: 어떻게 연구규모에 도달했는지를 설명한다.

예시 1

"연구 기간 동안 해당 지역의 환자 수가 표본 크기를 결정했다"[73].

예시 2

"해당 지역의 산후 우울증에 대한 설문 조사에서 19.8%의 유병률을 기록했다. 우울증이 있다고 가정할 때 정상 체중의 자녀를 둔 산모의 우울증을 20%로 가정하고 영양실조 아동을 둔 산모의 우울증에 대한 오즈비를 3으로 가정하면 80%의 검정력과 5%의 유의도를 가진 72개의 환자-대조군 세트(한 환자 당 한 명의 대조군)가 필요하다."[74].

설명

연구는 연구 질문에 의미 있게 답하기 위해 충분히 좁은 신뢰 구간으로 점 추정치를 얻을 수 있을 만큼 충분히 커야 한다. 작은 연관성과 연관성이 없는 것을 구별하려면 큰 표본이 필요하다. 소규모 연구는 종종 가치 있는 정보를 제공하지만, 신뢰 구간이 넓으면 신뢰 구간이 좁은 추정치를 제공하는 연구와 비교하여 현재 지식에 기여하는 바가 적을 수 있다. 또한 '흥미로운' 또는 '통계적으로 유의미한' 연관성을 보여주는 소규모 연구가 '유의미한' 결과가 없는 소규모 연구보다 더 자주 발표된다. 이러한 연구는 발견의 맥락에서 초기 신호를 제공할 수 있지만, 독자에게 잠재적인 약점에 대해 알려야 한다.

관찰 연구에서 표본 크기 결정의 중요성은 상황에 따라 다르다. 다른 목적으로 이미 사용 가능한 자료를 분석하는 경우, 자료 분석이 문헌에 실질적으로 기여할 수 있는 충분한 통계적 정밀도를 가진 결과를 산출할 수 있는지 여부가 주요 질문이며, 표본 크기 고려는 비공식적으로 이루어질 것이다. 새로운 연구를 계획할 때 공식적이고 선형적인 표본 크기 계산이 유용할 수 있다[75,76]. 이러한 계산은 일반적으로 생성되는 단일 숫자가 암시하는 것보다 더 많은 불확실성과 관련이 있다. 예를 들어, 관심 사건의 비율 또는 계산의 중심이 되는 기타 가정에 대한 추정치는 일반적으로 추측이 아니라면 부정확하다[77]. 최종 분석에서 얻은 정밀도는 다변량 분석에서 교란 변수를 포함하거나[78], 주요 변수를 측정할 수 있는 정밀도[79], 일부 개인을 제외함으로써 감소하기 때문에 사전에 결정할 수 없는 경우가 많다.

역학 연구에서 표본 크기에 대한 내용을 자세히 설명한 경우는 거의 없다[4,5]. 조사자가 적절한 공식 표본 크기 계산을 수행한 경우 이를 보고할 것을 권장한다. 다른 상황에서는 연구 규모를 결정한 고려 사항(예: 위의 첫 번째 예에서와 같이 사용 가능한 표본이 고정되어 있음)을 명시해야 한다. 통계적 유의성에 도달하기 전에 관찰 연구를 조기에 중단한 경우, 독자에게 이 사실을 알려야 한다. 연구 규모 또는 후향적 검정력 계산에 대한 사후 정당화로 독자를 귀찮게 하지 말아야 한다[77]. 독자의 관점에서 신뢰 구간은

궁극적으로 얻은 통계적 정밀도를 나타낸다 신뢰 구간은 통계적 불확실성에만 영향을 미치며 연구에 존재할 수 있는 모든 불확실성을 반영하지는 않는다는 점을 인식해야 한다(20번 항목 참조).

11. 양적 변수: 분석에서 양적변수를 어떻게 다루었는지 설명한다. 가능하다면 어떤 그룹화를 선택했고 그 이유가 무엇인지 기술한다.

예시

"글래스고 코마 척도가 8 미만인 환자는 심각한 부상을 입은 것으로 간주되고, 9 이상이면 뇌 손상이 덜 심각한 것으로 간주된다. 이 두 범주의 GCS와 부상 후 12개월 이내 사망 발생의 연관성을 조사했다."[80].

설명

연구자는 노출, 효과 변경인자 및 교란 요인에 대한 정량적 자료를 수집하고 분석하는 방법을 선택한다.

예를 들어, 연속 노출 변수를 그룹화하여 새로운 범주형 변수를 만들 수 있다(박스 4 참조). 그룹화 선택은 이후 분석에 중요한 영향을 미칠 수 있다[81,82]. 범주 수, 분절점, 범주 평균 또는 중앙값 등 정량 자료를 그룹화한 이유와 방법을 설명하는 것이 좋다. 자료를 표 형식으로 보고할 때는 각 범주별로 환자 수, 대조군 수, 고위험군 수, 위험에 노출된 시간 등을 제시해야 한다. 표는 효과 측정 추정치 또는 모델 피팅 결과만으로 구성되어서는 안된다.

박스 4. 그룹화

연속 자료를 그룹화할 수 있는 몇 가지 이유가 있다[86]. 자료를 수집할 때 몇 년에 걸친 회상을 기반으로 노출을 인위적으로 연속 측정값을 구하는 것보다 서수 변수를 사용하는 것이 더 나을 수 있다. 범주는 모든 변수를 유사한 스타일로 제시하거나 용량-반응 관계를 제시하는 등 프레젠테이션에도 유용할 수 있다.

그룹화는 분석을 단순화하기 위해(예: 선형성 가정을 피하기 위해) 수행될 수도 있다. 그러나 그룹화하면 정보가 손실되고 통계적 검정력이 떨어질 수 있다[87]. 특히 이분화를 사용할 경우 더욱 그렇다[82,85,88].

연속적인 교란 변수를 그룹화할 경우, 변수의 교란 효과 중 일부가 보정되지 않은 채로 남아있는 잔존 교란이 발생할 수 있다(박스 5 참조)[62,89]. 범주 수를 늘리면 검정력 손실과 잔존 교란을 줄일 수 있으며, 특히 대규모 연구에 적합하다. 소규모 연구에서는 제한된 수로 인해 그룹을 거의 사용할 수 없다.

연구자는 진단 또는 예후와 관련이 있거나 실용성 또는 통계적 근거에 따라 일반적으로 사용되는 값을 기준으로 그룹화를 위한 분절점을 선택할 수 있다. 사분위수를 사용하여 각 그룹에 동일한 수의 개인을 선택할 수도 있다[90]. 반면에 더 극단적인 외부 그룹을 선택하고 중간 그룹을 외부 그룹보다 크게 설정하여 결과와의 연관성에 대한 더 많은 통찰력을 얻을 수 있다[91]. 환자대조군연구의 경우, 대조군에서 분포를 도출하는 것이 모집단을 반영하기 위한 것이므로 선호된다. 여러 대안 중에서 사후에 분절점을 선택하는 경우 독자에게 알려야 한다. 특히, P값을 최소화하기 위해 분절점을 선택한 경우 실제 연관성의 강도가 과장될 수 있다[81].

그룹화된 변수를 분석할 때는 변수의 기본 연속성을 인식하는 것이 중요하다. 예를 들어, 정렬된 그룹에 걸쳐 위험의 가능한 추세를 조사할 수 있다. 일반적인 접근 방식은 그룹의 순위를 연속 변수로 모델링하는 것이다. 그룹 점수의 이러한 선형성은 그룹 간 간격이 동일한 경우(예: 10세 연령 그룹) 실제 선형 관계에

가까워지지만 그렇지 않은 경우에는 그렇지 않다. Il'yasov 등은[92] 메타분석을 용이하게 하고 용량-반응에 대한 본질적으로 가치 있는 정보를 제공하기 위해 표준 오차와 함께 범주형 및 연속형 효과 추정치를 모두 공개할 것을 권장한다. 한 분석이 다른 분석에 정보를 제공할 수 있으며 어느 쪽도 가정이 없는 것은 아니다. 저자는 종종 순서를 무시하고 참조 범주와 비교하여 각 범주에 대한 추정치(및 P값)를 개별적으로 고려한다. 이 방법은 설명에는 유용할 수 있지만 그룹 간 위험의 실제 추세를 감지하지 못할 수 있다. 추세가 관찰되는 경우 기울기에 대한 신뢰 구간은 관찰의 강도를 나타낼 수 있다.

조사자는 모든 정보를 유지하기 위해 노출을 연속적인 것으로 모델링할 수 있다. 모든 정보를 유지하기 위해, 이러한 선택을 할 때는 노출과 결과의 관계의 특성을 고려해야 한다. 선형 관계를 자동으로 가정하는 것은 잘못된 것일 수 있으므로 선형성에서 벗어날 수 있는 가능성을 조사해야 한다. 작성자는 분석 중에 탐색한 대체 모델(예: 로그 변환, 이차 항 또는 스플라인 함수 사용)을 언급할 수 있다. 노출과 결과 사이의 비선형 관계를 추정하는 방법에는 여러 가지가 있다[82-84]. 또한 주요 관심사인 정량적 노출에 대해 연속 분석과 그룹 분석을 모두 제시하는 것이 유익할 수 있다.

최근 조사에 따르면, 역학 연구 논문의 3분의 2가 정량적 노출 변수를 연구했다[4]. 50개 논문 중 42개(84%) 논문에서 노출은 몇 가지 정렬된 범주로 분류되었지만, 선택에 대한 근거가 명시되지 않은 경우가 많았다. 15개의 논문에서 선형 연관성을 사용하여 연속 노출을 모델링했지만, 선형성을 확인했다고

보고한 논문은 2개에 불과했다. 또 다른 조사에서는 심리학 문헌 중 이분화가 정당화된 논문은 110개 중 22개(20%)에 불과했다[85].

12. 통계적 방법:

12 (a) 교란 통제 방법을 포함해서 모든 통계적 방법을 밝힌다.

예시

"비교 그룹에 연령 또는 성별에 따른 교란이 존재하는지 평가하기 위해 만텔-헨젤 기법을 사용하여 보정 상대 위험도를 계산했다. 95% 신뢰 구간은 Greenland and Robins and Robins 등에 따른 분산값으로 보정 상대위험도를 계산하였다. [93].

설명

일반적으로 올바른 통계 분석은 하나만 있는 것이 아니며, 같은 질문을 다루지만 서로 다른 가정을 전제로 하는 여러 가지 가능성이 존재한다고 할 수 있다. 그럼에도 불구하고, 연구 프로토콜에서 최소한 주요 연구 목표에 대한 분석 방법을 미리 결정해야 한다. 종종 원래 예상했던 분석 대신 또는 추가 분석이 필요한 경우가 있으며, 이러한 분석은 자료의 내용에 따라 이루어진다. 연구를 보고할 때 저자는 독자에게 특정 분석이 자료를 보고 하게 되었는지 여부를 알려야 한다. 사전 지정 분석과 탐색적 분석의 구분이 때때로 모호할 수 있지만, 저자는 특정 분석에 대한 이유를 명확히 밝혀야 한다.

비교 대상 그룹이 일부 특성과 관련하여 유사하지 않은 경우, 계층화 또는 다변량 회귀분석을 통해 가능한 교란 변수를 보정해야 한다(박스 5 참조)[94]. 종종 연구 설계에 따라 어떤 유형의 회귀 분석을 선택할지

결정된다. 예를 들어, 코호트연구에서는 콕스 비례 위험 회귀가 일반적으로 사용되는 반면[95], 환자 대조군 연구에서는 로지스틱 회귀가 자주 선택되는 방법이다[96,97]. 최종 모형의 결과만 제시할 것이 아니라 변수 선택에 대한 구체적인 절차를 충분히 설명해야 한다[98,99]. 최종 모델에 포함할 잠재적 교란변수 목록을 좁히기 위해 모델을 비교하는 경우, 이 과정을 설명해야 한다. 하나 또는 두 개의 공변량이 자료 분석에서 명백한 교란의 상당 부분을 차지하는지 독자에게 알려주는 것이 도움이 된다. 결측값 대체 절차, 자료 변환, 기여 위험 계산과 같은 기타 통계 분석에 대해서도 설명해야 한다. 표준이 아니거나 새로운 접근법을 참조하고 사용된 통계 소프트웨어를 보고해야 한다. 기본 원칙으로, 통계적 방법은 '원본 자료에 접근할 수 있는 지식이 있는 독자가 보고된 결과를 확인할 수 있도록 충분히 상세하게 설명하는 것이 좋다'[100].

한 실증 연구에서 교란 보정을 보고한 169개 논문 중 93개(55%)만이 연속형 및 다범주 변수를 통계 모델에 입력하는 방법을 명확하게 명시했다[101]. 또 다른 연구에 따르면 교란 변수에 대한 통계 분석이 보정된 67건의 논문에서 교란 변수가 어떻게 선택되었는지 대부분 불분명했다[4].

12 (b). 하위그룹과 교호성을 평가한 모든 방법에 대해 기술한다.

예시

"연구된 3가지 생활습관 관련 위험 요인에 대한 감수성의 성별 차이는 Rothman에 따라 생물학적 교호성을 테스트하여 탐색했다. 4가지 범주(a^-b^- , a^-b^+ , a^+b^- , and a^+b^+)의 새로운 복합 변수를 성별과 관심 있는 이분법적 노출에 대해 재정의했다. a-, b-는 노출이 없음을 나타낸다 연령에 대한 보정 후 범주별로

상대위험도를 계산하였다. 교호성 효과는 절대 효과의 가산성(additivity)에서 벗어난 것으로 정의하고,

교호성으로 인한 초과 RR(RERI)을 계산했다:

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

여기서 $RR(a+b)$ 는 두 요인에 모두 노출된 사람의 상대위험도이며, $RR(a-b)$ 는 참조범주였다($RR = 1.0$).

95% CI는 호스머와 르메쇼가 제안한 대로 계산했다. RERI가 0이면 교호성이 없음을 의미한다[103].

설명

17번 항목에서 자세히 설명한 것처럼, 연구 모집단의 하위 그룹에 국한된 분석의 사용과 가치에 대해 많은 논쟁이 있다[4,104]. 그럼에도 불구하고 하위 그룹 분석은 종종 수행된다[4]. 독자는 어떤 하위 그룹 분석이 사전에 계획되었고, 어떤 하위 그룹 분석이 자료를 분석하는 동안 발생했는지 알아야 한다. 또한 그룹 간에 효과 또는 연관성이 다른지 여부를 조사하기 위해 어떤 방법을 사용했는지 설명하는 것도 중요하다(17번 항목 참조).

