

NECA 비둘림위험 평가도구 매뉴얼

- 부 록 -



NECA 비뿔림위험 평가도구 매뉴얼

- 부 록 -



목차

PART 1. 체계적문헌고찰 비뚤림위험 평가 도구

1. AMSTAR 2 : 원본 매뉴얼 번역	10
2. ROBIS : 원본 매뉴얼 번역	20
2.1. 1단계: 관련성의 평가(해당하는 경우)	21
2.2. 2단계: 문헌고찰 과정의 우려사항 파악	22
2.3. 3단계: 비뚤림 판단	46
2.4. ROBIS 사용에 있어 현실적인 이슈	52

PART 2. Cochrane RoB 2 : 원본 매뉴얼 번역

1. 서론	54
1.1 신호 질문	55
1.2. 비뚤림위험 평가	57
1.3. 관심 효과 지정	59
2. RoB 2 실행에 있어서 이슈	63
2.1. 다중 평가	63
2.2. 자료수집 프로세스	63
2.3. 비뚤림위험 평가의 제시	63
2.4. 신속 평가	64
3. 상세한 지침: 사전 고려사항	65
4. 무작위 배정 과정에서 발생하는 비뚤림	67
4.1. 배경	67
4.2. 무작위 배정 과정에서 발생하는 비뚤림의 경험적 근거	70
4.3. 현재 영역 사용의 실제	70
4.4. 신호 질문 및 비뚤림위험 판단 기준	76
5. 의도한 중재에서 이탈로 인한 비뚤림	81
5.1. 배경	81
5.2. 의도한 중재에서 이탈로 인한 비뚤림의 경험적 근거	88
5.3. 도구의 이 영역 사용하기	89
5.4. 신호 질문 및 비뚤림위험 판단 기준	91
6. 중재결과 자료의 결측으로 인한 비뚤림	102
6.1. 배경	102
6.2. 중재결과 자료 결측으로 인한 비뚤림의 경험적 근거	111
6.3. 이 영역 도구의 사용	111
6.4. 신호 질문 및 비뚤림위험 판단 기준	112

7. 중재결과 측정의 비뚤림	116
7.1 배경	116
7.2. 중재결과 측정에서 비뚤림의 경험적 근거	121
7.3. 이 영역 도구의 사용	121
7.4. 신호 질문 및 비뚤림의 위험성 판단 기준	123
8. 보고된 연구결과 선택의 비뚤림	127
8.1 배경	127
8.2. 보고된 연구결과를 선택할 때 비뚤림에 대한 경험적 근거	131
8.3. 도구의 이 영역 사용	131
8.4. 비뚤림의 위험을 판단하기 위한 신호 질문 및 기준	133
References	138

PART 3. ROBINS-I : 원본 매뉴얼 번역

1. 배경	148
1.1. 도구의 맥락	149
1.2. 목표 임상시험(target trial)과 관련해서 비뚤림위험을 평가하기	149
1.3. 비뚤림의 영역	150
1.4 연구설계	153
1.5. 비뚤림위험 평가는 특정 중재 효과와 관련이 되어야 함	154
1.6. 이 문서의 구조	154
2. 도구를 사용하기 위한 지침: 일반적인 고려사항	155
2.1. 프로토콜 단계	155
2.2. 개별 연구에서 사전에 고려할 사항	156
2.3. 신호 질문	162
2.4. 비뚤림위험에 대한 영역-수준의 판단	162
2.5. 비뚤림위험에 대해 전체적으로 판단하기	164
2.6 문헌고찰에서 여러 중재결과에 대해 비뚤림위험을 평가하기	166
3. 도구 사용 지침: 각 비뚤림 영역에 대한 자세한 지침	168
3.1. 세부 지침: 교란으로 인한 비뚤림	168
3.2. 세부 지침: 연구대상자 선택의 비뚤림	178
3.3. 세부 지침: 중재 분류의 비뚤림	183
3.4. 세부 지침: 의도한 중재에서 이탈되어 나타나는 비뚤림	186
3.5. 세부 지침: 결측치로 인한 비뚤림	196
3.6. 세부 지침: 중재결과 측정 비뚤림	200
3.7. 세부 지침: 보고된 연구결과 선택의 비뚤림	204
References	201
부록: 용어 해설	211

표목차

표 1-1. 영역 1에 대한 신호 질문들과 평가지침	24
표 1-2. 연구 적격 기준에 대한 우려 정도	25
표 1-3. 연구 적격 기준에 대한 평가 예시('높음'으로 평가)	26
표 1-4. 연구 적격 기준에 대한 평가 예시('낮음'으로 평가)	28
표 1-5. 영역 2에 대한 신호질문과 평가지침	30
표 1-6. 연구의 검색 및 선택에 대한 우려 정도	31
표 1-7. 연구 검색 및 선택에 대한 평가 예시('높음'으로 평가)	32
표 1-8. 연구 검색 및 선택 평가 예시('낮음'으로 평가)	34
표 1-9. 영역 3에 대한 신호 질문들과 평가지침	36
표 1-10. 자료수집과 연구평가에 대한 우려	37
표 1-11. 우려가 높은 것으로 판단된 자료 수집 및 연구 평가에 대한 평가 예	38
표 1-12. 우려가 낮은 것으로 판단된 자료 수집 및 연구 평가에 대한 평가 예	39
표 1-13. 영역 4에 대한 신호 질문들과 평가지침	41
표 1-14. 합성 및 연구 결과에 대한 우려	43
표 1-15. 합성에 대한 등급부여 예시('높음'으로 평가)	44
표 1-16. 낮은 우려로 판단된 합성의 등급부여 예시	45
표 1-17. 3단계에 대한 신호질문과 평가 방법	47
표 1-18. 연구를 확인하고 선택하기 위해 사용된 방법에 의해 개입된 비뚤림위험	47
표 1-19. 2단계 평가 중 확인된 우려 사항 요약	48
표 1-20. 비뚤림위험 평가 예시('높음'으로 평가)	49
표 1-21. 2단계 평가 중 확인된 우려 사항 요약	50
표 1-22. 비뚤림위험이 낮은 것으로 판단된 합성에 대한 등급부여 예시	51
표 2-1. 특정 중재 결과에 대한 전반적인 비뚤림위험 평가 방법	58
표 2-2. 배정순서 은폐에 대해 판단하는 최소 기준, 확장 기준	72
표 2-3. 무작위 배정 과정에서 발생하는 비뚤림위험 판단 기준	79
표 2-4 무작위 배정 과정에서 발생하는 비뚤림위험 판단에 대한 신호질문 맵핑	80
표 2-5. 의도한 중재에서 이탈로 인한 비뚤림(중재배정효과) 판단 기준	93