교호성은 한 요인이 다른 요인의 효과를 조절하는 상황과 관련이 있다(따라서 '효과 변경'이라고도 함). 두 요인의 공동 작용은 두 가지 방식으로 특성화할 수 있다. 가산 척도에서는 위험 차이로, 승수 척도에서는 상대위험도로 특성화할 수 있다(박스 8 참조).

박스 5. 교란

교란은 말 그대로 효과의 혼란을 의미한다. 어떤 연구에서 노출과 질병 위험 사이에 연관성이 있거나 연관성이 없는 것처럼 보일 수 있다. 연관성이 있거나 없는 것처럼 보이는 것은 실제로 그럴 수도 있지만, 노출과 연관된 다른 요인이 있기 때문일 수 있다. 이러한 다른 요인을 교란 요인 또는 교란자(confounder)라고 한다. 따라서 교란 요인은 노출의 잠재적인 '인과적' 연관성에 대한 잘못된 평가를 제공한다. 예를 들어, 중년이 되어 혈압이 높아지는 여성이 경구 피임약을 덜 처방받는 경우, 피임약을 사용하는 여성과 그렇지 않은 여성의 심혈관 질환 발생 빈도를 단순 비교하면 피임약이 심장 질환을 예방한다는 잘못된 인상을 줄 수 있다.

잠재적인 교란 요인에 대해 미리 고려해야 한다. 이를 통해 연구 설계에 정보를 제공하고 자세한 정보를 찾아야 하는 교란 요인을 식별하여 적절한 자료를 수집할 수 있다. 제한 또는 짝짓기를 사용할 수 있다. 위의 예에서 교란 요인인 혈압이 높지 않은 여성으로 제한할 수 있다. 혈압을 기준으로 짝짓기하는 것도 가능하지만 반드시 바람직한 것은 아니다(박스 2 참조). 분석 단계에서 연구자는 층화 또는 다변량 분석을 사용하여 교란 변수의 영향을 줄일 수 있다. 층화 분석은 자료를 교란 요인에 대한 계층(예: 혈압 계층)으로 나누고, 각 계층 내에서 연관성 추정치를 평가하고, 모든 계층에 대한 가중 평균으로 결합된 연관성 추정치를 계산하는 것으로 구성된다. 다변량 분석은 동일한 결과를 얻을 수 있지만 더 많은 변수를 동시에 고려할 수 있다. 다변량 분석은 더 유연하지만 노출과 질병 사이의 관계에 대한 수학적 형태에 대한 추가적인 가정을 포함할 수 있다. 관찰 연구에서는 교란변수를 고려하는 것이 중요하지만, 교란변수를 보정한 분석이 연관성의 '인과적 부분'을 확립한다고 가정해서는 안 된다. 잔존 교란(교란을 통제하려는 시도가 실패한 후에도 남아있는 교란[102]), 무작위 추출 오류, 선택 바이어스, 정보 바이어스로 인해 결과가 여전히 왜곡될 수 있다(박스 3 참조).

12 (c). 결측치를 어떻게 다루었는지 기술한다

예시

"결측 자료 분석 절차에서는 무작위 결측(MAR) 가정을 사용했다. STATA에서 다중 다변량 결측값 대체의 MICE(연쇄 방정식에 의한 다변량 결측값 대체) 방법을 사용했다. 다변량 로지스틱 회귀 분석에서 결측값이 적절히 결측값 대체된 자료 사본 10개를 각각 독립적으로 분석했다. 변수 추정치의 평균을 구하여 단일 평균 추정치를 구하고 루빈의 규칙에 따라 표준 오차를 보정했다[106].

설명

박스 6. 결측된 자료: 문제점 및 가능한 해결 방법

결측된 자료를 처리하는 일반적인 접근 방식은 특정 분석에 필요한 모든 변수에 대한 완전한 자료를 가진 개인으로 분석을 제한하는 것이다. 이러한 '완전한 환자' 분석은 많은 상황에서 편향되지 않지만, 편향될 수 있으며 항상 비효율적이다[108]. 자료가 결측된 개인이 전체 표본의 전형이 아닐 경우 바이어스가 발생한다.

비효율성은 분석을 위한 표본 크기가 줄어들기 때문에 발생한다.

반복 측정에 마지막 관측치를 이월하여 사용하면 결과의 예고를 경험한 사람이 선택적으로 탈락할 경우 시간이 지남에 따라 추세가 왜곡될 수 있다[109]. 교란 요인에 대해 결측된 범주 지표를 삽입하면 잔존

교란이 증가할 수 있다[107]. 각 결측값을 가정 또는 추정값으로 대체하는 추정은 관심 있는 연관성을 약화시키거나 과장할 수 있으며, 아래에 설명된 정교한 방법을 사용하지 않으면 표준 오차가 너무 작아질 수 있다.

루빈은 관측값이 결측될 확률에 대한 모델을 기반으로 결측 자료 문제의 유형학을 개발했다[108,110]. 특정 관측값이 결측될 확률이 관측 가능한 변수의 값에 의존하지 않는 경우 자료는 완전 무작위 결측(MCAR)으로 설명된다. 관찰된 자료가 주어졌을 때 관찰이 결측될 확률이 결측된 자료의 실제 값과 무관한 경우 자료는 무작위 결측(MAR)이라고 한다. 예를 들어, 어린 아이일수록 폐활량 측정값이 결측되기 쉽지만, 연령을 고려한 후 결측 확률이 실제 관찰되지 않은 폐 기능과는 무관하다고 가정한다. 그러면 연령을 포함한 모델에서 결측된 폐 기능 측정값은 MAR이 된다. 사용 가능한 자료를 고려한 후에도 결측 확률이 여전히 결측된 값에 의존하는 경우 자료가 무작위가 아닌 결측(MNAR)이다. 자료가 MNAR인 경우 유효한 추론을 위해서는 자료 결측을 초래한 메커니즘에 대한 명시적인 가정이 필요하다.

무작위 자료 결측(MAR)을 처리하는 방법은 크게 세 가지로 분류된다[108,111]: 확률 기반 접근법[112], 가중치 추정[113], 다중 결측값 대체[111,114]. 이 세 가지 접근법 중 다중 결측값 대체는 가장 일반적으로 사용되며 특히 여러 변수에 결측값이 있는 경우 유연성이 뛰어나다[115]. 이러한 접근법을 사용한 결과는 전체 환자 분석의 결과와 비교하고 중요한 차이점을 논의해야 한다. 결측 자료 분석에서 가정한 타당성은 일반적으로 검증할 수 없다. 특히 자료가 MNAR이 아니라 MAR이라는 것을 증명하는 것은 불가능하다. 따라서 이러한 분석은 민감도 분석의 정신으로 보는 것이 가장 좋다(항목 12e 및 17 참조).

자료 결측은 관찰 연구에서 흔히 발생한다. 연구 참여자에게 보내는 설문지가 항상 완전하게 작성되는 것은 아니며, 참여자가 모든 후속 방문에 참석하지 않을 수도 있고, 일상적인 자료원 과 임상

데이터베이스가 불완전한 경우가 많다. 자료 결측의 보편성과 중요성에도 불구하고 자료 결측 문제에 대해 자세히 보고한 논문은 거의 없다[5,107]. 연구자는 결측된 자료를 해결하기 위해 여러 접근법 중 하나를 사용할 수 있다. 박스 6에서 다양한 접근법의 장점과 한계를 설명한다. 저자는 각 관심 변수(노출, 결과, 교란 요인)와 분석의 각 단계별로 결측된 값의 수를 보고할 것을 권장한다. 저자는 가능한 경우 결측값에 대한 이유를 제시하고, 연구 참여자의 흐름을 설명할 때 자료 결측으로 인해 제외된 개인 수를 명시해야 한다(13번 항목 참조). 결측된 자료를 설명하는 분석의 경우, 저자는 분석의 특성(예: 다중 결측값 대체)과 가정(예: 무작위 결측, 박스 6 참조)을 설명해야 한다.

12 (d). 코호트연구—추적관찰 소실을 어떻게 다루었는지 기술한다.

예시

"적극적인 추적관찰을 시행한 치료 프로그램에서 추적관찰을 중단한 환자와 1년간 추적관찰을 시행한 환자의 기저 CD4 세포 수(는 비슷했지만 중앙값 115 세포/ ℓ vs 123 세포/ ℓ), 적극적인 추적관찰을 시행하지 않은 프로그램에서 추적관찰을 중단한 환자의 CD4 세포 수는 추적관찰을 시행한 환자보다 유의하게 낮았다(중앙값 64 세포/ ℓ vs 123 세포/ ℓ). (... .) 소극적 추적 관찰이 포함된 치료 프로그램은 후속 분석에서 제외되었다"[116].

설명

코호트연구는 개인별 추적 관찰 시간과 관심 질병 발병 시간을 기반으로 하는 생명표 방법이나 기타

접근법을 사용하여 분석한다. 관찰 기간이 종결된 시점에서 질병이 없는 개인 중 추적 관찰 시간은 결과 발생 확률과 관련이 없는 것으로 가정한다. 이는 추적 관찰이 정해진 날짜 또는 특정 연령에 종료되는 경우에 해당된다. 참가자가 해당 날짜 이전에 연구를 철회하면 추적조사 손실이 발생한다. 추적조사 손실이 있는 개인 또는 질병 발병 위험이 높은 사람에게서 선택적으로 발생하는 경우 연구의 유효성을 저해할 수 있다('정보 중도절단', 'informative censoring'). 위의 예에서, 적극적인 추적 관찰이 이루어지지 않은 치료 프로그램에서 추적 관찰이 중단된 환자는 관찰 중인 환자보다 CD4 헬퍼 세포 수가 적었으므로 사망 위험이 더 높았다[116].

연구 종료에 도달한 사람과 추적 관찰에서 탈락한 사람을 구별하는 것이 중요하다. 안타깝게도 통계 소프트웨어는 일반적으로 두 가지 상황을 구분하지 못하며, 두 경우 모두 관찰 기간이 끝나면 추적 관찰 시간이 자동으로 잘린다('중도절단'). 따라서 연구자는 연구 계획 단계에서 추적 조사 손실에 어떻게 대처할지 결정해야 한다.

소수의 환자가 손실된 경우, 연구자는 추적 관찰이 불완전한 개인을 제외하거나 추적 관찰 손실 날짜 또는 연구 종료 시점에 살아있는 상태에서 중도 포기한 것으로 처리할 수 있다. 저자는 추적 관찰에서 손실된 환자 수와 어떤 중도절단 전략을 사용했는지 보고할 것을 권장한다.

12 (d). 환자 대조군 연구: 환자군과 대조군의 짝짓기 방법에 대해 설명한다.

예시

"맥네마 검사, 쌍검정, 조건부 로지스틱 회귀 분석을 사용하여 치매 환자와 대조군을 비교하여 심혈관

위험 요인, 자발적 뇌색전 발생, 경동맥 질환, 정맥 대 동맥 순환 단락에 대해 비교했다'[117].

설명

개별적으로 짝짓기된 환자대조군연구에서 짝짓기를 무시하고 확률 비율을 조잡하게 분석하면 일반적으로 단일성(unity)에 편향된 추정치가 도출된다(박스 2 참조). 따라서 짝짓기 분석이 필요한 경우가 많다. 이는 직관적으로 계층 분석으로 이해할 수 있는데, 각 환자는 일치하는 대조군 집합을 가진 하나의 계층으로 간주된다. 이 분석은 일치하는 변수가 비슷함에도 불구하고 환자가 대조군보다 더 자주 노출되는지 여부를 고려하는 데 중점을 둔다. 조사자는 '일치하는' 2×2 테이블에서 맨텔-헨젤 방법을 사용하여 이러한 계층 분석을 수행할 수 있다. 가장 간단한 형태로 확률 비율은 노출 변수에 대해 불일치하는 쌍의 비율이 된다. 연령 및 성별과 같이 보편적인 속성을 가진 변수에 대해 짝짓기가 수행된 경우, 분석에서 개별적인 개인 간 짝짓기를 유지할 필요는 없으며 연령 및 성별 범주의 간단한 분석으로 충분하다[50]. 그러나 이웃, 형제 관계 또는 친구 관계와 같은 다른 짝짓기 변수의 경우, 짝짓기된 각 집합을 고유한 계층으로 간주해야 한다. 개별 짝짓기 연구에서 가장 널리 사용되는 분석 방법은 조건부 로지스틱 회귀 분석으로, 각 환자와 그 대조군을 함께 고려한다. 조건부 방법은 환자마다 대조군의 수가 다르거나 일치하는 변수 외에 다른 변수를 보정해야 할 때 필요하다. 독자가 분석에서 짝짓기 설계가 적절하게 고려되었는지 판단할 수 있도록 작성자는 자료 분석에 사용된 통계적 방법을 자세히 설명하는 것이 좋다. 짝짓기를 고려해도 추정치에 거의 영향을 미치지 않는 경우, 저자는 짝짓기되지 않은 분석을 제시할 수 있다.

12 (d). 단면연구: 해당되는 경우, 표본추출 전략을 고려한 분석 방법을 설명한다.

예시

"복잡한 표본 설계에 기반한 추정치의 표본 오차를 추정하기 위해 테일러 확장 방법을 사용하여 표준 오차(SE)를 계산했다. (... .) 이완기 혈압에 대한 전체 설계 효과는 남성 1.9, 여성 1.8로 나타났으며 수축기 혈압의 경우 남성 1.9, 여성 2.0으로 나타났다"[118].

설명

대부분의 단면연구는 미리 지정된 추출 전략을 사용하여 모집단에서 참가자를 선택한다. 그러나 추출은 단순한 무작위 표본 추출보다 더 복잡할 수 있다. 여기에는 여러 단계와 참여자의 지역(예: 지역 또는 마을)별 군집 전략이 포함될 수 있다. 비례 계층화는 특정 특성을 가진 하위 그룹이 올바르게 대표되도록 할 수 있다. 불균형 계층화는 특정 관심 하위 그룹을 과도하게 추출하는 데 유용할 수 있다.

복잡한 표본에서 도출된 연관성 추정치는 단순한 무작위 표본에서 도출된 추정치보다 정확도가 다소 떨어질 수 있다. 표준 오차 또는 신뢰 구간과 같은 정밀도 측정은 단순한 무작위 추출 대신 더 복잡한 추출 전략을 사용할 경우 얼마나 많은 정밀도를 얻거나 잃는지를 설명하는 비율 측정치인 설계 효과를 사용하여 보정해야 한다[119]. 대부분의 복잡한 추출 기법은 정밀도를 감소시켜 1보다 큰 설계 효과를 초래한다.