표 2-6. 의도한 중재에서 이탈로 인한 비돌림(중재 배정 효과) 위험 판단에 대한 신호질문 맵핑	94
표 2-7. 의도한 중재에서 이탈로 인한 비돌림(중재 준수 효과) 판단 기준	98
표 2-8. 의도한 중재에서 이탈로 인한 비돌림(중재 준수 효과) 위험 판단에 대한 신호질문 맵핑	100
표 2-9. 중재결과 자료의 결측으로 인한 비돌림 판단 기준	114
표 2-10. 중재결과 자료의 결측으로 인한 비돌림위험 판단에 대한 신호질문 맵핑	115
표 2-11. 결과자료 측정의 비돌림위험 판단 기준	125
표 2-12. 결과자료 측정의 비돌림위험 판단 신호질문 맵핑	126
표 2-13. 보고된 연구결과 선택의 비돌림위험 판단	136
표 2-14. 보고된 연구결과 선택의 비돌림위험 판단 신호질문 맵핑	137
표 3-1. ROBINS-I 도구에 포함된 비돌림 영역	152
표 3-2. 특정 중재결과에서 전체 RoB 판단하기	165
표 3-3. 특정 중재결과에 대해 전체 RoB 판단하기	166
표 3-4. 서로 다른 중재결과에 대한 다른 RoB 판단 예시	167
표 3-5. 교란으로 인한 비돌림에 대한 비돌림위험 판단하기	178
표 3-6. 연구대상자의 선택에서 비돌림위험 판단하기	182
표 3-7. 중재 분류의 비돌림에 대한 비돌림위험 판단하기	184
표 3-8. 의도한 중재에서 이탈되어 나타나는 비돌림에 대한 비돌림위험 판단하기	195
표 3-9. 결측치로 인한 비돌림의 비돌림위험 판단하기	199
표 3-10. 중재결과 측정 비돌림의 비돌림위험 판단하기	203
표 3-11. 보고된 연구결과 선택 비돌림의 비돌림위험 판정	207

Box 목차

Box 1-1. 체계적 문헌고찰의 여러 종류에 대한 목표질문과 PICO 형식의 예시	21
Box 2-1. 무작위 임상시험에서 비뚤림의 경험적 근거: 메타 역학(meta-epidemiology)의 역할	55
Box 2-3. 적절한 배정 순서 은폐의 예(Schulz and Grimes(39)에서 편집)	72
Box 2-4. 무작위 배정 과정에서 발생하는 비뚤림위험	77
Box 2-5. 의도한 중재에서 이탈이 있는 연구의 예	86
Box 2-6. The RoB 2 tool: 의도한 중재에서 이탈로 인한 비뚤림(중재 배정 효과)	91
Box 2-7. ROB 2 도구: 의도한 중재에서 이탈로 인한 비뚤림위험(중재 준수 효과)	96
Box 2-8. 중재결과 자료의 결측으로 인한 비뚤림	112
Box 2-9. 중재 효과 추정치에 대한 비차등적 측정 오류의 의미	116
Box 2-10. 중재결과 측정의 비뚤림위험	123
Box 2-11. RoB 2 도구: 보고된 연구결과 선택의 비뚤림위험	133
Box 3-1. ROBINS-I 도구(1단계): 프로토콜 단계	156
Box 3-2. ROBINS-I 도구(2단계, 파트 1): 개별 연구에서: 평가 셋업하기	159
Box 3-3. ROBINS-I 도구(2단계, 파트 2): 개별 연구에서: 교란 영역과 공동 중재의 평가	160
Box 3-4. ROBINS-I tool(2단계, 파트 3): 교란으로 인한 비뚤림위험	176
Box 3-5. ROBINS-I 도구(2단계, 파트 4): 연구대상자 선택의 비뚤림위험	181
Box 3-6. ROBINS-I 도구(2단계, 파트 5): 중재분류의 비뚤림위험	185
Box 3-7. ROBINS-I 도구(2단계, 파트 6): 의도한 중재에서 이탈되어 나타나는 비뚤림위험	193
Box 3-8. ROBINS-I 도구(2단계, 파트 7): 결측치로 인한 비뚤림위험	198
Box 3-9. ROBINS-I 도구(2단계, 파트 8): 중재결과 측정 비뚤림위험	202
Box 3-10. ROBINS-I 도구(2단계, 파트 9): 보고된 연구결과 선택의 비뚤림위험	208

그림 목차

그림 2-1. 무작위 배정과정에서 발생하는 비뚤림위험 판단을 위한 알고리즘	80
그림 2-2. 의도한 중재에서 이탈로 인한 비뚤림위험 판단 알고리즘(중재 배정 효과)	95
그림 2-3. 의도한 중재에서 이탈로 인한 비뚤림위험 판단 알고리즘(중재 준수 효과)	101
그림 2-4. 중재결과 자료의 결측으로 인한 비뚤림위험 판단 알고리즘	115
그림 2-5. 결과자료 측정의 비뚤림위험 판단 알고리즘.	126
그림 2-6. '보고된 연구결과 선택의 비뚤림'과 '중재결과 비보고 비뚤림'의 예시	128
그림 2-7. 보고된 연구결과 선택의 비뚤림위험 판단 알고리즘	137

PART 1

체계적문헌고찰 비뚤림위험 평가 도구

1. AMSTAR 2 : 원본 매뉴얼 번역	10
2. ROBIS: 원본 매뉴얼 번역	20



1. AMSTAR 2 : 원본 매뉴얼 번역

AMSTAR 2의 문항들은 내용만 보면 그 뜻을 알 수 있도록 되어 있다. 그러나 기저에 있는 이슈는 복잡할 수 있고, 특히 광범위한 증재에 걸쳐 판단이 내려질 때 다양한 해석이 가능할 수 있다. 이 문서자료는 AMSTAR 2의 사용에 대한 추가 지침을 제공한다. 다시 한 번 강조하지만 이 문서는 지침이다. 이 문서는 우리가 잘 정의된(대개 임상적) 증재에 대해 문헌고찰이 수행되는 경우 기준을 어떻게 적용해야 하는지에 대한 것이다. 물론, 개별 사용자는 개별 영역을 다룰 때 혹은 전체적인 평가를 할 때 이 지침에서 벗어나는 것이 필요할 수 있다. 그런 경우에도 그러한 이탈을 문서화 하는 것이 필요하다.

AMSTAR 2는 전체적인 '점수'를 매기도록 되어 있지 않다. 높은 점수 자체로 부적절한 문헌 검색 혹은 체계적인 문헌고찰에 포함된 개별 연구의 비뮌림위험(ROB)을 평가하지 못하는 것 같은 특정 영역의 결정적인 약점이 감추어 질 수도 있다. 체계적 문헌고찰에 대한 전반적인 등급을 매길 때 중요한 영역의 결함을 고려하는 것이 중요하다. 왜냐하면 그러한 결함은 체계적 문헌고찰에 대한 확신도를 크게 약화시킬 수 있기 때문이다.

문항 1: 체계적 문헌고찰의 연구질문과 포함기준에는 PICO의 구성요소가 포함되었는가?

PICO(인구, 증재, 대조군 및 증재결과)를 연구질문 체계화를 위한 틀로 사용하는 것이 일반적이다. 만일 시점이 연구에서 관련 임상 결과를 포착할 가능성을 결정하는 데 중요한 경우(예: 증재의 영향은 몇 년 후에만 예상됨), 시간 프레임을 추가해야 한다. PICO를 통해 체계적 문헌고찰 보고서에 기술해야 하는 요소를 확인할 수 있고, 평가자가 연구 선택과 결함 가능성을 판단할 수 있다. 또한 PICO를 통해 문헌고찰 사용자가 증재결과의 적용 가능성을 결정할 수 있다. 체계적 문헌고찰 저자들이 항상 PICO 요소를 명시적으로 기술하는 것은 아니지만 초록, 서론, 방법 파트를 주의 깊게 읽음으로써 확인할 수 있어야 한다.

'예' : 평가자는 PICO의 4가지 요소가 보고서의 어딘가에 기술되어 있음을 확인한다.