저자는 복잡한 추출 전략을 보정하는 데 사용된 방법을 명확하게 명시하여 독자가 선택한 추출 방법이 얻은 추정치의 정밀도에 어떤 영향을 미치는지 이해할 수 있도록 하는 것이 좋다. 예를 들어 클러스터

추출의 경우, 설계 효과를 보고하면 자료 수집의 용이성과 정밀도 손실 사이의 암묵적인 절충점을

투명하게 알 수 있다. 이 예에서 남성의 설계 효과가 1.9로 계산된 것은 결과 추정치가 동일한 정밀도를

가지려면 실제 표본 크기가 단순 무작위 표본 추출보다 1.9배 더 커야 함을 나타낸다

12 (e). 모든 민감도 분석에 대해 기술한다

예시

"자료가 불충분한 사망 환자(38/148 25.7%)의 비율이 생존 환자(15/437 3.4%)에 비해 상대적으로 높았기

때문에(...), 이로 인해 결과가 편향되었을 가능성이 있다. 따라서 민감도 분석을 수행했다. 연구 그룹에서

경구 피임약을 사용하는 여성의 비율이 전체(사망자 19.1%, 생존자 11.4%)에 적용된다고 가정한 다음,

노출된 결측된 환자가 모두 2세대 피임약을 사용하거나 모두 3세대 피임약을 사용한다는 두 가지 극단적인

시나리오를 적용했다[120].

설명

민감도 분석은 주요 결과가 대체 분석 전략 또는 가정을 통해 얻은 결과와 일치하는지 여부를 조사하는 데

유용한다[121]. 검토할 수 있는 문제에는 분석에 포함할 기준, 노출 또는 결과의 정의[122], 보정이 필요한

교란 변수, 결측 자료 처리[120,123], 노출, 질병 및 기타 변수의 부정확하거나 일관되지 않은 측정으로

인한 선택 바이어스 또는 바이어스 가능성, 양적 변수 처리와 같은 특정 분석 선택(항목 11 참조) 등이

있다. 여러 바이어스나 가정의 영향을 동시에 모델링하기 위해 정교화된 방법이 점점 더 많이 사용되고

있다[124- 126].

1959년 CornField 등은 흡연에 따른 폐암의 상대위험도 9가 비흡연자보다 흡연자에서 최소 9배 이상 유병률이 높아야 하기 때문에 생각할 수 있는 교란 요인으로 인한 결과일 가능성은 극히 낮다는 것을 보여주었다[127]. 이 분석은 그러한 요인이 존재할 가능성을 배제하지는 않았지만, 그러한 요인이 존재해야 하는 유병률을 확인했다. 최근 소아 백혈병과 송전선 근처 거주 사이의 연관성을 설명할 수 있는 그럴듯한 교란 요인을 식별하기 위해 동일한 접근 방식이 사용되었다[128]. 보다 일반적으로 민감도 분석은 연관성을 왜곡하는 데 필요한 교란, 선택 바이어스 또는 정보 바이어스의 정도를 식별하는 데 사용할 수 있다. 민감도 분석의 한 가지 중요한 사용법은 연구 결과 노출과 결과 사이의 연관성이 거의 또는 전혀 나타나지 않고 교란 또는 영(0)에 대한 기타 바이어스가 존재할 가능성이 있는 경우이다.

결과

연구 참여자 모집, 연구 집단에 대한 설명, 주요 결과 및 보조 분석에 이르기까지 발견된 내용에 대해 사실적인 설명이 결과에 보고된다. 저자의 견해와 의견을 반영하는 해석이나 담론적 텍스트가 없어야 한다.

13. 참여자:

13 (a). 각 단계별 인원에 대해 밝힌다 – 잠재적 적격자 수, 적격에 대해 조사한 수, 적격 확인 수, 연구 포함된 수, 추적 완료자 수, 분석자 수

예시

"표본으로 추출한 105개의 독립형 술집과 선술집 중 13개 업소는 더 이상 영업을 하지 않고 9개 업소는 식당에 위치하여 83개의 적격 업소가 남았다. 22개의 경우 6회 이상 시도했지만 소유주와 전화 연락이 닿지 않았다. 36개 술집의 주인은 연구 참여를 거부했다. (...) 25개의 참여 술집과 선술집에는 124명의 바텐더가 근무하고 있었으며, 67명의 바텐더가 주당 최소 1일 주간 근무를 하고 있었다. 주간 바텐더 중 54명 (81%)이 기본 인터뷰와 폐활량 측정을 완료했으며, 이들 중 53명(98%)이 후속 조치를 완료했다 [129].

설명

연구 참여자 모집 과정에 대한 자세한 정보는 여러 가지 이유로 중요하다. 연구에 포함된 사람들은 대상 인구집단과 여러 면에서 다른 경우가 많다. 이로 인해 대상 인구집단의 경험을 반영하지 않는 유병률 또는 발생률 추정치가 나올 수 있다. 예를 들어, 성행위에 대한 우편 설문조사에 참여하기로 동의한 사람들은 거부한 사람들보다 교회에 덜 자주 출석하고, 덜 보수적인 성적 태도를 가지고 있으며, 첫 성관계 연령이 더 어려웠고, 담배를 피우거나, 술을 마실 가능성이 더 높았다[130]. 이러한 차이로 인해 우편 설문조사 결과가 인구의 성적 자유주의와 활동을 과대평가할 수 있다. 이러한 응답 바이어스(박스 3 참조)은 연구 참여자와 실제 연구에 포함된 사람 사이에 연관성이 다를 경우 노출-질병 연관성을 왜곡할 수 있다. 또 다른 예로, 일부 환자대조군연구에서 관찰된 젊은 산모의 나이와 자녀의 백혈병 사이의 연관성은 환자군과 대조군에 젊은 여성이 차별적으로 참여한 것이 원인 일 수 있다[131,132]. 건강한 자녀를 둔 젊은 여성은 건강하지 않은 자녀를 둔 여성보다 참여 가능성이 낮았다[133].

참여율이 낮다고 해서 연구의 타당성이 크게 훼손되는 것은 아니지만, 참여율과 불참 사유에 대한 투명한 정보는 필수적이다. 또한 참여율, 응답률 또는 후속 조치 비율에 대한 보편적으로 합의된 정의가 없기 때문에 독자는 저자가 이러한 비율을 어떻게 계산했는지 이해할 필요가 있다[134].

이상적으로 연구자는 대상 모집단 선정부터 분석에 참여자 자료를 포함시키는 것까지 연구 참여자 모집의 각 단계에서 고려한 개인 수에 대한 설명을 제공해야 한다. 연구 유형에 따라 여기에는 잠재적 적격자로 간주되는 사람의 수, 적격성 평가 인원, 적격자로 판명된 인원, 연구에 포함된 인원, 조사된 인원, 추적 관찰된 인원, 분석에 포함된 인원이 포함될 수 있다. 위의 예에서와 같이 연구 참여자의 추출이 두 단계 이상으로 수행되는 경우(다단계 추출) 다른 추출 단위에 대한 정보가 필요할 수 있다. 환자대조군연구의 경우, 저자는 환자 그룹과 대조군에 대해 참여자의 흐름을 별도로 설명하는 것이 좋다[135]. 대조군은 입원 환자 및 지역사회 거주자 등 여러 출처에서 선택할 수 있다. 이 경우 각 유형의 대조군에 대한 참가자 수를 별도로 설명하는 것이 좋다. 올슨과 동료들은 무작위 전화걸기 및 기타 방법을 통해 모집한 대조군에 대한 유용한 보고 지침을 제안했다[136].

최근 10년에 발표된 역학 연구에 대한 설문조사에 따르면 일반 역학, 공중 보건 및 의학 저널에 발표된 107건의 환자 대조군 연구 중 47건(59%), 코호트연구 154건 중 49건(32%), 단면연구 86건 중 51건(59%)에서 참여자에 대한 일부 정보가 제공되었다[137]. 역학 연구에서 참여 및 불참에 대한 보고가 불완전하거나 없는 것이 다른 두 건의 문헌 조사에서도 확인되었다[4,5]. 마지막으로, 최근 수십 년 동안 역학 연구 참여가 감소했을 수 있다는 근거가 있으며[137,138], 이는 투명한 보고의 필요성을 강조한다[139].

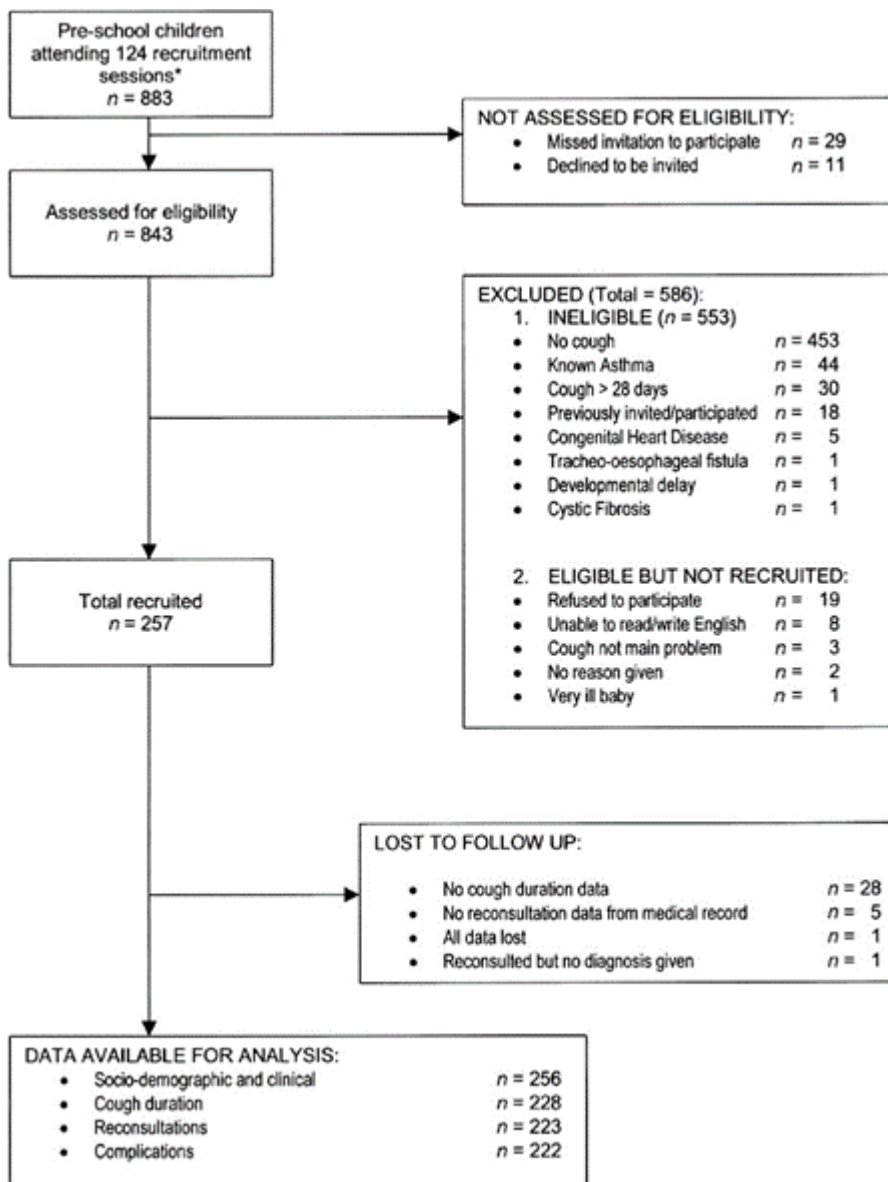
13 (b). 각 단계에서 불참 이유를 제시한다.

예시

"불참의 주요 이유는 참여자가 너무 아프거나 인터뷰 전에 사망 한 경우(환자 30%, 대조군 1%),
무응답(환자 2%, 대조군 21%), 거부(환자 10%, 대조군 29%), 기타 이유(전문의 또는 일반의의 거부,
비영어권, 정신 장애, (환자 7%, 대조군 5%)였다. [140]."

설명

사람들이 더 이상 연구에 참여하지 않은 이유 또는 통계 분석에서 제외된 이유를 설명하면 독자가 연구
집단이 대상 집단을 대표하는지, 바이어스가 있을 가능성이 있는지 판단하는 데 도움이 된다. 예를 들어,
단면 건강 설문조사에서 건강 상태와 관련이 없을 것 같은 이유(예: 잘못된 주소로 인해 초대장이 전달되지
않은 경우)로 인한 불참은 추정치의 정확성에 영향을 미치지만 바이어스를 유발하지는 않는다. 반대로,
질병이나 건강 상태가 양호하다고 인식되어 설문조사에 불참하는 사람이 많으면 조사 결과가 인구의 질병
유병률을 과소평가하거나 과대평가할 수 있다.



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

위의 예에서와 같이 유익하고 잘 구조화된 순서도는 긴 설명이 필요할 수 있는 정보를 쉽고 투명하게

전달할 수 있다[142]. 이 다이어그램에는 사건 수와 같은 주요 결과가 유용하게 포함될 수 있다. 특히

복잡한 관찰 연구의 경우 흐름도를 사용하는 것이 좋지만, 특정 형식을 제안하지는 않는다.

2. 설명 자료:

14 (a). 연구 참여자의 특성(예: 인구통계학적, 임상적, 사회적)과 노출 및 잠재적 교란 요인에 대한 정보를 제공한다.

Example

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

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

HCV, Hepatitis C virus.

^aMales <60 g ethanol/day, females <30 g ethanol/day.

^bMales >60 g ethanol/day, females >30 g ethanol/day.

Table adapted from Osella et al. [143].

설명

독자는 연구 결과의 일반화 가능성을 판단하기 위해 연구 참여자 및 노출에 대한 설명이 필요하다. 잠재적

교란 요인에 대한 정보(측정 여부 및 방법 포함)는 연구 유효성에 대한 판단에 영향을 미친다. 각 연구

그룹의 연속형 변수는 평균과 표준 편차를 제공하거나 자료가 비대칭 분포인 경우 중앙값과 백분위수 범위(예: 25번째 및 75번째 백분위수)를 제시하여 요약하는 것이 좋다. 소수의 범주(예: 질병의 1단계에서 4단계까지)를 연속형 변수로 제시해서는 안 되며, 각 범주에 대한 숫자와 비율을 제시하는 것이 바람직하다(박스 4 참조). 그룹을 비교하는 연구에서는 위의 예와 같이 그룹별로 기술적 특성과 숫자를 제시해야 한다.

표준 오차 및 신뢰 구간과 같은 추론적 측정값은 특성의 변동성을 설명하는 데 사용해서는 안 되며, 기술적(descriptive) 표에서 유의성 검정은 피해야 한다. 또한 P값은 분석에서 보정할 교란변수를 선택하는 데 적절한 기준이 아니며, 결과에 큰 영향을 미치는 교란변수의 작은 차이도 중요할 수 있다[144,145].

코호트연구에서는 노출이 다른 특성 및 잠재적 교란 요인과 어떻게 관련되는지 문서화하는 것이 유용할 수 있다. 환자대조군연구에서는 환자군과 대조군을 비교하여 잠재적 교란 요인을 판단할 수 없다. 대조군은 원천(source) 집단을 대표하며 일반적으로 여러 측면에서 환자군과 다를 수 있다. 예를 들어 경구 피임약과 심근경색 관련성에 대한 연구에서 심근경색이 있는 젊은 여성 표본은 대조군보다 높은 혈청 콜레스테롤, 흡연, 양성 가족력 등 해당 질환의 위험 요인을 더 자주 가지고 있었다[146]. 경구 피임약 처방이 이러한 위험 요인의 존재에 따라 이루어지지 않는 한(예: 위험 요인이 사건 이후에야 확인되었기 때문에) 경구 피임약의 효과 평가에 영향을 미치지 않는다(박스 5 참조). 환자대조군연구에서 잠재적 교란 요인의 존재 여부에 대해 노출된 그룹과 노출되지 않은 그룹을 동등하게 비교하려면, 대조군이 충분히 크고, 원천 모집단을 대표해야 한다[121,147].

14 (b). 관심 있는 각 변수에 대해 결측된 자료가 있는 참가자 수를 제시한다.

예시

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

Table adapted from Hay et al. [141].

설명

결측된 자료는 결과의 일반화 가능성에 바이어스를 초래하거나, 영향을 줄 수 있으므로 저자는 노출, 잠재적 교란 요인 및 기타 환자의 중요한 특성에 대한 결측된 자료의 양을 독자에게 알려야 한다(항목 12c 및 박스 6 참조). 코호트연구의 경우, 불완전한 추적조사는 연구 결과에 바이어스를 초래할 수 있으므로, 저자는 추적조사 손실 정도를 이유와 함께 보고해야 한다(항목 12d 및 13 참조)[148]. 표와 그림을 사용하여 결측된 자료의 양을 열거할 것을 권장한다.

14 (c). 코호트연구: 추적 관찰 기간(예: 평균 및 총기간)을 요약한다.

예시

"4366년(중앙값 5.4년, 최대 8.3년)의 추적 관찰 기간 동안 알츠하이머병 202명을 포함하여 265명의
피험자가 치매 진단을 받았다."[149].

설명

독자는 사용 가능한 결과 자료에 대한 추적 관찰 기간과 범위를 알아야 한다. 저자는 평균 또는 중앙값
또는 둘 모두를 사용하여 평균 추적 관찰 기간에 대한 요약을 제시할 수 있다. 독자는 평균에 연구 참여자
수를 곱하여 총 인년 수를 계산할 수 있다. 또한 저자는 독자에게 추적 관찰 시간의 분포를 보여주기 위해
최소 및 최대 시간 또는 분포의 백분위수를 제시할 수 있다. 총 추적 관찰 기간(인년)을 보고하거나 수집된
잠재적 자료의 비율을 제시할 수도 있다[148]. 이러한 모든 정보는 두 개 이상의 노출 범주에 속한
참가자에 대해 별도로 제시될 수 있다. 암 저널에 실린 132편의 논문(대부분 코호트연구) 중 거의 절반이
추적 관찰 기간에 대한 요약을 제공하지 않았다[37].

15 결과 자료:

코호트연구: 시간 경과에 따른 사건 발생 건수 또는 요약 측정값을 보고한다.

예시

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

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

CI, confidence interval.

Table adapted from Kengeya-Kayondo et al. [150].

환자 대조 연구: 각 노출 범주의 보고 번호 또는 노출 요약 측정값을 제시한다.

Example

Table. Exposure among Liver Cirrhosis Cases and Controls

	Cases (<i>n</i> = 40)	Controls (<i>n</i> = 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].

단면연구: 결과 발생 건수 또는 요약 측정값을 보고한다.

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

Example

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

Categorized <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].

설명

노출(위험 요인)과 결과 사이의 가능한 연관성을 다루기 전에 저자는 관련된 기술자료를 보고해야 한다.

기술 자료를 나타내는 동일한 표에 연관성 측정치를 제시하는 것이 가능하고 의미 있을 수 있다(항목 14a

참조). 사건 발생을 결과로 하는 코호트연구의 경우, 관심 있는 각 결과에 대한 총 사건 수 혹은 1인당

사건 발생률을 보고하는 것을 고려한다. 추적 관찰 시간에 따라 사건의 위험이 변하는 경우, 사건의 수와

비율을 적절한 추적 관찰 간격으로 제시하거나 카플란-마이어 수명 표 또는 도표로 제시한다. 시간에 따른

측정값(예: 평균 및 표준 편차)은 표나 그림으로 제시하는 것이 더 나을 수 있다. 단면적 연구의 경우 가장

흔한 결과 사건이나 요약 측정값에 대해 동일한 유형의 정보를 제시하는 것이 좋다. 환자대조군연구의

경우 환자군과 대조군의 빈도 또는 정량적 요약을 별도로 보고하는 데 중점을 둔다[154]. 모든 설계에

대해 자료가 분석되지 않더라도 지속적인 결과 또는 노출을 범주별로 표로 작성하는 것도 도움이 될 수

있다. .

16. 주요 결과

16 a) 비보정 추정치를 제시한다. 가능하면 교란변수 보정 추정치와 정밀도를 제시한다(95% 신뢰구간).

보정 변수를 제시하고 보정변수 선택 이유를 밝힌다.

예시 1

"우리는 처음에 다음과 같은 변수를 만텔-헨첼 계층화 분석에 의한 잠재적 교란변수로 고려했다. (...) 최종 로지스틱 회귀 모델에 포함된 변수는 만텔-헨첼 보정 후 확률 비율에 10%의 변화를 가져온 (...) 변수였다"[155]

Example 2

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

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

설명

많은 경우 비보정 혹은 최소 보정 분석 결과와 완전 보정 분석 결과를 함께 제시할 수 있다. 보정되지

않은 분석 결과를 주요 자료와 함께 제공하는 것이 좋다(예: 노출된 환자 및 대조군 수). 이를 통해 독자는

연관성 측정의 이면에 있는 자료를 이해할 수 있다(15번 항목 참조). 보정 분석의 경우, 공변량에서 결측된

값으로 인해 이 숫자가 달라질 수 있으므로 분석 참여자 수를 보고한다(항목 12c 참조). 추정치는 신뢰

구간과 함께 제공해야 한다. 독자는 비보정연관성 측정값과 잠재적 교란 요인에 대해 보정된 연관성

측정값을 비교하여 얼마나 많이, 어떤 방향으로 변화했는지 판단할 수 있다. 독자는 '보정' 결과가 연관성 주제 측정의 인과적 부분과 동일하다고 생각할 수 있지만, 보정 결과도 추출 오류, 선택 바이어스, 정보 바이어스 또는 잔존 교란이 없다고 할 수는 없다(박스 5 참조). 따라서 보정 결과를 해석할 때 세심한 주의를 기울여야 하며, 결과의 타당성은 종종 중요한 교란변수에 대한 완전한 지식, 정확한 측정, 통계 모델의 적절한 사양(항목 20 참조)이 중요하기 때문이다[157,158]. 저자는 고려된 모든 잠재적 교란변수와 통계 모델에서 변수를 제외하거나 포함시키는 기준을 설명해야 한다. 변수를 제외할지 아니면 포함할지에 대한 결정은 인과 관계에 대한 지식 또는 명시적 가정에 따라 이루어져야 한다. 부적절한 결정, 예를 들어 노출과 질병 사이의 인과 경로에 있는 변수를 포함함으로써 바이어스를 유발할 수 있다(매개 변수에 의한 영향의 정도를 평가하는 것이 목적이 아닌 경우). 모델에 변수를 포함하기로 한 결정이 추정치의 변화에 근거한 것이라면, 어떤 변화가 그 포함을 정당화할 만큼 충분히 중요하다고 간주되었는지 보고하는 것이 중요하다. 교란변수를 선택하기 위해 '후방 삭제' 또는 '전방 포함' 전략을 사용했다면, 그 과정을 설명하고 귀무가설을 거부할 수 있는 유의 수준을 제시해야 한다. 참고로, 통계적 유의성 테스트에만 근거하여 교란변수를 선택하는 것을 권장하지 않는다[147,159,160]. 최근 역학 연구 보고의 질에 대한 연구에 따르면 대부분의 논문에서 신뢰 구간이 보고된 것으로 나타났다[4]. 그러나 교란 변수의 선택에 대해 설명한 저자는 거의 없었다[4,5].

16 (b). 연속변수를 범주화 했으면 범주와 범위를 제시한다.

Example

Table. Polychlorinated Biphenyls in Cord Serum

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

Table adapted from Sagiv et al. [161].

설명

연속 자료를 분류하는 것은 분석에서 중요한 부분이며(박스 4 참조), 결과 제시에도 영향을 미친다. 표에서는 각 그룹(예: 환자 및 대조군)에 대해 개별적으로 관련성이 있는 경우 위험에 노출된 사람, 시간 등 각 노출 범주에 대한 결과를 제공해야 한다. 사용된 범주의 세부 사항은 연구 비교 및 메타 분석에 도움이 될 수 있다. 체질량지수 임계값[162]과 같은 기존의 임계점을 사용하여 자료를 그룹화한 경우, 최고 및 최저 범주를 제외하고 그룹 경계(즉, 값의 범위)를 쉽게 도출할 수 있다. 사분위수 범주를 사용하는 경우, 자료에서 범주 경계를 유추할 수 없다. 최소한 범주 경계는 보고해야 하며, 자료의 범위와 범주 내 평균 또는 중앙값을 보고하는 것도 도움이 된다.

16 (c). 적절하다면, 상대위험도를 절대위험도로 변환한다.

예시

"10년간 HRT[호르몬 대체 요법]를 사용하면 에스트로겐 단독 제제 사용자 1000명당 5건(95% CI 3-7)의

추가 유방암이 발생하고 에스트로겐-프로게스테겐 복합제 사용자 1000명당 19(15-23) 건의 추가 암이 발생하는 것으로 추정된다"[163].

설명

노출과 질병 사이의 연관성을 조사한 연구 결과는 일반적으로 위험도, 비율 또는 확률의 비율 같은 상대적인 수치로 보고된다(박스 8 참조). 상대 측정치는 노출과 질병 사이의 연관성의 강도를 파악한다. 상대위험도가 1에서 멀리 떨어져 있으면 교란으로 인한 연관성일 가능성이 적다[164,165]. 상대 효과 또는 연관성은 절대적 측정치보다 연구와 인구집단 간에 일관성이 있는 경향이 있지만, 종종 특정 환자에서는 연관성이 없을 수도 있다. 예를 들어, 관상동맥 심장 질환의 고전적 위험도는 국가마다 크게 다르지만 북아일랜드, 프랑스, 미국 및 독일에 거주하는 남성의 상대 위험도는 유사하였다[166,167]. 대조적으로, 심혈관 질환 사망률의 위험 요인으로서 고혈압에 대한 연구에서 자료는 일정한 비의 비율보다 위험도 차이가 더 적합했다 [168]. 로지스틱 [169] 및 비례 위험 (Cox) 회귀 [170]를 포함하여 널리 사용되는 통계 모델은 비율 측정에 기반한다. 이러한 모델에서는 비율 효과 측정값의 정합성 이탈만 쉽게 식별할 수 있지만, 상호 작용으로 인한 상대적 초과 위험(RERI, 항목 12b 및 박스 8 참조)과 같이 위험 차이의 추가성 이탈을 평가하는 측정값은 비율 측정값에 기반한 모델에서 추정할 수 있으며, 많은 상황에서 노출과 관련된 절대 위험은 상대 위험보다 더 중요한 관심사이다. 예를 들어, 약물의 부작용에 초점을 맞추는 경우 단위 사용 시간(예: 일, 주 또는 년) 당 추가 환자 수를 알고 싶을 것이다. 위의 예시는 10년간 호르몬 대체 요법을 사용한 여성 1000명당 유방암 추가 발생 건수를 보여준다[163]. 기여 위험 또는 인구 기여 비율과 같은 측정은 노출을 제거할 경우 질병을 얼마나 예방할 수 있는지 측정하는 데

유용할 수 있다. 이러한 측정치는 통계적 불확실성 측정치(예: 예시에서와 같은 신뢰 구간)와 함께 제시하는 것이 바람직하다. 저자는 위험 요인과 질병 사이의 인과 관계를 포함하여 이러한 맥락에 서 만들어진 강력한 가정을 알고 있어야 한다(박스 7 참조)[171]. 의미적 모호성과 복잡성 때문에 저자는 기여성 위험을 계산하는 데 어떤 방법을 사용했는지 자세히 보고해야 하며, 이상적으로는 사용된 공식을 제공해야 한다[172]. 최근 주요 의학 저널에 게재된 222개 논문의 초록을 조사한 결과, 비율을 포함한 무작위 임상시험 초록의 62%에서 절대 위험도가 제공되었지만 코호트연구 초록의 21%에서만 제공된 것으로 나타났다[173]. 1966년부터 1997년까지 Medline의 무료 텍스트 검색에 따르면 제목 또는 초록에 기여 위험을 언급한 항목은 619개였으며, 상대 위험 또는 오즈비를 사용한 항목은 18,955개로 1에서 31의 비율을 사용했다[174].