문항 2. : 체계적 문헌고찰 방법론이 실제 문헌고찰을 시행하기 전에 확립되었으며, 보고서에는 프로토콜로부터 중대한 이탈이 있는 경우 이에 대한 정당화(합당한 이유)가 제시되었나?

체계적 문헌고찰은 관찰 연구의 한 형태이기 때문에 문헌고찰이 시작되기 전에 문헌고

찰 방법을 합의해야 한다. 잘 개발된 프로토콜을 준수하면 문헌고찰의 비뮐림위험이 감소한다. 저자들은 그들이 서면프로토콜을 통해 작업했다는 사실이 객관적으로 입증가능하도록 하여야 한다. 그러한 근거에는 등록처(예: SPRIPO - <https://www.crd.york.ac.uk/PROSPERO/>), 공개 간행물 저널(예: BMJ Open) 또는 기관 사무처 또는 연구 윤리 위원회에 제출일이 기록된 제출 서류들이 있다.

문헌고찰에 앞서 연구질문과 문헌고찰 연구 방법을 미리 계획해야 하고, 최소한 보고서에 이를 명시해야 한다(‘일부 예’).

‘예’로 체크하려면, 문헌고찰이 진행되기 전에 등록처 또는 다른 독립 기관(예: 연구 윤리 위원회 또는 기관 사무처)에 서면 프로토콜이 등록되어 있어야 한다.

평가자는 문헌고찰의 공개된 보고서를 등록된 프로토콜과 비교해야 한다. 프로토콜에서 이탈이 있는 경우 문헌고찰 저자가 이를 보고하고 정당화하는지 여부를 결정해야 한다. 설명되지 않은 명백한 불일치가 있는 경우 등급을 하향평가 해야 한다.

문항 3: 문헌고찰 저자는 문헌고찰에 포함될 연구설계 선택에 대해 설명하였나?

체계적인 문헌고찰에 포함하기 위한 연구 유형 선택이 자의적이어서는 안 된다. 저자들은 그들이 전략을 따랐다는 것을 표시해야 한다. 일반적인 규칙은 RCT로 제한하면, 문헌고찰이 치료의 효과에 대해 불완전하게 평가할 것인가를 물어보는 것이다(이것은 저자들이 실제로 기록한 것에서 미루어 추측해야 할 수도 있음). 이러한 질문은 다시 말해 관련 RCT가 없거나[보통 위해] 사용 가능한 RCT 에 중재결과 결측, 통계적 검정력 부족, 인구집단 제한 또는 중재나 대조군의 대표성 부족 등의 문제가 있는지를 보는 것이다. 이 일반적인 질문에 대한 대답이 ‘예’인 경우 중재 비무작위 연구를 포함하는 것이 정당화된다. 반대로, RCT만으로 문헌고찰을 제한하는 것을 정당화하기 위해서는, 그것만으로 관심 효과에 대한 완전한 그림을 제공할 수 있다고 주장해야 한다. RCT가 필요한 중재결과 자료를 제공할 수 없는 경우 또는 RCT의 문헌고찰이 이미 출판이 되었고, NRSI의 문헌고찰이 이미 알려진 것을 보완할 경우, NRSI로만 문헌고찰을 제한하는 것은 정당화된다. RCT와 NRSI를 모두 포함하는 전략은 중재와 관련된 효과와 위해를 완전히 파악할 수 있다면 정당화될 수 있다. 만일 그렇다면 이 두 가지 유형의 연구를 독립적으로 평가하고 결합할 것(메타 분석이 적절한 경우)을 권고한다. 이 영역은 다소 소홀히 다루어진 것도 사실이며, 지침이 있더라도 문헌고찰이 현재의 평가 기준을 충족하는지를 판단하기 어려울 수 있다. 연구설계의 선택에 대한 정당성은 전체 연구 보고서를 주의 깊게 읽은 것에서 추론해야 할 수 있다.

문항 4: 문헌고찰 저자는 포괄적인 문헌 검색 전략을 사용하였는가?

적어도 두 개의 서지 데이터베이스를 검색해야 한다. 보고서에는 조사 연도 및 데이터베이스 이름(예: Central, EMBASE 및 MEDLINE)이 제시되어야 한다. 키워드, MESH 용어를 보고하고, 필요에 따라 전체 검색 전략을 제시할 수 있어야 한다. 기존 문헌고찰, 전문 저장소에 있는 프로토콜을 검토하여 검색을 보완해야 한다. 또는 특정 연구 분야의 전문가를 포함하거나 자문을 받고, 발견된 연구의 참조 목록을 검토하여 보완해야 한다. 때때로 연구결과를 확인 하거나 업데이트본이나 수정본을 얻기 위해 원본 연구 저자들에게 접근해야 할 필요가 있다. 모든 언어의 출판물을 찾아야 하며, 언어 제한이 있으면 정당화해야 한다. 우리는 경우에 따라 회색 문헌을 검색할 필요성을 강조해 왔다. 회색 문헌은 때때로 웹 사이트(예: 정부, 비정부기구, 보건의료 기술 기관)에서만 이용할 수 있는 정책보고서, 프로그램 평가 보고서와 함께 중요하다. 회색문헌은 동료평가가 되지 않았을 수 있지만, 혹시 그런 평가가 있는지 확인해 보아야 한다. 회색 문헌이 중요하게 여겨지는 경우, 저자들은 적절한 자료원을 검색해 보아야 하는데 임상시험 등록처, 학술대회 초록집, 학위 논문, 개별 웹사이트(예: 대학, 리서치게이트) 등이 그런 곳이다. 게다가 의료 중재에 대한 임상시험 중 상당수는 동료심사 학술지에 출판되지 않아서 회사 스폰서나 연구자로부터 직접 입수해야 할 수도 있다. ‘예’에 체크하려면 문헌고찰 저자가 다른 검색의 모든 측면이 만족스러워야 한다.

문항 5: 문헌고찰 저자는 연구 선택을 중복으로 수행하였는가?

모범 사례가 되려면 두 명의 문헌고찰 저자가 체계적인 문헌고찰에 포함시킬 연구의 적격성을 결정하여야 한다. 그것은 연구의 특성(제목, 초록, 전문으로 부터)을 확인하여 연구질문 요소를 만족하는지 살펴보는 것이다. 응답 옵션의 하나로 우리는 연구들 표본에서 시행한 연구 선택에 대한 평가자간 일치도 결과를 통해 문헌고찰 저자의 수행 정도를 평가할 수 있다. 연구 선택에 이견이 생겼을 때 합의 과정이 있어야 한다. 한 명이 모든 연구 선택을 수행하려면 연구 표본에 대한 합의를 확인하여 두 사람 사이에 카파 점수에서 ‘강한’ 일치도(0.80 이상)를 보여야 한다.

문항 6: 문헌고찰 저자는 자료추출을 중복으로 수행하였는가?

문항 5에서와 같이, 자료 추출을 수행하는 독립적인 평가자가 적어도 두 명 이상 있어야 한다. 의견 불일치가 생겼을 때는 합의 과정을 이용해야 한다. 한 명이 자료 추출을 수행하고자 하는 경우에는, 두 번째 검토자와 연구 샘플에 대한 일치도를 확인하고 카파 점수 0.80 이상을 달성해야 한다.

문항 7: 문헌고찰 저자는 배제 연구에 대한 목록과 합당한 배제사유를 제공하였는가?