박스 7. 연관성, 효과 및 영향 측정

관찰 연구는 인구집단에서 건강 문제의 규모와 분포를 설명하기 위해서 수행될 수 있다. 특정 시점에 질병에 걸린 사람의 수(유병률) 또는 정해진 기간 동안 질병에 걸린 사람의 수(발병률)를 조사할 수 있으며, 발병률은 질병에 걸린 사람의 비율(누적 발병률) 또는 추적 관찰 기간 동안의 1인당 비율(발병률)로 표현할 수 있다. 다양한 발생률을 설명하기 위해 사망률, 출생률, 발병률, 환자 사망률 등의 특정 용어가 사용된다. 마찬가지로, 점 유병률 및 기간, 연간 또는 평생 유병률과 같은 용어는 다양한 유형의 유병률을 설명하는 데 사용된다 [30]. 다른 종류의 관찰 연구는 인과 관계를 다룬다. 이러한 연구들은 조사 대상 위험 요인에 노출된 사람과 노출되지 않은 사람 사이의 관심 대상 사건의 위험, 비율 또는 유병률을 비교하는 데 중점을 둔다.

이러한 연구는 종종 '상대위험도'를 추정하는데, 이는 위험비(누적 발생률의 비율)와 비율(발생률의 비율)을 의미할 수 있다. 환자대조군연구에서는 원인 집단(대조군)의 일부만 포함된다. 결과는 환자군과 대조군 간의 노출 확률의 비율로 표현된다. 이 오즈비는 환자와 대조군의 추출에 따른 위험 또는 비율 비율의 추정치를 제공한다(박스 1 참조)[175,176]. 단면연구의 유병률 또는 유병률 오즈비는 일부 상황에서 유용할 수 있으며[177], 결과를 상대적 및 절대적 용어로 표현하는 것이 종종 도움이 될 수 있다. 예를 들어, 영국 남성 의사를 대상으로 한 연구에서 50년 추적 관찰 기간 동안 폐암으로 인한 사망률은 흡연자의 경우 연간 10만 명당 249명인 반면 비흡연자의 경우 연간 10만 명당 17명으로 14.6(249/17)의 배율을 보였다[178]. 관상동맥심장질환(CHD)의 경우, 해당 비율은 연간 10만 명당 1001명과 619명으로, 비율비는 1.61(1001/619)이었다. 흡연이 사망에 미치는 영향은 CHD보다 폐암에서 훨씬 더 강하게 나타난다. 흡연의 절대적인 영향을 고려하면 그림이 달라진다. 발생률의 차이는 폐암의 경우 연간 10만 명당 232명(249-17명), 만성 폐쇄성 폐질환의 경우 382명(1001-619명)이었다. 따라서 흡연을 하는 의사들 사이에서 흡연은 폐암보다 만성 폐쇄성 폐질환으로 인한 사망을 유발할 가능성이 더 높았다. 노출을 제거함으로써 인구의 질병 부담을 얼마나 예방할 수 있을까? 한 연구에 따르면 전체 폐암의 91%, 만성 폐쇄성 폐질환의 40%, 2000년 남성 전체 사망의 33%가 흡연에 기인한 것으로 나타났다[179]. 인구 기여 비율은 일반적으로 특정 노출로 인한 환자의 비율로 정의되지만 여러 개념(통일된 용어 없음)이 존재하며 다른 요인을 보정하기 위한 잘못된 접근 방식이 때때로 사용된다[172,180]. 보고의 의미는 무엇인가? 상대적 측정치는 연관성의 강도를 강조하며 병인 연구에 가장 유용하다. 노출과의 인과관계가 제시되고 연관성이 효과로 해석되는 경우, 공중보건 정책의 가능한 영향을 측정하기 위해 상대위험도 추정치를 절대적 위험 측정치로 변환할 수 있다(항목 16c 참조)[181]. 그러나 저자는 이러한 맥락에서 만들어진 강력한 가정을 인식해야 한다[171]. 개별 상황에 적합한

개념과 방법을 결정할 때 주의가 필요하다.

17. 다른 분석: 다른 분석에 대해 보고한다 – 하부집단분석, 교호작용(interactions), 민감도분석

Example 1

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

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

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

Example 2

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

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

CI, confidence interval.

*Adjusted for age, sex, cardiovascular drug use, and unmeasured binary confounder.

Table adapted from Wei et al. [184].

설명

관찰 연구에서는 주 분석 외에도 다른 분석이 수행되는 경우가 많다. 이러한 분석은 특정 하위 그룹, 위험 요인 간의 잠재적 상호 작용, 기여 위험 계산을 다루거나 민감도 분석에서 연구 변수에 대한 다른 정의 사용 등이 있다.

하위 그룹 분석과 관련된 위험성과 분석 다양성에 대한 논쟁이 있다[4,104]. 우리의 의견으로는, 전체 결과에서 효과가 거의 또는 전혀 없을 때 하위 그룹별 연관성을 보는 근거를 찾는 경향이 너무 크다.

반면에 전체 연관성이 여러 하위 그룹, 특히 각 하위 그룹에서 충분한 자료를 확보할 수 있을 만큼 연구 규모가 큰 경우 미리 지정된 여러 하위 그룹에서 일관되게 나타나는지 살펴보는 것도 가치가 있다. 두 번째 쟁점 사항으로 자료 분석 중에 발생한 관심 하위 그룹에 관한 것이다. 이는 중요한 결과일 수도 있지만 우연히 발생할 수도 있다. 향후 다른 자료에 대한 분석을 통해 초기의 흥미로운 결과가 시간의 경과에 따라 어느 정도 지속될 수 있는지 알 수 없기 때문에 수행된 모든 하위 그룹 분석에 대해 알리는 것은 가능하지도 않고 필요하지도 않다고 주장하기도 한다[9]. 어떤 분석이 계획된 분석이고 어떤 분석이 계획되지 않은 분석인지 보고할 것이 필요하다(항목 4,12b 및 20 참조). 이를 통해 독자는 발견에서 검증 또는 반박에 이르는 연속체에서 연구의 위치를 고려하여 다중성의 의미를 판단할 수 있다.

세 번째 논쟁 영역은 위험 요인 간의 공동 효과와 교호성을 평가하는 방법, 즉 덧셈 모형 혹은 곱셈 모형으로 평가해야 하는지, 그것도 아니면 가장 적합한 통계 모델로 척도를 결정해야 하는지이다(항목 12b 및 박스 8 참조). 가능한 경우 위의 첫 번째 예[183] 또는 Martinelli 등의 연구에서와 같이 각 노출의 개별 효과와 공동 효과를 표로 작성하여 보고하는 것이 좋다[185]. 이러한 표는 독자에게 덧셈 상호

작용뿐만 아니라 곱셈 상호 작용을 평가할 수 있는 충분한 정보를 제공한다(이러한 계산 방법은 박스 8에 나와 있다). 개별 및 공동 효과에 대한 신뢰 구간은 독자가 자료의 강도를 판단하는 데 도움이 될 수 있다. 또한 상호 작용으로 인한 초과 상대위험도(RERI)와 같은 상호 작용 측정치에 대한 신뢰 구간은 상호 작용 테스트 또는 동질성 테스트와 관련이 있다. 한 가지 반복되는 문제는 저자가 하위 그룹 간의 P값 비교를 사용하여 효과 변경자(effect modifiers)에 대한 잘못된 주장을 하는 것이다. 예를 들어, 한 범주(예: 남성)에서는 통계적으로 유의미한 연관성이 있지만 다른 범주(예: 여성)에서는 그렇지 않다고 해서 그 자체로 효과 변경의 근거가 되는 것은 아니다. 마찬가지로, 각 점 추정치에 대한 신뢰 구간 간극이 겹친다는 것을 근거로 교호성이 없다고 추론하는 경우가 있는데 이는 부적절하다. 유효한 추론을 하려면 연관의 크기가 하위 그룹 간에 차이가 있는지 직접 평가해야 한다.

민감도 분석은 통계 분석에서 선택한 항목의 영향을 조사하거나 결측된 자료 또는 가능한 바이어스에 대한 결과의 강건성을 조사하는 데 유용하다(항목 12b 참조). 이러한 분석의 보고 수준과 관련하여 판단이 필요하다. 민감도 분석이 다수 수행된 경우, 모든 민감도 분석에 대해 상세한 결과를 제시하는 것은 비현실적일 수 있다. 민감도 분석이 수행되었으며 제시된 주요 결과와 일치한다고 보고하는 것으로 충분할 수 있다. 조사된 문제가 주요 관심사이거나 효과 추정치가 상당히 다른 경우 상세히 제시해야 한다[59,186].

Pocock 등은 관찰 연구를 보고한 73편의 논문 중 43편에 하위 그룹 분석이 포함되어 있으며, 대다수가 그룹 간 차이를 주장했지만 교호성에 대한 공식적인 평가를 보고한 논문은 8편에 불과했다고 하였다(항목12b 참조)[4].

고찰

고찰 파트에서는 연구의 타당성과 의미에 대해 다룬다[191]. 설문 조사에 따르면 고찰 파트는 연구 결과와 그 의미에 대한 불완전하거나 편향된 평가와 저자의 결과를 뒷받침하는 내용이 많이 나타나고 있다[192,193]. 고찰을 구조화하면 저자가 독자에게 결과 내용을 안내하면서 결과에 대한 부당한 추측과 과도한 해석을 피하는 데 도움이 될 수 있다[194,195]. 예를 들어, *Annals of Internal Medicine*[196]에서는 저자가 다음을 제시하여 고찰 파트를 구성할 것을 권장한다 : (1) 주요 결과에 대한 간략한 개요; (2) 가능한 메커니즘과 설명; (3) 다른 출판 연구의 관련 결과와의 비교; (4) 연구의 한계; (5) 진료과 연구에 미치는 영향에 대한 내용. 다른 사람들도 비슷한 제안을 했다[191,194]. 연구 권고 사항에 대한 부분과 연구의 한계에 대한 부분은 서로 밀접하게 연결되어야 한다. 연구자들은 '더 많은 연구가 필요하다'고 단호하게 말하기보다는 후속 연구가 자신의 연구를 개선할 수 있는 부분을 제시해야 한다[197,198]. 우리는 적절한 부제목을 사용하여 고찰 파트를 구성할 것을 권장한다.

18. 주요 결과: 연구 목적에 비추어 주요 결과를 요약한다.

예시

'우리는 소수 민족이 더 높은 수준의 심혈관 질환(CVD) 위험 요소와 관련이 있고, 그 연관성은 사회경제적 상태에 의해 실질적으로 설명될 것이라는 가설을 세웠다. 우리의 가설은 확인되지 않았다. 연령과 SES를

보정한 후에도 백인 여성과 흑인 및 멕시코계 미국인 여성 사이에는 체질량 지수, 혈압, 당뇨병, 신체 활동 부족에서 매우 유의미한 차이가 남아 있었다. 또한 우리는 SES에 의한 CVD 위험 요소에 큰 차이가 있음을 발견했는데, 이는 소수 민족 여성과 SES가 낮은 백인 여성 모두의 고위험 상태를 보여주는 결과이다[199].

설명

연구의 주요 결과에 대한 간략한 요약으로 고찰을 시작하는 것이 좋다. 짧은 요약은 독자에게 주요 결과를 상기시키고 저자가 제공한 해석과 의미가 결과에 의해 뒷받침되는지 여부를 평가하는 데 도움이 될 수 있다.

19. 제한점: 잠재적 바이어스나 비정밀을 고려하면서 연구의 제한점에 대해 고찰한다. 잠재적

바이어스의 방향성과 크기를 고찰한다

예시

"비만 수준이 증가함에 따라 상담 시행률도 증가하기 때문에 우리의 추정치는 실제 시행률을 과대평가할 수 있다. 전화 설문조사 역시 상담의 실제 시행률을 과대평가할 수 있다. 전화가 없는 사람은 전화가 있는 사람과 비슷한 수준의 과체중을 가지고 있지만, 전화가 없는 사람은 교육 수준이 낮은 경향이 있으며 이는 우리 연구에서 상담 수준이 낮은 것과 관련된 요소이다. 또한, 체중에 대한 질문에 응답을 거부한 사람들뿐만 아니라 참여를 거부한 사람들로 인해 발생할 수 있는 바이어스도 우려된다. 더욱이, 자료는 단면적으로 수집되었기 때문에 환자의 체중 감량 시도에 앞서 상담이 선행되었다고 추론할 수 없다." [200].

설명

연구의 한계를 확인하고 논의하는 것은 과학적 보고의 필수적인 부분이다. 결과에 영향을 미칠 수 있는 바이어스와 교란의 원인을 확인하는 것뿐만 아니라 잠재적인 바이어스의 방향과 크기, 상대적 중요성을 논의하는 것도 중요하다(항목 9 및 박스 3 참조).

저자는 결과의 비정밀성에 대해서도 논의해야 한다. 연구 규모(항목 10), 노출 측정, 교란요인 및 결과(항목 8)를 포함하여 연구의 여러 측면과 관련하여 비정밀성이 발생할 수 있다. 노출의 실제 값을 정확하게 측정할 수 없으면 단일성을 향하는 바이어스가 발생할 수 있다. 위험 요소를 덜 정확하게 측정할수록 바이어스가 커진다. 이 효과는 '감쇠'[201,202] 또는 최근에는 '회귀 희석 바이어스'[203]으로 설명되었다. 그러나 서로 연관되어 있는 위험 요인들이 서로 다른 정도로 비정밀하게 측정되면 이와 관련된 보정 상대위험도는 단일성(unity)으로 편향되거나 단일성에서 멀어질 수 있다[204-206].

한계를 논의할 때 저자는 제시된 연구를 타당성, 일반화 가능성 및 정밀도 측면에서 다른 연구와 비교할 수 있다. 이 접근 방식에서 각 연구는 독립되어 있지 않고, 서로 다른 연구에 영향을 미치는 것으로 볼 수 있다[207]. 놀랍게도 연구의 중요한 한계에 대한 논의가 출판된 보고서에서 생략되는 경우가 있다. The Lancet에 독창적인 연구 논문을 발표한 저자들을 대상으로 한 설문 조사에 따르면 연구의 중요한 약점이 조사 설문지에 보고되었지만 출판된 논문에는 보고되지 않은 것으로 나타났다[192].

20. 해석: 연구목적, 제한점, 유사연구 결과, 다양한 분석, 다른 관련 근거를 고려하면서 주의 깊게 결과에 대한 해석을 수행한다.

예시

"심근경색으로 인한 사망과 2세대 경구 피임약 사용 사이의 연관성에 대한 모든 설명은 추측적이어야 한다. 직접적인 생물학적 메커니즘을 제안하는 공개된 근거는 없으며 관련 결과가 있는 다른 역학 연구도 없다. (...) 절대 위험의 증가는 매우 작으며 아마도 주로 흡연자에게 적용될 것이다. 확증적인 증거가 부족하고 분석이 상대적으로 적은 숫자를 기반으로 하기 때문에 해당 주제에 대한 더 많은 증거가 필요하다. 우리는 이러한 결과를 바탕으로 처방을 변경하는 것을 권장하지 않다."[120].

설명

고찰 파트의 핵심은 연구 결과의 해석이다. 과잉 해석은 흔하고 인간적이다. 우리가 객관적인 평가를 하기 위해 열심히 노력할 때에도, 심사자들은 우리가 어떤 면에서 너무 멀리 나아갔다고 올바르게 지적하는 경우가 많다. 결과를 해석할 때 저자는 발견과 검증의 스펙트럼에서 어디에 속하고, 탈락과 비참여같은 바이어스의 잠재적 원인을 고려해야 한다(항목 9, 12 및 19 참조). 교란(항목 16a), 관련 민감도 분석 결과, 다중분석 및 하위 그룹 분석 문제(항목 17)를 적절히 고려해야 한다. 저자는 또한 측정되지 않은 변수나 정밀하지 않은 교란변수 측정으로 인한 잔여 교란도 고려해야 한다. 예를 들어, 사회경제적 상태는 많은 건강 결과와 연관되어 있으며 종종 비교 대상 그룹마다 다를 수 있다. SES를 측정하는 데 사용되는 변수는 정의되지 않고 측정되지 않은 기타 노출에 대한 대용이며, 실제 교란요인은 정의에 따라 오류로 측정될 수 있다[208]. 저자는 신뢰 구간에 반영된 통계적 불확실성보다 더 큰 추정치의 실제 불확실성 범위를 다루어야 한다. 통계적 불확실성은 연구의 설계, 구현 및 측정 방법에서 발생하는 다른 불확실성을 고려하지 않는다[209].

인과관계에 대한 사고와 결론을 안내하기 위해 일부에서는 1965년 Bradford Hill이 제안한 기준이 도움이 될 수 있다[164]. 노출과의 연관성은 얼마나 강한가? 질병이 발병하기 전에 발생했나? 다양한 연구와 환경에서 연관성이 일관되게 관찰되나? 실험실 및 동물 연구를 포함한 실험 연구에서 뒷받침하는 증거가 있나? 노출의 추정 효과는 얼마나 구체적이며 용량-반응 관계가 있나? 연관성이 생물학적으로 타당한가? 그러나 이러한 기준을 기계적으로 적용해서는 안된다. 예를 들어 일부에서는 2 또는 3 미만의 상대 위험도를 무시해야 한다고 주장했다[210,211]. 이는 Cornfield et al.의 주장을 뒤집은 것이다. 큰 상대적 위험의 강도에 대해(항목 12b 참조) [127] 인과적 영향은 상대 위험도가 9일 때 더 가능성이 높지만, 3 미만의 위험도가 반드시 허위라는 뜻은 아니다. 예를 들어, 자궁 내 방사선 조사 후 소아 백혈병 위험이 약간 증가하는 것은 다른 설명이 분명하지 않은 의료 절차의 부작용과 관련되기 때문에 신뢰할 수 있다[212]. 게다가 방사선의 발암성 영향도 잘 알려져 있다. 일주일에 2~4개의 계란을 섭취하는 것과 관련된 난소암 위험이 두 배로 증가한다는 것은 즉시 신뢰할 수 없다. 왜냐하면 식습관은 SES뿐만 아니라 많은 생활 방식 요인과 연관되어 있기 때문이다[213]. 대조적으로, 다양한 유형의 경구 피임약 사이의 혈전증 위험 차이에 대해 많이 논의된 역학적 지식의 신뢰성은 무작위 교차 시험에서 발견된 응고의 차이로 인해 크게 향상되었다[214]. 다양한 유형의 연구에서 나온 기존 외부 증거에 대한 논의가 항상 포함되어야 하지만, 위험의 작은 증가를 보고하는 연구에서는 특히 중요할 수 있다. 또한 저자는 자신의 결과를 유사한 연구와 연관시키고 이상적으로는 체계적인 문헌고찰을 참조하여 새로운 연구가 기존 증거에 어떻게 영향을 미치는지 설명해야 한다.

21. 일반화 가능성: 연구 결과의 일반화 가능성(외적타당도)에 대해 고찰한다.

예시

"우리의 추정치가 다른 HIV-1 감염 환자에게 얼마나 적용할 수 있을까? 예후 모델을 개발하는 데 사용된 자료밖의 다른 자료에 적용하면 예측 모델의 정확도가 낮아지는 경향이 있기 때문에, 이는 중요한 질문이다. 우리는 모델 복잡성에 페널티를 적용하고 추정 절차에서 생략된 집단에 가장 잘 일반화된 모델을 선택하여 이 문제를 해결했다. 우리 데이터베이스에는 다양한 환경에서 치료를 받은 유럽과 북미의 여러 국가의 환자가 포함되어 있다. 환자의 범위는 남성과 여성, 청소년부터 노인까지 광범위했으며, 주요 노출 범주가 잘 표현되어 있었다. 기준치에서 면역결핍의 심각도는 측정할 수 없는 것부터 매우 심각한 것까지, 바이러스 수치는 감지할 수 없는 것부터 극도로 높은 것까지 다양했다." [215]

설명

외부 타당성 또는 적용 가능성이라고도 불리는 일반화 가능성은 연구 결과를 다른 상황에 적용할 수 있는 정도를 말한다[216]. 외적타당도는 명확하게 명시된 조건과 관련해서만 의미가 있다[217]. 연령, 성별, 민족, 질병의 중증도 및 동반 질환과 관련하여 연구에 등록한 사람들과 다른 개인, 그룹 또는 인구 집단에 결과를 적용할 수 있을까? 노출의 성격과 수준이 비교 가능하고 결과의 정의가 다른 환경이나 모집단과 관련이 있을까? 수년 전에 추적 연구를 통해 수집된 자료가 오늘날에도 여전히 관련이 있을까? 한 국가의 의료 서비스 연구 결과가 다른 국가의 의료 시스템에 적용 가능할까? 연구 결과가 외적 타당도를 가지고 있는지 여부에 대한 질문은 종종 연구 환경, 참가자의 특성, 조사된 노출 및 평가된 결과에 따라 판단의 문제이다. 따라서 저자가 독자에게 환경과 위치, 자격 기준, 노출 및 측정 방법, 결과 정의, 모집 및 추적

기간에 대한 적절한 정보를 제공하는 것이 중요하다. 비참여 정도와 결과가 나타나는 노출되지 않은 참가자의 비율도 관련이 있다. 모집단에 따라 종종 달라지는 절대 위험과 노출 확산에 대한 지식은 결과를 다른 설정 및 모집단에 적용할 때 도움이 된다(박스 7 참조).

기타 정보

22. 자금원: 있다면, 현재 논문의 기초가 된 연구의 자금원과 자금지원자의 역할을 제시한다.

설명

일부 저널에서는 저자에게 재정적 및 기타 이해 상충의 유무를 공개하도록 요구한다[100,218]. 여러 조사에서는 자금 출처와 연구 논문의 결론 사이에 강한 연관성이 있음을 보여준다[219-222]. 무작위 임상시험의 결론은 효과 크기를 보정한 후에도 영리 단체의 자금 지원을 받은 연구의 경우 실험 약물을 더 많이 권고하였(OR 5.3). 다른 연구에서는 자금을 지원한 연구에 대한 담배 및 통신 산업의 영향이 있었다[224-227]. 스폰서가 정부 또는 비영리 조직인 경우에도 과도한 영향력을 행사하는 사례가 있다. 저자 또는 자금 제공자는 다음 사항에 영향을 미치는 이해 상충을 가질 수 있다: 연구 설계 [228]; 노출 선택 [228,229], 결과 [230], 통계 방법 [231], 건강결과 [230] 및 연구 [232]의 선택적 보고. 결과적으로 자금 제공자의 역할을 자세히 설명해야 한다 : 연구의 어떤 부분에서 직접 책임을 졌는지(예: 설계, 자료 수집, 분석, 원고 초안 작성, 출판 결정)[100]. 부당한 영향을 미치는 다른 출처로는 고용주(예: 학계 연구원의 경우 대학 관리자, 정부 감독자, 특히 정부 연구원의 경우 정치적 지명자), 자문 위원회, 소송

당사자 및 특수 이익 단체가 있다.

결론

STROBE Statement는 역학 관찰 연구 보고에 대한 권고안을 제공하는 것이 목표이다. 좋은 보고는 연구의 강점과 약점을 드러내고 연구 결과의 건전한 해석과 적용을 촉진한다. STROBE Statement는 또한 관찰 연구를 계획하는 데 도움이 될 수 있으며 동료 심사자와 편집자가 원고를 평가하는 데 지침이 될 수 있다.

우리는 관찰 연구에서 투명하고 완전하게 하는 보고의 중요성을 논의하고, 체크리스트에 포함된 다양한 항목의 근거를 설명하고, 좋은 보고라고 생각하는 출판된 기사의 예를 제공하기 위해 이 설명 기사를 썼다. 여기에 제시된 자료가 작성자와 편집자가 STROBE를 사용하는 데 도움이 되기를 바란다.

우리는 연구 보고에 대한 STROBE 및 기타 권장 사항[13,233,234]이 지속적인 평가, 개선 및 필요한 경우 변경이 필요한 문서로 보아야 함을 강조한다[235,236]. 예를 들어, 병렬 그룹 무작위 시험 보고를 위한 CONSORT statement은 1990년대 중반에 처음 개발되었다[237]. 그 이후로 그룹 구성원들은 정기적으로 만나 권장 사항 수정의 필요성을 검토했다. 개정판은 2001년에 등장했으며[233] 추가 버전이 개발 중이다. 마찬가지로, 이 기사와 STROBE 체크리스트에 제시된 원칙은 새로운 증거와 비판적 의견이 축적됨에 따라 변경될 수 있다. STROBE 웹 사이트(<http://www.strobe-statement.org/>)는 체크리스트, 이 설명 문서 및 역학 연구의 올바른 보고에 대한 정보의 개선을 위한 고찰과 제안을 위한 포럼을 제공한다. 몇몇 저널에서는 저자들에게 STROBE Statement를 따르도록 요청한다 (현재 목록은 <http://www.strobe-statement.org/> 참조). 우리는 다른 저널에도 STROBE 선언문을 채택하도록 요청하고 당사 웹사이트를

통해 연락하여 이를 알려드린다. STROBE 권장 사항을 출판하는 저널은 오픈 액세스로 제공한다. 따라서 STROBE Statement는 생물의학 커뮤니티에서 널리 접근할 수 있다.

박스 8. 교호성(효과 변경): 관절효과 분석

노출과 질병 위험의 연관성이 다른 노출에 따라 달라질 때 교호성(interaction)이 존재한다. 교호성을 평가하고 보고할 때의 문제 중 하나는 노출 효과가 상대 위험도(또는 비율) 또는 위험도 차이(또는 비율 차이)의 두 가지 방법으로 측정될 수 있다는 것이다. 상대 위험도를 사용하면 곱셈(multiplicative) 모델이 되고, 위험도 차이를 사용하면 덧셈(additive) 모델이 된다[187,188]. 곱셈 또는 덧셈 모델 모두 아닌 '통계적 상호 작용'과 덧셈 모델에서 벗어나 측정되는 '생물학적 상호 작용'이 때때로 구별된다[189]. 그러나 덧셈 모델이나 곱셈 모델 모두 특정 생물학적 메커니즘을 나타내지 않는다.

모델 선택에 관계없이 주요 목표는 두 노출의 결합 효과가 (다른 노출이 없는 경우) 개별 효과와 어떻게 다른지 이해하는 것이다. Human Genomic Epidemiology Network(HuGENet)는 다양한 유형의 교호성을 평가할 수 있는 개별 및 공동 효과를 투명하게 제시하기 위한 레이아웃을 제안했다[183].

제안을 설명하기 위해 경구 피임약 및 인자 V Leiden 돌연변이에 관한 연구 자료가 사용되었으며 이 예는 항목 17에도 사용되었다. 경구 피임약과 인자 V Leiden 돌연변이 모두 정맥 혈전증의 위험을

증가시킨다. 이들의 개별 효과와 공동 효과는 2x4 표(항목 17의 예 1 참조)에서 계산할 수 있다.

여기서 OR 1은 경구 피임약을 사용하지 않는 인자 V Leiden이 없는 여성의 기준선을 나타낸다.

어려운 점은 사례 관리 연구와 같은 일부 연구 설계와 로지스틱 또는 Cox 회귀 모델과 같은 여러 통계

모델이 상대 위험도를 추정하고 본질적으로 곱셈 모델로 이어진다는 것이다. 이 경우 상대 위험도는

추가 규모로 변환될 수 있다. 항목 17의 예 1에서 개별 OR은 3.7과 6.9이다. 결합 승산비는 34.7이다.

이러한 자료를 승법 모델로 분석하면 결합 승산비는 25.7(3.73 - 6.9)이 될 것으로 예상된다. 관찰된

결합 효과 34.7은 곱셈 척도(34.7/25.7)에서 예상되는 것보다 1.4배 더 크다. 이는 곱셈 상호 작용의

승산비이다. 이는 로지스틱 회귀 모델에서 추정된 상호 작용 계수의 역로그와 같다. 가산 모델에서

결합 승산비는 9.6(3.7 + 6.9 - 1)이 될 것으로 예상된다. 관찰된 결합 효과는 가산성과 크게 다르다.

차이는 25.1(34.7 - 9.6)이다. 승산비가 상대적 위험(또는 비율)으로 해석될 때 후자의 양(25.1)은

RERI(교호성으로 인한 상대적 초과 위험)이다[190]. 이는 기준 값(OR=1과 동일)이 정맥 혈전증의 기준

발생률, 예를 들어 1/10,000 여성년을 나타낸다고 상상하면 더 쉽게 이해될 수 있다. 이는 개별노출과

공동노출이 있을 때 증가한다.

ORCID

Soo-Young Kim: <https://orcid.org/0000-0002-3205-9408>

Hyun-Chang Kim: <https://orcid.org/0000-0001-7867-1240>

Yoonjoo Suh: <https://orcid.org/0000-0002-0202-8352>

Authors' contributions

Translation: Kim SY

Scientific review: Kim HC

Review for Korean: Suh Y

Conflict of interest

The authors have declared that no competing interests exist. The translators and reviewers have declared that no competing interests exist.

Funding

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

Data availability

Not applicable.

Acknowledgments

Not applicable.