이 문항에서는 문헌고찰 저자가 각 연구의 배제에 대한 정당성을 가지는지, 그리고 잠재적으로 관련이 있는 연구의 전체 목록을 제공하는지를 평가한다. 부적절함/부적합한 인구 집단, 중재 및 비교중재에 기초하여 다양한 이유로 연구가 배제될 수 있다. 배제를 비뚤림위험에 따라 하면 안 되며, 비뚤림위험은 별도로, 문헌고찰 과정에서 나중에 다루어야 한다. 합당하지 않은 배제는 문헌고찰 결과를 편향시킬 수 있으며 문헌고찰 초기 단계에서 포용적인(inclusive) 접근법을 권장한다. 이 문항에서는 문헌고찰 저자가 잠재적으로 관련이 있는 연구의 전체 목록을 제공하고 각 연구의 배제에 대한 합당한 사유를 제공할 것을 요구한다.

문항 8: 문헌고찰저자는 포함된 연구들의 세부사항을 적절히 기술하였는가?

연구대상자, 중재, 비교중재, 중재결과, 연구설계, 분석 및 세팅에 대한 설명이 제공되어야 한다. 이러한 설명은 평가자 또는 사용자가 연구가 적절하게 선택되었는지(PICO 구조와 관련하여) 그리고 연구의 인구집단과 중재가 자신의 임상진료와 정책과 관련이 있는지에 대해 판단하기에 충분해야 한다. 또한 중재효과의 이질성을 연구하기 위한 틀도 제공하여야 한다(예: 용량, 연령 범위, 임상 세팅 등).

문항 9: 문헌고찰저자는 문헌고찰에 포함된 개별 연구의 비뚤림위험(ROB)을 평가하기 위해 만족스러운 도구를 사용하였는가?

이는 특히 NRSI(중재 비무작위 연구)를 포함하는 체계적 문헌고찰 평가에 있어 중요한 부분이다. 핵심적인 내용은 문헌고찰 저자들이 연구결과를 요약하고 해석할 때 비뚤림위험을 고려했는지 여부다.

문헌고찰이 무작위배정 비교임상시험으로 제한되는 경우 검토자가 개별 RCT에 대한 ROB에 대해 적절한 평가를 수행했는지 여부를 코크란 핸드북에 따라 판단해야 한다. 여전히 비무작위 연구의 ROB 평가는 도전적이다.

문헌고찰 저자는 ROB 평가에 체계적인 접근법을 사용해야 하며, 가급적 적절히 개발된 평가 도구를 사용해야 한다. 표준적인 도구가 아니라면 매우 심각한 방법론적 결함만을 감지할 수 있다는 점을 알아야 한다. 뉴캐슬 오타와 척도, SIGN, MMAT(Mixed Methods Evaluation Tool) 등 여러 ROB 평가 도구가 흔히 사용된다. 가장 포괄적인 평가 도구는 최근에 도입된 코크란 도구인 ROBINS-I이다. 하지만 문헌고찰이 수행될 당시에는 이 도구가 없었을 수 있다.

AMSTAR 2를 개발하면서 우리는 RCTs:(<http://handbook.cochrane.org/>)

chapter_8/8_5_the_cochrane_collaborations_tool_for_assessing_risk_of_bias.htm)과 NRSI(www.riskofbias.info)와 함께 Cochrane ROB 도구라는 그림을 그렸다. 두 도구의 영역 평가분항에 이 도구에 있는 내용이 있다. 문헌고찰 저자가 어떠한 도구를 사용했더라도, 도구의 문항 9에 나열된 내용을 만족시켜야 한다.

여기에 제시된 지침이 포괄적이지 않을 수 있다. AMSTAR 2는 가장 일반적으로 인식되는 비뚤림영역만을 다룬다는 사실을 유의해야 한다. 비뚤림위험을 심층적으로 평가하기 위해서는 전문가 도움이 필요하다. 문헌고찰 저자가 RoB를 어떻게 평가하였는지 평가할 때, 만일 현재 팀에서 그런 조언을 얻기 힘들면 자문해줄 방법론 전문가와 내용 전문가를 찾아봐야 할 수 있다. 그들에게 어떤 교란 요인이 중요한지, 고려중인 문헌고찰과 관련이 있는 선택 비뚤림과 측정 비뚤림을 확인하는 방법에 대한 구체적인 조언을 들어야 한다. 또한 교란 요인에 대한 보정 방법에 대한 조언도 구해야 한다.

NRSI를 포함하는 체계적 문헌고찰과 가장 관련이 있는 영역의 목록은 아래와 같다. 그것은 ROBINS-I/ACROBAT-NRSI 도구에도 포함되어 있다.

교란. 교란은 분석 중에 두 가지 관련 증재 또는 노출(예: 흡연과 음주)의 영향이 분리되지 않은 경우에 발생한다. 교란으로 인해 다른 변수 때문에 효과가 있음에도 특정 변수가 효과가 있는 것처럼 보일 수 있다. 증재의 교란 요인 중 중요한 것은 치료의 적응증에 대한 것이다. 예를 들어, 동반질환이 더 많은 환자에게 실험 약물을 대조 약물보다 더 많이 처방하고, 이러한 동반질환이 관심 있는 증재결과와 관련이 있는 경우이다. 이러한 평가를 위해서는 일반적으로 개별 연구에서 보고된 기저상태 자료를 봐야 한다. 기저상태 교란 요인은 여러 가지 방법으로 해결 가능한데, 설계 과정(예: 성향 점수에 의한 짝짓기), 보정(예: 로지스틱 회귀), 기타(예: 도구변수 분석, ‘tracer’ 노출) 등 여러 가지가 있다. 정교한 방법으로 보정을 해도 모든 교란 변수를 보정할 수 없다고 보아야 한다. 따라서 아무리 통계적으로 유의미해도 약한 연관성이라면 해석에 주의를 기울여야 한다.

표본 선택 비뚤림. 이는 대상자가 노출과 증재결과 사이의 진정한 관련성을 직접 왜곡하는 편향된 방식으로 표본이 추출될 때 발생한다. 교란의 경우와 같이 제 3의 요소가 필요하지는 않다. 예를 들어, 금연 클리닉에 의뢰된 환자 중 연구대상자를 모집하여 흡연과 심장질환과의 연관성을 연구해서는 안 된다. 관심 노출을 조사할 대상자를 선정할 때는 그들이 증재결과와 관련이 없어야 한다. 마찬가지로, 증재결과를 가진 대상자의 선택은 노출 상태와 관련이 없어야 한다. 선택시점도 중요할 수 있다. 대상자로 등록되기 전에 일정기간 동안 약

물을 사용한 적이 있는 경우(기존 사용자), 이들은 유해 중재결과의 위험이 더 낮은 집단이 될 것이다.

이러한 이유 때문에, 약물역학연구에서는 신규 약물사용자를 모집한다. 다른 시간적 표본 비뚤림(불멸의 시간 비뚤림, 개시 비뚤림)이 때때로 중요하다. 자세한 내용은 사용자가 ROBINS-I 지침 문서를 참조하는 것이 좋다.

노출 및 중재결과 측정의 비뚤림 : 정확한 기록이 실시간으로 이루어지지 않으면 노출 또는 치료 내용이 잘못 분류될 수 있다. 일반적으로 현대의 약물역학연구는 처방전 또는 조제 기록을 소비에 대한 대리 지표로 사용한다. 그러나 처방된 치료에 대한 준수가 100%가 아니므로 이 방법으로 실제 소비를 잘못 분류할 수 있다. 이 오류가 비차등적이라면, 무효점(null)을 향하는 비뚤림이 될 것이다. 그러나 일부 연구 분야에서는 회상에 의존한다(예 : 자외선 노출과 흑색종). 이로 인해 차등적 오분류가 발생할 수 있다. 예를 들어, SIDS로 사망한 어린이의 부모는 사건발생 전에 아기에게 투여한 모든 약물에 대한 회상이 강화될 수 있다. 중재결과 측정 또한 오분류에 의해 영향을 받을 수 있으며, 이것이 비차등적인 경우, 일반적으로 무효점으로 향하는 비뚤림이 될 것이다. 그러나 비차등적 오분류도 비뚤림을 유발할 수 있다. 예를 들어, 다리 초음파가 경구 피임약을 복용하는, 부종과 통증이 있는 다리를 가진 여성에서 자주 수행된다면, 초음파 클리닉에서 대상자를 선택하는 것은 DVT와 경구 피임약 간의 관련성에 대한 연구가 편향될 수 있다.

중재결과 및 분석에 대한 선택적 보고 : 대규모 관찰연구는 정의된 인구집단에서 발생하는 많은 중재결과를 기록하는 인구집단 데이터베이스를 분석할 수 있다. 만약 중재결과가 사전에(바람직하게는 등록된 프로토콜에) 지정되어 있지 않다면, 연구자들은 다중 중재결과를 분석하고 노출된 대상자와 비노출 대상자간에 다르게 나온 중재결과를 선택적으로 보고하려는 유혹을 받을 수 있다.

또한, 비무작위 자료세트를 분석할 수 있는 몇 가지 잠재적인 방법이 있다(예 : 중재를 분류하는 다른 방법 또는 다양한 다변수 보정 모델 포함). 분석 프로토콜이 연구 수행 이전에 구체화되지 않은 경우, 다른 분석에서는 분명하지 않은, 유의미한 통계적 차이를 나타내는 것으로 보이는 한 세트의 분석을 선택할 수 있다. 문헌고찰 저자는 연구저자가 결과를 사전에 지정하고 분석하였는지 여부를 결정해야 한다. 향후 더 많은 연구가 수행되기 전에 등록될 것이기 때문에 앞으로는 이러한 분석이 용이해질 것이다.

문항 10. 문헌고찰 저자는 문헌고찰에 포함된 연구들의 자금 출처에 대해 보고하였는가?

여러 조사에 따르면, 상업적으로 후원되는 연구가 독립적으로 자금을 지원받은 연구보다 후원자의 제품에 유리한 연구결과가 나올 가능성이 더 높다. 문헌고찰 저자는 문헌고찰에 포함된 각 연구에 대한 자금 출처를 상세히 기술하거나 연구 보고서에 정보가 제공되지 않았다는 것을 기술하는 것이 중요하다. 이 정보에 따라 상업적으로 자금을 조달한 연구결과와 독립적으로 자금을 조달한 연구결과를 별도로 분석할 수도 있다.

문항 11. 메타분석을 수행하였다면, 문헌고찰 저자는 이에 대한 합당한 이유를 제시하였고, 연구결과의 통계학적 결함을 위해 적절한 방법을 사용하였는가?(이 문항은 합성기법으로 메타 분석이 보고 된 경우에만 작성하십시오.)

문헌고찰 저자는 포함된 연구 자료에 대한 메타 분석을 수행하기로 한 의사결정의 기반이 되는 원칙을 문헌고찰 프로토콜에 명확히 제시해야한다.

그러한 원칙에는 단일 통합 효과(예 : 다수의 호환 가능한, 검정력이 낮은 연구에서 얻은)를 얻는 것과 연구의 비교성 정도(인구집단, 중재, 비교중재 측면에서)를 평가하여 결합가능한지 살펴보는 것이 포함된다.

메타 분석이 고려되는 경우, RCT의 경우 고정 또는 변량 효과 모형을 사용하기로 결정할 이유를 설명해야 하고 이질성을 조사하기 위해 사용하려는 방법을 설명해야한다. NRSI 연구에서, 인구집단은 수십 또는 수백 명의 소규모 코호트에서 수십만 명 인구집단, 수천 개 사건에 이르기까지 규모가 다양하다. 이러한 연구결과를 규모가 작은 RCT의 연구결과와 결합될 경우, 비무작위 연구들의 자료가 통합 효과추정치를 결정할 것이다.

또한, NRSI의 결과는 다양한 비플림에 의해 영향을 받을 수 있으며, 전체 통합 추정치는 정밀하지만 비플림이 있을 수 있다. 문헌고찰 저자는 다른 연구설계에 대해 통합된 추정치를 별도로 보고해야 한다. NRSI의 경우, 통합은 매우 정밀하고 '통계적으로 유의미하지만' 효과 추정치는 편향될 수 있다. 그러나 신뢰 구간은 비플림이 없다는 가정 하에 계산된다(즉, 추정치는 동일한 대상자수를 가진 높은 질의 RCT에서 얻은 것처럼 정확함).

NRSI가 연구질문이 동일한 고품질 RCT 정도로 비플림위험이 낮은 경우는 드물기 때문에, NRSI를 기반으로 한 통합 추정치와 신뢰구간을 주의 깊게 검토해야 한다. 이 문제는 NRSI의 다양한 비플림위험과 비플림위험에 대한 불확실성을 고려할 때 중요하다.

이질성은 모든 메타분석에서 중요한 문제이다. 비무작위 연구의 경우 다른 연구에서 더 다양한 방법이 사용될 가능성이 크기 때문에, 이질성 문제는 NRSI에 대한 문헌고찰에서 특히 중요하다. 일반적인 이질성의 원인[다른 비교군, 중재결과의 기저 위험 변이 또는 연구

집단의 다른 특성, 다른 중재(예 : 용량 효과, 맥락/ 세팅, 실무자 경험) 및 다양한 중재결과 정의) 이외에도, 연구대상자의 출처, 자료의 완전성, 자료 관리 및 분석 방법에서의 이질성을 고려하는 것이 중요하다. 교란요인의 중재 효과에 대해 통계적 보정을 하면 보정되지 않은 추정치와는 상당히 다른 추정치가 도출되기도 한다. 일반적으로, NRSI 고찰 결과를 결합할 때는 원자료가 아니라 보정된 효과추정치를 대상으로 해야 한다. 만약 저자들이 그렇게 하지 않았다면 이에 대한 합당한 설명이 있어야 한다. 그러나 연구들 마다 공변량 세트(또는 상이한 방식으로 측정되거나 나름대로 적용된 공변량)를 달리해서 보정된 치료효과를 보고할 가능성이 매우 높다; 이러한 다양성은 잠재적 이질성의 또 다른 원인이 되기도 한다.

문항 12: 메타분석을 수행하였다면, 문헌고찰 저자는 개별 연구의 비뚤림위험이 메타분석 연구결과나 다른 근거 합성에 미칠 잠재적 영향을 평가하였는가?

문헌고찰 저자가 높은 질의 RCT만 선택한 경우, 연구결과에 대한 비뚤림의 잠재적 영향에 대한 논의는 거의 필요 없을 수 있다. 그러나 다양한 질을 가진 RCT를 포함시킨 경우, 비뚤림위험의 영향을 회귀분석으로 평가하거나, 비뚤림위험이 낮은 연구만으로 통합효과크기를 추정해야 한다.