Supplementary materials

Not applicable.

References

1. Glasziou P, Vandenbroucke JP, Chalmers I (2004) Assessing the quality of research. *BMJ* 328: 39–41.
2. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G (2001) Distribution of study designs in four major

US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 51: 8–11.

3. Scales CD Jr., Norris RD, Peterson BL, Preminger GM, Dahm P (2005) Clinical research and statistical methods in the urology literature. *J Urol* 174: 1374–1379.
4. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, et al. (2004) Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 329: 883.
5. Tooth L, Ware R, Bain C, Purdie DM, Dobson A (2005) Quality of reporting of observational longitudinal research. *Am J Epidemiol* 161: 280–288.
6. von Elm E, Altman DG, Egger M, Gøtzsche PC, Pocock SJ, et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med* e296. doi:10.1371/journal.pmed.0040296.
7. Mihailovic A, Bell CM, Urbach DR (2005) Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg* 48: 148–151.
8. Rushton L (2000) Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup Environ Med* 57: 1–9.
9. Rothman KJ (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1: 43–46.
10. Moonesinghe R, Khoury MJ, Janssens AC (2007) Most published research findings are false—but a little replication goes a long way. *PLoS Med* 4: e28. doi:10.1371/journal.pmed.0040028
11. Jenicek M (1999) *Clinical Case Reporting. Evidence-Based Medicine*. Oxford: Butterworth-Heinemann. 117 pp.
12. Vandenbroucke JP (2001) In defense of case reports and case series. *Ann Intern Med* 134: 330–334.
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 138: 40–44.
14. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, et al. (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 93: 387–391.
15. Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, et al. (2006) A road map for efficient and

reliable human genome epidemiology. *Nat Genet* 38: 3–5.

16. Rodrigues L, Kirkwood BR (1990) Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol* 19: 205–213.
17. Rothman KJ, Greenland S (1998) Case-Control Studies. In: Rothman KJ, Greenland S editors. *Modern epidemiology*. 2nd ed: Lippincott Raven. pp. 93–114.
18. Forand SP (2004) Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. *Environ Health* 3: 7.
19. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 342: 1878–1886.
20. Gøtzsche PC, Harden A Searching for non-randomised studies. Draft chapter 3. Cochrane Non-Randomised Studies Methods Group, 26 July 2002. Available: <http://www.cochrane.dk/nrsmg>. Accessed 10 September 2007.
21. Lohse N, Hansen AB, Pedersen G, Kronborg G, Gerstoft J, et al. (2007) Survival of persons with and without HIV infection in Denmark, 1995– 2005. *Ann Intern Med* 146: 87–95.
22. American Journal of Epidemiology (2007) Information for authors. Available: http://www.oxfordjournals.org/aje/for_authors/index.html. Accessed 10 September 2007.
23. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ (1990) More informative abstracts revisited. *Ann Intern Med* 113: 69–76.
24. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, et al. (1994) Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ* 150: 1611–1615.
25. Hartley J, Sydes M (1996) Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *J Inform Sci* 22: 27–37.
26. Viner RM, Cole TJ (2005) Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *BMJ* 330: 1354.

27. McCauley J, Kern DE, Kolodner K, Dill L, Schroeder AF, et al. (1995) The “battering syndrome”: prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med* 123: 737–746.
28. McEvoy SP, Stevenson MR, McCartt AT, Woodward M, Haworth C, et al. (2005) Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 331: 428.
29. Vandenbroucke JP (1991) Prospective or retrospective: what’s in a name? *BMJ* 302: 249–250.
30. Last JM (2000) *A Dictionary of Epidemiology*. New York: Oxford University Press.
31. Miettinen OS (1985) *Theoretical Epidemiology: principles of occurrence research in medicine*. New York: Wiley. pp. 64–66.
32. Rothman KJ, Greenland S (1998) Types of Epidemiologic Studies. In: Rothman KJ, Greenland S editors. *Modern epidemiology*. 2nd ed: Lippincott Raven. pp. 74–75.
33. MacMahon B, Trichopoulos D (1996) *Epidemiology, principles and methods*. 2nd ed. Boston: Little, Brown. 81 p.
34. Lilienfeld AM (1976) *Foundations of Epidemiology*. New York: Oxford University Press.
35. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, et al. (1995) Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med* 332: 912–917.
36. Goodman KJ, O’Rourke K, Day RS, Wang C, Nurgalieva Z, et al. (2005) Dynamics of *Helicobacter pylori* infection in a US-Mexico cohort during the first two years of life. *Int J Epidemiol* 34: 1348–1355.
37. Altman DG, De Stavola BL, Love SB, Stepniowska KA (1995) Review of survival analyses published in cancer journals. *Br J Cancer* 72: 511–518.
38. Cerhan JR, Wallace RB, Folsom AR, Potter JD, Munger RG, et al. (1993) Transfusion history and cancer risk in older women. *Ann Intern Med* 119: 8–15.
39. Beane Freeman LE, Dennis LK, Lynch CF, Thorne PS, Just CL (2004) Toenail arsenic content and cutaneous melanoma in Iowa. *Am J Epidemiol* 160: 679–687.
40. Canto JG, Allison JJ, Kiefe CI, Fincher C, Farmer R, et al. (2000) Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *N Engl J Med* 342:

1094– 1100.

41. Metzkor-Cotter E, Kletter Y, Avidor B, Varon M, Golan Y, et al. (2003) Long-term serological analysis and clinical follow-up of patients with cat scratch disease. *Clin Infect Dis* 37: 1149–1154.
42. Johnson ES (1990) Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. *J Occup Med* 32: 250–254.
43. Berkson J (1946) Limitations of the application of fourfold table analysis to hospital data. *Biom Bull* 2: 53.
44. Feinstein AR, Walter SD, Horwitz RI (1986) An analysis of Berkson's bias in case-control studies. *J Chronic Dis* 39: 495–504.
45. Jick H, Vessey MP (1978) Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 107: 1–7.
46. Hackam DG, Mamdani M, Li P, Redelmeier DA (2006) Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet* 367: 413–418.
47. Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, et al. (2004) MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet* 364: 963–969.
48. Costanza MC (1995) Matching. *Prev Med* 24: 425–433.
49. Sturmer T, Brenner H (2002) Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *Am J Epidemiol* 155: 593–602.
50. Rothman KJ, Greenland S (1998) Matching. In: Rothman KJ, Greenland S, editors. 2nd ed. *Modern epidemiology*. Lippincott Raven. pp. 147–161.
51. Szklo MF, Nieto J (2000). *Epidemiology, Beyond the Basics*. Sudbury (MA): Jones and Bartlett. pp. 40–51.
52. Cole P, MacMahon B (1971) Attributable risk percent in case-control studies. *Br J Prev Soc Med* 25: 242–244.
53. Gissler M, Hemminki E (1996) The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. *Eur J Obstet Gynecol Reprod*

Biol 69: 73–75.

54. Gefeller O, Pfahlerberg A, Brenner H, Windeler J (1998) An empirical investigation on matching in published case-control studies. *Eur J Epidemiol* 14: 321–325.
55. Artama M, Ritvanen A, Gissler M, Isojarvi J, Auvinen A (2006) Congenital structural anomalies in offspring of women with epilepsy—a population- based cohort study in Finland. *Int J Epidemiol* 35: 280–287.
56. Ebrahim S (2004) Cohorts, infants and children. *Int J Epidemiol* 33: 1165– 1166.
57. Walker M, Whincup PH, Shaper AG (2004) The British Regional Heart Study 1975–2004. *Int J Epidemiol* 33: 1185–1192.
58. Wieland S, Dickersin K (2005) Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol* 58: 560–567.
59. Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L (2005) Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology* 16: 155–163.
60. Winkelmayr WC, Stampfer MJ, Willett WC, Curhan GC (2005) Habitual caffeine intake and the risk of hypertension in women. *JAMA* 294: 2330– 2335.
61. Lukanova A, Soderberg S, Kaaks R, Jellum E, Stattin P (2006) Serum adiponectin is not associated with risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 15: 401–402.
62. Becher H (1992) The concept of residual confounding in regression models and some applications. *Stat Med* 11: 1747–1758.
63. Brenner H, Blettner M (1997) Controlling for continuous confounders in epidemiologic research. *Epidemiology* 8: 429–434.
64. Phillips MR, Yang G, Zhang Y, Wang L, Ji H, et al. (2002) Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet* 360: 1728–1736.
65. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, et al. (2006) Prospective study of type 2 diabetes mellitus and risk of primary open- angle glaucoma in women. *Ophthalmology* 113:

1081–1086.

66. Craig SL, Feinstein AR (1999) Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol* 22: 51–56.
67. Rogler LH, Mroczek DK, Fellows M, Loftus ST (2001) The neglect of response bias in mental health research. *J Nerv Ment Dis* 189: 182–187.
68. Murphy EA (1976) *The logic of medicine*. Baltimore: Johns Hopkins University Press.
69. Sackett DL (1979) Bias in analytic research. *J Chronic Dis* 32: 51–63.
70. Johannes CB, Crawford SL, McKinlay JB (1997) Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. *Am J Epidemiol* 146: 429–438.
71. Bloemenkamp KW, Rosendaal FR, Buller HR, Helmerhorst FM, Colly LP, et al. (1999) Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med* 159: 65–70.
72. Feinstein AR (1985) *Clinical epidemiology: the architecture of clinical research*. Philadelphia: W.B. Saunders.
73. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA (2003) Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *Am J Trop Med Hyg* 68: 519–526.
74. Anoop S, Saravanan B, Joseph A, Cherian A, Jacob KS (2004) Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. *Arch Dis Child* 89: 325–329.
75. Carlin JB, Doyle LW (2002) Sample size. *J Paediatr Child Health* 38: 300–304.
76. Rigby AS, Vail A (1998) Statistical methods in epidemiology. II: A commonsense approach to sample size estimation. *Disabil Rehabil* 20: 405–410.
77. Schulz KF, Grimes DA (2005) Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 365: 1348–1353.
78. Drescher K, Timm J, Jockel KH (1990) The design of case-control studies: the effect of confounding on sample size requirements. *Stat Med* 9: 765–776.

79. Devine OJ, Smith JM (1998) Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med* 17: 1375–1389.
80. Linn S, Levi L, Grunau PD, Zaidise I, Zarka S (2007) Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. *Ann Epidemiol* 17: 142–147.
81. Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994) Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 86: 829–835.
82. Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 25: 127–141.
83. Greenland S (1995) Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology* 6: 450–454.
84. Royston P, Ambler G, Sauerbrei W (1999) The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 28: 964–974.
85. MacCallum RC, Zhang S, Preacher KJ, Rucker DD (2002) On the practice of dichotomization of quantitative variables. *Psychol Methods* 7: 19–40.
86. Altman DG (2005) Categorizing continuous variables. In: Armitage P, Colton T editors. *Encyclopedia of biostatistics*. 2nd ed. Chichester: John Wiley. pp. 708–711.
87. Cohen J (1983) The cost of dichotomization. *Applied Psychological Measurement* 7: 249–253.
88. Zhao LP, Kolonel LN (1992) Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *Am J Epidemiol* 136: 464–474.
89. Cochran WG (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 24: 295–313.
90. Clayton D, Hills M (1993) Models for dose-response (Chapter 25). *Statistical Models in Epidemiology*. Oxford: Oxford University Press. pp. 249–260.
91. Cox DR (1957) Note on grouping. *J Am Stat Assoc* 52: 543–547.
92. Il'yasova D, Hertz-Picciotto I, Peters U, Berlin JA, Poole C (2005) Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes Control*

16: 383–388.

93. Berglund A, Alfredsson L, Cassidy JD, Jensen I, Nygren A (2000) The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. *J Clin Epidemiol* 53: 1089–1094.
94. Slama R, Werwatz A (2005) Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique* 53 Spec No 2: 2S65–80.
95. Greenland S (1998) Introduction to regression modelling (Chapter 21). In: Rothman KJ, Greenland S editors. *Modern epidemiology*. 2nd ed: Lippincott Raven. pp. 401–432.
96. Thompson WD (1994) Statistical analysis of case-control studies. *Epidemiol Rev* 16: 33–50.
97. Schlesselman JJ (1982) Logistic regression for case-control studies (Chapter 8.2). *Case-control studies Design, conduct, analysis*. New York, Oxford: Oxford University Press. pp. 235–241.
98. Clayton D, Hills M (1993) Choice and interpretation of models (Chapter 27). *Statistical Models in Epidemiology*. Oxford: Oxford University Press. pp. 271–281.
99. Altman DG, Gore SM, Gardner MJ, Pocock SJ (1983) Statistical guidelines for contributors to medical journals. *Br Med J* 286: 1489–1493.
100. International Committee of Medical Journal Editors (1997) Uniform requirements for manuscripts submitted to biomedical journals [Electronic version updated February 2006, available at <http://www.icmje.org/>] *N Engl J Med* 336: 309–315.
101. Mullner M, Matthews H, Altman DG (2002) Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 136: 122–126.
102. Olsen J, Basso O (1999) Re: Residual confounding. *Am J Epidemiol* 149: 290.
103. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarod K, et al. (2006) Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? *Am J Kidney Dis* 47: 396–405.
104. Gotzsche PC (2006) Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ* 333: 231–234.
105. Szklo MF, Nieto J (2000) Communicating Results of Epidemiologic Studies (Chapter 9). *Epidemiology, Beyond the Basics*. Sudbury (MA): Jones and Bartlett. pp. 408–430.
106. Chandola T, Brunner E, Marmot M (2006) Chronic stress at work and the metabolic syndrome:

prospective study. *BMJ* 332: 521–525.

107. Vach W, Blettner M (1991) Biased estimation of the odds ratio in case- control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol* 134: 895–907.

108. Little RJ, Rubin DB (2002) A taxonomy of missing-data methods (Chapter 1.4.) In: *Statistical Analysis with Missing Data*. New York: Wiley. pp. 19– 23.

109. Ware JH (2003) Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 348: 2136–2137.

110. Rubin DB (1976) Inference and missing data. *Biometrika* 63: 581–592.

111. Schafer JL (1997) *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

112. Lipsitz SR, Ibrahim JG, Chen MH, Peterson H (1999) Non-ignorable missing covariates in generalized linear models. *Stat Med* 18: 2435–2448.

113. Rotnitzky A, Robins J (1997) Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med* 16: 81–102.

114. Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

115. Barnard J, Meng XL (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 8: 17–36.

116. Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boulle A, et al. (2006) Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 367: 817–824.

117. Purandare N, Burns A, Daly KJ, Hardicre J, Morris J, et al. (2006) Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. *BMJ* 332: 1119–1124.

118. Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J (2001) Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens* 19: 1717–1725.

119. Lohr SL (1999) *Design Effects (Chapter 7.5). Sampling: Design and Analysis*. Pacific Grove (CA): Duxbury Press.

120. Dunn NR, Arscott A, Thorogood M (2001) The relationship between use of oral contraceptives

and myocardial infarction in young women with fatal outcome, compared to those who survive: results from the MICA case-control study. *Contraception* 63: 65–69.

121. Rothman KJ, Greenland S (1998) Basic Methods for Sensitivity Analysis and External Adjustment. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed: Lippincott Raven. pp. 343–357.

122. Custer B, Longstreth WT Jr., Phillips LE, Koepsell TD, Van Belle G (2006) Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. *BMC Cancer* 6: 152.

123. Wakefield MA, Chaloupka FJ, Kaufman NJ, Orleans CT, Barker DC, et al. (2000) Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ* 321: 333–337.

124. Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 98: 47–54.

125. Lash TL, Fink AK (2003) Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 14: 451–458.

126. Phillips CV (2003) Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 14: 459–466.

127. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, et al. (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 22: 173–203.

128. Langholz B (2001) Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? *Bioelectromagnetics* 5: S19–31.

129. Eisner MD, Smith AK, Blanc PD (1998) Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA* 280: 1909–1914.

130. Dunne MP, Martin NG, Bailey JM, Heath AC, Bucholz KK, et al. (1997) Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *Int J Epidemiol* 26: 844–854.

131. Schuz J, Kaatsch P, Kaletsch U, Meinert R, Michaelis J (1999) Association of childhood cancer with

factors related to pregnancy and birth. *Int J Epidemiol* 28: 631–639.

132. Cnattingius S, Zack M, Ekblom A, Gunnarskog J, Linet M, et al. (1995) Prenatal and neonatal risk factors for childhood myeloid leukemia. *Cancer Epidemiol Biomarkers Prev* 4: 441–445.

133. Schuz J (2003) Non-response bias as a likely cause of the association between young maternal age at the time of delivery and the risk of cancer in the offspring. *Paediatr Perinat Epidemiol* 17: 106–112.

134. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD (1995) Response rates among control subjects in case-control studies. *Ann Epidemiol* 5: 245–249.

135. Schulz KF, Grimes DA (2002) Case-control studies: research in reverse. *Lancet* 359: 431–434.

136. Olson SH, Voigt LF, Begg CB, Weiss NS (2002) Reporting participation in case-control studies. *Epidemiology* 13: 123–126.

137. Morton LM, Cahill J, Hartge P (2006) Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 163: 197–203.

138. Olson SH (2001) Reported participation in case-control studies: changes over time. *Am J Epidemiol* 154: 574–581.

139. Sandler DP (2002) On revealing what we'd rather hide: the problem of describing study participation. *Epidemiology* 13: 117.

140. Hepworth SJ, Schoemaker MJ, Muir KR, Swerdlow AJ, van Tongeren MJ, et al. (2006) Mobile phone use and risk of glioma in adults: case-control study. *BMJ* 332: 883–887.

141. Hay AD, Wilson A, Fahey T, Peters TJ (2003) The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. *Fam Pract* 20: 696–705.

142. Egger M, Juni P, Bartlett C (2001) Value of flow diagrams in reports of randomized controlled trials. *JAMA* 285: 1996–1999.

143. Osella AR, Misciagna G, Guerra VM, Chiloiro M, Cuppone R, et al. (2000) Hepatitis C virus (HCV) infection and liver-related mortality: a population-based cohort study in southern Italy. The Association for the Study of Liver Disease in Puglia. *Int J Epidemiol* 29: 922–927.

144. Dales LG, Ury HK (1978) An improper use of statistical significance testing in studying

covariables. *Int J Epidemiol* 7: 373–375.

145. Maldonado G, Greenland S (1993) Simulation study of confounder- selection strategies. *Am J Epidemiol* 138: 923–936.

146. Tanis BC, van den Bosch MA, Kemmeren JM, Cats VM, Helmerhorst FM, et al. (2001) Oral contraceptives and the risk of myocardial infarction. *N Engl J Med* 345: 1787–1793.

147. Rothman KJ, Greenland S (1998) Precision and Validity in Epidemiologic Studies. In: Rothman KJ, Greenland S editors. *Modern epidemiology*. 2nd ed: Lippincott Raven. pp. 120–125.

148. Clark TG, Altman DG, De Stavola BL (2002) Quantification of the completeness of follow-up. *Lancet* 359: 1309–1310.

149. Qiu C, Fratiglioni L, Karp A, Winblad B, Bellander T (2004) Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. *Epidemiology* 15: 687–694.

150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, Ruberantwari A, Wagner HU, et al. (1996) Incidence of HIV-1 infection in adults and socio- demographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. *Int J Epidemiol* 25: 1077–1082.

151. Mastrangelo G, Fedeli U, Fadda E, Valentini F, Agnesi R, et al. (2004) Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect* 112: 1188–1192.

152. Salo PM, Arbes SJ Jr., Sever M, Jaramillo R, Cohn RD, et al. (2006) Exposure to *Alternaria alternata* in US homes is associated with asthma symptoms. *J Allergy Clin Immunol* 118: 892–898.

153. Pocock SJ, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 359: 1686– 1689.

154. Sasieni P (1992) A note on the presentation of matched case-control data. *Stat Med* 11: 617–620.

155. Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA (2002) A nested case- control study of residential and personal magnetic field measures and miscarriages. *Epidemiology* 13: 21–31.

156. Tiihonen J, Wahlbeck K, Lonnqvist J, Klaukka T, Ioannidis JP, et al. (2006) Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first

hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. *BMJ* 333: 224.

157. Christenfeld NJ, Sloan RP, Carroll D, Greenland S (2004) Risk factors, confounding, and the illusion of statistical control. *Psychosom Med* 66: 868–875.

158. Smith GD, Phillips A (1990) Declaring independence: why we should be cautious [erratum: *J Epidemiol Community Health* 1991 Mar;45(1);88]. *J Epidemiol Community Health* 44: 257–258.

159. Greenland S, Neutra R (1980) Control of confounding in the assessment of medical technology. *Int J Epidemiol* 9: 361–367.

160. Robins JM (2001) Data, design, and background knowledge in etiologic inference. *Epidemiology* 12: 313–320.

161. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA (2007) Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology* 18: 120–129.

162. World Health Organization (2007) Body Mass Index (BMI). Available: http://www.euro.who.int/nutrition/20030507_1. Accessed 10 September 2007.

163. Beral V (2003) Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 362: 419–427.

164. Hill AB (1965) The environment and disease: Association or causation? *Proc R Soc Med* 58: 295–300.

165. Vineis P (2003) Causality in epidemiology. *Soz Präventivmed* 48: 80–87.

166. Empana JP, Ducimetiere P, Arveiler D, Ferrieres J, Evans A, et al. (2003) Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur Heart J* 24: 1903–1911.

167. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, et al. (1999) Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. *Lancet* 353: 1547–1557.

168. Cambien F, Chretien JM, Ducimetiere P, Guize L, Richard JL (1985) Is the relationship between

blood pressure and cardiovascular risk dependent on body mass index? *Am J Epidemiol* 122: 434–442.

169. Hosmer DW, Taber S, Lemeshow S (1991) The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 81: 1630–1635.

170. Tibshirani R (1982) A plain man's guide to the proportional hazards model. *Clin Invest Med* 5: 63–68.

171. Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. *Am J Public Health* 88: 15–19.

172. Uter W, Pfahlerberg A (2001) The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res* 10: 231–237.

173. Schwartz LM, Woloshin S, Dvorin EL, Welch HG (2006) Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ* 333: 1248.

174. Nakayama T, Zaman MM, Tanaka H (1998) Reporting of attributable and relative risks, 1966–97. *Lancet* 351: 1179.

175. Cornfield J (1951) A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst* 11: 1269–1275.

176. Pearce N (1993) What does the odds ratio estimate in a case-control study? *Int J Epidemiol* 22: 1189–1192.

177. Rothman KJ, Greenland S (1998) Measures of Disease Frequency. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed. Lippincott Raven. pp. 44–45.

178. Doll R, Hill AB (2004) The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. *BMJ* 328: 1529–1533; discussion 1533.

179. Ezzati M, Lopez AD (2003) Estimates of global mortality attributable to smoking in 2000. *Lancet* 362: 847–852.

180. Greenland S (1998) Applications of Stratified Analysis Methods. In: Rothman KJ, Greenland S, editors. *Modern epidemiology*. 2nd ed. Lippincott Raven. pp. 295–297.

181. Rose G (2001) Sick individuals and sick populations. *Int J Epidemiol* 30: 427–432; discussion 433–424.

182. Vandenbroucke JP, Koster T, Briet E, Reitsma PH, Bertina RM, et al. (1994) Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet* 344: 1453–1457.
183. Botto LD, Khoury MJ (2001) Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. *Am J Epidemiol* 153: 1016–1020.
184. Wei L, MacDonald TM, Walker BR (2004) Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Ann Intern Med* 141: 764–770.
185. Martinelli I, Taioli E, Battaglioli T, Podda GM, Passamonti SM, et al. (2003) Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med* 163: 2771–2774.
186. Kyzas PA, Loizou KT, Ioannidis JP (2005) Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 97: 1043–1055.
187. Rothman KJ, Greenland S, Walker AM (1980) Concepts of interaction. *Am J Epidemiol* 112: 467–470.
188. Saracci R (1980) Interaction and synergism. *Am J Epidemiol* 112: 465–466.
189. Rothman KJ (2002) *Epidemiology. An introduction*. Oxford: Oxford University Press. pp. 168–180.
190. Rothman KJ (1986) *Interactions Between Causes*. Modern epidemiology. Boston: Little Brown. pp. 311–326.
191. Hess DR (2004) How to write an effective discussion. *Respir Care* 49:1238–1241.
192. Horton R (2002) The hidden research paper. *JAMA* 287: 2775–2778.
193. Horton R (1995) The rhetoric of research. *BMJ* 310: 985–987.
194. Docherty M, Smith R (1999) The case for structuring the discussion of scientific papers. *BMJ* 318: 1224–1225.
195. Perneger TV, Hudelson PM (2004) Writing a research article: advice to beginners. *Int J Qual Health Care* 16: 191–192.
196. *Annals of Internal Medicine*. Information for authors. Available at:

http://www.annals.org/shared/author_info.html. Accessed 10 September 2007.

197. Maldonado G, Poole C (1999) More research is needed. *Ann Epidemiol* 9: 17–18.
198. Phillips CV (2001) The economics of ‘more research is needed’. *Int J Epidemiol* 30: 771–776.
199. Winkleby MA, Kraemer HC, Ahn DK, Varady AN (1998) Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988–1994. *JAMA* 280: 356–362.
200. Galuska DA, Will JC, Serdula MK, Ford ES (1999) Are health care professionals advising obese patients to lose weight? *JAMA* 282: 1576–1578.
201. Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15: 72–101.
202. Fuller WA, Hidirolou MA (1978) Regression estimates after correcting for attenuation. *J Am Stat Assoc* 73: 99–104.
203. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, et al. (1990) Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 335: 765–774.
204. Phillips AN, Smith GD (1991) How independent are “independent” effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 44: 1223–1231.
205. Phillips AN, Smith GD (1992) Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med* 11: 953–961.
206. Greenland S (1980) The effect of misclassification in the presence of covariates. *Am J Epidemiol* 112: 564–569.
207. Poole C, Peters U, Il'yasova D, Arab L (2003) Commentary: This study failed? *Int J Epidemiol* 32: 534–535.
208. Kaufman JS, Cooper RS, McGee DL (1997) Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology* 8: 621–628.
209. Greenland S (1990) Randomization, statistics, and causal inference. *Epidemiology* 1: 421–429.
210. Taubes G (1995) Epidemiology faces its limits. *Science* 269: 164–169.

211. Temple R (1999) Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA* 281: 841–844.
212. Greenberg RS, Shuster JL Jr. (1985) Epidemiology of cancer in children. *Epidemiol Rev* 7: 22–48.
213. Kushi LH, Mink PJ, Folsom AR, Anderson KE, Zheng W, et al. (1999) Prospective study of diet and ovarian cancer. *Am J Epidemiol* 149: 21–31.
214. Kemmeren JM, Algra A, Meijers JC, Tans G, Bouma BN, et al. (2004) Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V Leiden mutation: a randomized trial. *Blood* 103: 927–933.
215. Egger M, May M, Chene G, Phillips AN, Ledergerber B, et al. (2002) Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet* 360: 119–129.
216. Campbell DT (1957) Factors relevant to the validity of experiments in social settings. *Psychol Bull* 54: 297–312.
217. Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. *Ann Intern Med* 130: 515–524.
218. Krinsky S, Rothenberg LS (2001) Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics* 7: 205–218.
219. Bekelman JE, Li Y, Gross CP (2003) Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 289: 454–465.
220. Davidson RA (1986) Source of funding and outcome of clinical trials. *J Gen Intern Med* 1: 155–158.
221. Stelfox HT, Chua G, O'Rourke K, Detsky AS (1998) Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med* 338: 101–106.
222. Lexchin J, Bero LA, Djulbegovic B, Clark O (2003) Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 326: 1167–1170.
223. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL (2003) Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 290: 921–928.

224. Barnes DE, Bero LA (1998) Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 279: 1566–1570.
225. Barnes DE, Bero LA (1996) Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law* 21: 515–542.
226. Glantz SA, Barnes DE, Bero L, Hanauer P, Slade J (1995) Looking through a keyhole at the tobacco industry. The Brown and Williamson documents. *JAMA* 274: 219–224.
227. Huss A, Egger M, Hug K, Huwiler-Muntener K, Roosli M (2007) Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environ Health Perspect* 115: 1–4.
228. Safer DJ (2002) Design and reporting modifications in industry- sponsored comparative psychopharmacology trials. *J Nerv Ment Dis* 190: 583–592.
229. Aspinall RL, Goodman NW (1995) Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ* 311: 844–846.
230. Chan AW, Hrobjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 291: 2457– 2465.
231. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003) Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 326: 1171–1173.
232. Scherer RW, Langenberg P, von Elm E (2005) Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: MR000005. Available: <http://www.cochrane.org/reviews/en/mr000005.html>. Accessed 10 September 10, 2007.
233. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357: 1191–1194.
234. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, et al. (2000) Meta-analysis of

observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283: 2008–2012.

235. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, et al. (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134: 663–694.