NRSI의 경우, 문헌고찰 저자들은 비뚤림위험이 낮거나 증등도인 연구, 또는 낮은 연구만으로 통합효과크기를 추정해야 한다. 메타 분석(또는 회귀분석과 같은 다른 자료합성 기술)이 수행되지 않은 경우에도, 저자는 개별 연구 결과에 대한 ROB의 영향에 대한 의견을 제시해야한다.

문항 13. 문헌고찰저자가 문헌고찰 결과를 해석/논의할 때 개별 연구의 비뚤림위험을 고려하였는가?

메타 분석이 수행되지 않았더라도 문헌고찰 저자는 문헌고찰 연구결과와 해석에 있어 ROB의 영향에 대해 논의해야한다. 이는 항상 중요하지만, 문헌고찰에 비뚤림위험 정도가 다양한 RCT와 NRSI를 포함시킨 경우에 특히 중요하다. 이 논의는 통합추정치에 대한 ROB의 영향에만 국한되지 않아야하며, 개별 연구들의 연구결과 차이를 설명할 수 있는지 여부도 고려해야한다. 임상치료나 정책에 영향을 줄 수 있는 권고안을 제시 할 경우, 저자는 ROB를 명시적으로 고려해야한다.

문항 14: 문헌고찰저자는 문헌고찰 연구결과에서 발견된 이질성에 대해 만족스러운 설명을 하였는가?

RCT보다 NRSI의 연구결과에는 이질성의 잠재적 원인이 더 많이 있다. 이 도구에서는 연구설계의 차이, 분석 방법의 차이, 인구집단의 차이, 약물의 경우 용량을 포함하여 중재 강도의 차이 등을 고려한다. 문항 9에 나열된 PICO 요소와 비뮌림영역 모두가 연구결과에서 이질성의 중요한 잠재적 원인으로 간주되어야 한다. 문헌고찰 저자는 이러한 가능성을 탐색하고 연구결과 결론 및 권고안에 대한 이질성의 영향에 대해 논의해야한다.

문항 15: 양적 합성을 하였다면, 문헌고찰저자는 출판비뮌림(소규모연구 비뮌림)에 대한 적절한 조사를 수행하고, 문헌고찰 결과에 미칠 수 있는 영향에 대해 고찰하였는가?

이것은 매우 중요한 문제이지만 문헌고찰 저자와 평가자가 완전히 해결하기 어려울 수 있다. 일반적으로 통계 검정 또는 그래프 제시가 사용되며, 그 결과가 유의하면 출판 비뮌림이 있음을 나타낸다. 그러나 검정이 민감하지 않기 때문에 검정결과가 유의하지 않다고, 출판 비뮌림이 없음을 보장하지는 않는다. 출판 비뮌림의 중요성은 어느 정도 맥락과 세팅에 달려 있다. 예를 들어, 방법론적으로 강건한 산업체 후원 연구(예 : 약물, 의료기기 등)는 유사하지만 산업체에 독립적인 연구에 비해 출판 비뮌림의 영향을 받을 가능성이 더 크다.

중요한 것은 저자가(필요와 세팅에 따라) 심층적이고 집중적인 문헌 검색을 통해 출판 비뮌림을 조사하기 위해 최선을 다했는지, 연구결과에 대한 해석 및 고찰에서 출판 비뮌림의 영향에 대한 인식을 보여주었는지, 그리고 그들이 얻은 연구결과를 무효화하는데 필요한 누락된 '무효(null)' 연구의 수를 결정하기 위해 민감도 분석을 수행하였는지 여부이다.

문항 16: 문헌고찰 저자는 문헌고찰 수행을 위한 자금지원을 포함하여 잠재적 이해상충에 대해 보고하였는가?

위(ROB에서)에서 언급 한 바와 같이, 이해관계에 의해 지원된 연구는 그렇지 않은 연구보다 중재에 유리한 연구결과를 생성할 수 있다. 체계적 문헌고찰에서도 동일한 가정이 적용되며 저자는 직접적인 자금 출처를 보고해야한다.

일반적으로 학술지에는 이해상충에 대한 정보가 있다. 그러나 문헌고찰 저자의 이해 상충에 대한 평가는 그것만 확인해서는 안 된다. 그들은 다른 유대관계를 보고해야 한다. 문헌고찰은 자금 지원으로부터는 자유로울 수 있지만, 문헌고찰 저자가 체계적 문헌고찰에 포함된 제품을 제조하는 회사와 관련이 있을 수 있다. 전문가적인 이해상충은 강력하지만, 거의 보고되지 않아 식별하기가 어렵다. 연구자들이 연구 분야에 오랫동안 있어서 생기는 신념이

있을 수 있다. 이 유형의 잠재적인 이해 상충은 평가하기 어려울 수 있지만, 문헌고찰 저자가 문헌고찰 중인 분야에서 다양한 논문을 출판하였으며 그들의 연구가 해당 체계적 문헌고찰에 포함되어 있다는 사실에서 추론될 수 있다. 이해상충 때문에 다른 비돌림 영역의 결함이 생길 수 있고, 만일 그렇다면 이 문항은 항상 개별적으로 평가해야한다.

2. ROBIS: 원본 매뉴얼 번역

이 도구는 3단계로 완성된다. ① 관련성 평가(해당하는 경우) ② 문헌고찰 과정에서의 우려사항 확인 ③ 문헌고찰에서 비뚤림위험 판단. 신호 질문은 문헌고찰에 잠재적 비뚤림에 대한 특정한 우려를 평가하는데 도움이 되도록 포함되어 있다. 이러한 신호 질문의 등급(rating)은 평가자가 전반적인 비뚤림위험을 판단하는데 도움이 된다.

비뚤림위험의 정의

ROBIS는 문헌고찰에서 비뚤림위험과(해당하는 경우) 연구질문에 대한 문헌고찰의 관련성을 모두 평가한다. 구체적으로 ① 요약치 추정과 문헌고찰 결론의 비뚤림위험을 최소화한 문헌고찰 방법의 정도와 ② 문헌고찰이 다루는 연구질문이 사용자가 다루고 있는 연구질문(예: overview 작성자 또는 진료지침 개발자)과 일치하는 정도를 다룬다. 비뚤림은 문헌고찰의 설계, 수행 또는 분석에 있어 체계적인 결함이나 한계가 결과를 왜곡할 경우 발생한다. 문헌고찰 질문이 overview/지침의 질문과 일치하지 않는 경우 문헌고찰의 근거는 제한된 관련성을 가질 수 있다.

목표 독자(target audience)

ROBIS는 세 부류의 사용자 그룹을 염두에 두고 개발되었다:

- 체계적 문헌고찰의 overview 저자
- 진료지침 개발자
- 출판된 문헌고찰에서 비뚤림위험을 평가하거나 프로토콜 단계에서 문헌고찰을 계획할 때 비뚤림위험을 최소화하고자 하는 검토자

체계적 문헌고찰에서 비뚤림위험을 평가하고자 하는 사람에게도 도움이 될 수 있지만, 이러한 목적을 위해 설계된 것은 아니다. 다른 ROBIS의 잠재적인 사용자에는 의사 결정을 지원하는 조직(예: NICE, IQWiG), EBM에 관심이 있는 임상, 학술지 편집인 및 원고 검토자가 있다.

2.1. 1단계: 관련성의 평가(해당하는 경우)

평가자가 먼저 자신이 답변하려고 하는 질문(예: overview 또는 진료지침)을 보고한다- 우리는 이를 '목표질문(target question)'이라고 부른다. 효능 혹은 효과 문헌고찰의 경우 PICO(연구대상자, 중재, 비교중재, 중재결과) 측면에서 이를 정의해야 한다. 다양한 유형의 질문(예: 진단 검사, 예후 요인, 원인 또는 예측 모델)에 대한 문헌고찰의 경우 적절한 대안 범주가 제공된다(Box 1-1 참조). 평가자는 ROBIS를 사용하여 평가할 체계적 문헌고찰에 대해 PICO 또는 이에 준하는 항목을 작성한 후 두 질문(목표질문과 체계적 문헌고찰 질문)이 일치하는지를 묻는다. 하나 이상의 범주(PICO 또는 동등한 범주)가 일치하지 않을 경우 '아니오'로 평가되어야 한다. 범주 간에 부분 일치하는 경우, '부분적'으로 평가되어야 한다. 예를 들어, 목표질문이 성인과 관련이 있지만 체계적 문헌고찰은 60세 이상의 연구대상자로 제한되는 경우가 그런 경우이다. 문헌고찰이 단독적으로 평가되고 있고 목표질문이 없다면, ROBIS의 이 단계는 생략할 수 있다.

Box 1-1. 체계적 문헌고찰의 여러 종류에 대한 목표질문과 PICO 형식의 예시

문헌고찰 종류	PICO 형식	예시
중재	환자/인구집단: 중재: 비교중재: 중재결과:	만성 C형 간염에 감염된 성인 pegylated 인터페론을 이용한 트리플 항바이러스 요법 이중 항바이러스 요법 지속적인 바이러스 반응(sustained virologic response)
원인	환자/인구집단: 노출과 비교노출: 중재결과 (outcome):	성인 체질량지수(BMI) 대장암
진단	환자군: 중재 검사: 참고 표준 검사: 대상 질환:	직장암 의심 증상을 가진 성인 내시경 초음파 외과적 조직학(surgical histology) 직장암 초기(T0)
예후	환자군: 예상되는 중재결과: 모델의 사용 목적: 의도한 시점:	태아 성장 제한이 있거나 없는 임신부로서, 막(membrane)의 조기파열 증거가 없고, 선천성 또는 구조적 이상의 증거가 없음. 유해한 임신 결과 (저체중 또는 과체중, 신생아 사망, 주산기 사망률) 예측 임신 후기(37주 이상)

타당성 평가 예시

진단 문헌고찰에서:

범주	목표질문	평가 대상 문헌고찰
환자:	직장암 의심 증상을 가진 성인	불확실
중재 검사:	내시경 초음파	내시경 초음파
참고 표준 검사:	외과적 조직학	수술
대상 질환:	직장암 초기(T0)	직장암 초기(T0)

타당성 평가	이유
<p>문헌고찰에 의해 다루어진 질문이 답을 알고자 하는 질문과 일치합니까?(예: overview 혹은 진료지침에서)</p> <p>불확실 (unclear)</p>	<p>이 질문은 중재 검사, 참고 표준 검사 및 대상 질환의 점수에 일치하지만 인구집단이 문헌고찰에서 정의되지 않았기 때문에 문헌고찰에서 다른 질문이 목표질문과 일치하는지 여부는 불확실하다.</p>

2.2. 2단계: 문헌고찰 과정의 우려사항 파악

2단계는 비플림이 체계적 문헌고찰의 결과에 영향을 끼칠 수 있는 영역을 찾아내는 것을 목표로 한다. 여기에는 연구 적격 기준, 연구 검색 및 선택, 자료 수집 및 연구 평가, 합성 및 결론 등 4개 영역의 평가로 이루어져 있다. ROBIS의 이 단계는 최종 단계에서 전반적인 비플림의 위험을 판단하는데 도움이 되는 잠재적 관심 영역을 찾아낸다. 각 영역은 세 부분으로 구성된다: 판단을 뒷받침하는데 사용되는 정보, 신호 질문, 그리고 비플림위험에 대한 우려의 판단. 각 영역들은 순차적으로 고려되어야 하며 독립된 단위로 평가해서는 안된다. 예를 들어 영역 2(연구의 검색 및 선택)를 평가할 때, 평가자가 영역 1에 명시된 연구 문제와 관련하여 검색 전략의 용어와 구조를 고려해야 한다.

신호 질문은 ‘그렇다’, ‘아마도 그렇다’, ‘아마도 아니다’, ‘아니다’, ‘정보 없음’으로 답하며, ‘그렇다’는 낮은 우려를 나타낸다. 이후 해당 영역의 비플림에 대한 우려 수준은 ‘낮음’, ‘높음’, 또는 ‘불확실’로 판단한다. 이 등급은 ‘그렇다’, ‘아니다’ 또는 ‘불확실’의 등급으로 표현하지 않는 이유는 ‘그렇다’가 모든 신호 질문 문항의 표현과는 대조적으로 나쁜 것을 의미하기 때문이다. 또한 이 등급은 영역 3의 비플림위험 등급과 일치한다. 영역의 모든 신호

질문에 대한 답변이 ‘그렇다’ 또는 ‘아마도 그렇다’인 경우, 우려 수준은 낮은 것으로 판단할 수 있다. 신호 질문이 ‘아니다’ 또는 ‘아마도 아니다’인 경우, 비독립에 대한 우려 가능성이 존재한다. ‘정보 없음’ 범주는 판단을 하기에 보고된 자료가 불충분한 경우에만 사용해야 한다. 판단에 도달하는 데 사용되는 정보(판단에 대한 뒷받침)를 기록함으로써 투명하게 등급을 매기고, 필요한 경우, 문헌고찰 저자들과 협의 후 독립적으로 평가를 완료한다. ROBIS 사용자가 평가를 완료하기 위해서는 주제 내용과 방법론에 대한 전문 지식이 모두 필요할 수 있다.

2.2.1. 영역 1: 연구 적격 기준

(1) 배경

첫 번째 영역은 일차 연구 적격 기준이 사전에 지정되고 명확하며 문헌고찰 질문에 적절하지 여부를 평가하는 것을 목표로 한다. 체계적 문헌고찰은 명확하게 초점을 맞춘 질문이나 목표에서 시작해야 한다. 이는 일차 연구가 문헌고찰 대상에 포함될 수 있는지 여부를 결정하기 위해 사용되는 기준의 사전 설명서에 반영되어야 한다. 이 사전 설명서는 어떤 연구를 포함시킬지에 대한 결정이, 연구 자체의 특성과 결론에 대한 기존의 지식으로 이루어지는 것이 아니라, 일관성있게 이루어지도록 하는 것을 목적으로 한다. 문헌고찰의 수행과 보고를 사전에 예약하는 프로토콜이나 등록 문서를 이용할 수 있는 경우에만 적격 기준을 적절하게 사전에 정했는지(그리고 문헌고찰에서 준수했는지) 여부를 평가할 수 있다. 이러한 문서를 사용할 수 없는 경우 평가자는 이 영역에 대한 판단을 문헌고찰의 결론에서 보고한 것에 기초할 필요가 있을 텐데, 이 경우 이러한 기준이 실제로 사전에 규정되었는지, 저자들이 문헌고찰 과정 중에 무엇을 했는지 또는 문헌고찰 과정에서 결정되거나 수정되었는지 여부를 알 수가 없다.

적격 기준은 문헌고찰 질문의 측면과 문헌고찰이 고려하는 연구설계를 포함한 연구 유형에 대한 추가 세부사항을 합친 것이다. 이상적인 경우, 적격 연구의 특성은 관련된 인구집단/연구대상자, 평가된 중재/노출 및 비교중재에 따라 표현될 것이다. 연구 특성에 대해 제공된 정보는 저자가 연구 적격에 대해 문헌고찰 과정 중에 판단을 내릴 수 있는 모호성을 피하도록 가능한 한 구체적이어야 한다. 저자들은 종종 적격 기준에 제한을 가하여 적합하다고 판단할 수도 있고 그렇지 않을 수도 있다. 예를 들어 특정한 관심 중재결과를 보고하는 것, 특정(연구의) 질 기준을 충족하는 것, 또는 특정 날짜 범위 내에서 수행되는 것과 같은 경우처럼 특정한 연구 특성에 따라 포함된 연구를 제한할 수 있다. 또한 완전한 형태로 출

판되거나 영어로 이용 가능한 정보 등 그들이 포함할 정보의 출처에 제한을 가할 수도 있다. 포함된 연구에 제약을 가하는 경우, 사전에 이를 명시하고 정당화해야 한다. 어떤 상황에서 제한사항이 포함될 수 있는 연구의 선택에 비뚤림을 초래할 수 있다.

표 1-1에는 영역 1에 포함된 신호 질문이 각 질문에 대한 등급부여(rating) 지침과 함께 요약되어 있다. 표 1-2에는 영역 1에 대한 영역 레벨 등급의 개요를 제시하였다. 평가자가 이 영역에 대한 높거나 낮은 우려를 가지고 있는 경우 문헌고찰의 등급 예시를 표 3과 4에 제시하였다.

표 1-1. 영역 1에 대한 신호 질문들과 평가지침

신호 질문	등급부여 지침
1.1 문헌고찰이 사전에 정의된 연구 목적과 적격 기준을 준수하였는가?	<p>체계적 문헌고찰은 연구가 포함되기에 적절한지 여부를 결정하기 위해 사용되는 기준에 반영되는 명확하게 포커스 된 질문이나 목표를 가지고 시작해야 한다. 문헌고찰 프로토콜에서 미리 정해야 하는 세부사항은 문헌고찰 유형에 따라 다르지만 일반적으로 적격 기준에는 연구설계, 연구대상자 및 중재/노출의 유형을 포함해야 한다. 모든 제외 기준도 미리 정해야 한다.</p> <p>이 정보를 제공하는 프로토콜이 있는 경우 이 질문에 대한 대답은 '그렇다'일 것이다. 프로토콜은 없지만 사전에 정의된 목표와 세부적인 적격 기준에 대한 정보가 제공되고, 이러한 목표들이 사전에 정해져 문헌고찰 중에 준수되었다고 믿을 만한 충분한 이유가 있는 경우, 평가자는 이 질문에 대해 '아마도 그렇다'라고 대답하는 것을 고려할 수 있다.</p> <p>적격 기준 또는 중재결과에 대한 사후 변경은 문헌고찰 목표에 충실해야 하며 적절히 정당화되고 문서화되어야 한다. 사전에 출판된 프로토콜이 없는데, 사전에 정의된 목표와 적격 기준에 대한 정보가 오직 문헌고찰 출판물에서만 사후적으로 이용 가능하여, 이러한 세부 사항이 사전에 정해져 시작부터 준수되었다고 믿을 만한 이유가 없으면 이 질문에 '아마도 아니다'라고 답해야 한다. 이러한 세부 사항 중 일부 또는 전부가 누락된 경우 이 질문에 '아니다'라고 대답해야 한다.</p>
1.2 적격 기준이 문헌고찰의 질문에 부합하는가?	<p>적격 기준은 문헌고찰 질문에서 비롯되어야 하며, 포함된 연구가 질문에 적합한지에 대한 판단이 가능하도록 세부사항을 제공해야 한다. 필요한 정보는 주제별로 다를 것이다. 예를 들어, 적합성을 판단하기 위해 평가자는 연명 범위와 연구대상자의 진단, 연구가 수행되었던 세팅, 약물의 용량 또는 노출 빈도의 측면에서 인구집단에 대한 명확한 설명이 필요할 것이다. 이 질문에 답하기 위해 평가자는 일부 내용에 대한 지식이 요구될 가능성이 높다.</p>
1.3 적격 기준이 모호하지 않은가?	<p>적격 연구의 특성에 대한 구체적인 정보가 제공되어야 하며, 가능한 한 연구 유형, 인구집단, 중재, 비교중재 및 중재결과에 대한 모호성을 피해야 한다. 기준은 정해진 기준을 사용하여 문헌고찰을 재현할 수 있을 정도로 충분히 상세해야 한다.</p>

	<p>체계적 문헌고찰에서 여러 가지 중요한 세부 사항이 적격 기준에서 일반적으로 누락되고 있다. 예를 들어, 연구대상자의 진단에 대한 세부 사항을 들 수 있는데, 진단은 여러 가지 다른 방법을 통해 이루어질 수 있으며, 그 중 일부는 다른 것들 보다 더 유효하거나 정확할 수 있다.</p> <p>문헌고찰 저자는 문헌고찰 과정 중에 잠재적인 비독립을 피하기 위해 문헌고찰 질문에 적합한 진단 방법을 미리 결정해야 한다.</p> <p>마찬가지로 약물 용량, 투여 빈도, 동시 치료 등과 같은 특성을 포함하여 증재/노출 및 비교증재에 대한 구체적인 세부사항이 제시되어야 한다. 평가자는 이 질문에 답하기 위해 일부 내용에 대한 지식을 요구할 가능성이 높지만, 명시된 적격 기준에 대한 구체적인 의문(query)이 있는 경우 일반적으로 '아니다' 또는 '아마도 아니다'라고 판단할 수 있다.</p>
<p>1.4 연구 특성에 따른 적격 기준의 제한이 적절한가?</p>	<p>연구 특성에 따라 적용되는 모든 제한사항은 명확하게 기술되어야 하며 확실한 근거를 제시해야 한다. 이러한 세부사항을 통해 평가자는 그러한 제약이 적절한지 판단할 수 있다. 예를 들어 연구설계, 연구 출판시점, 연구의 크기, 연구 질의 일부 측면 및 이용 가능한 증재결과 측정 등이 있을 수 있다.</p> <p>이 질문은 적격 기준이 문헌고찰 질문에 적합한지 여부를 다루는 위의 질문과는 다르다. 충분한 정보를 이용할 수 있고, 평가자가 제한사항이 적절하다고 합리적으로 만족하는 경우, 이 질문에 '그렇다' 또는 '아마도 그렇다'라고 대답할 수 있다. 연구 특성을 둘러싼 제약이 정당화되지 않고 이러한 제약이 적절한지 판단하기에 충분한 정보가 없는 경우, 이 질문에 '아마도 아니다' 또는 '아니다'로 대답해야 한다. 적격 기준이 충분히 상세하고, 연구 특성에 대한 제한이 명시적으로 보고되지 않은 경우에는 부과된 것이 없다고 가정할 수 있으며, 질문에 '그렇다'라고 답해야 한다.</p>
<p>1.5 정보의 출처에 따른 적격 기준의 제한이 적절한가?</p>	<p>정보의 출처에 기초하여 적용되는 모든 제한사항을 명확하게 기술하고 확실한 근거를 제시해야 한다. 이러한 세부사항을 통해 평가자는 그러한 제약이 적절한지 판단할 수 있다.</p> <p>예를 들어, 출판 상태 또는 형식(format), 언어 및 자료 가용성이 있을 수 있다. 이 질문은 검색 제한에 관한 영역 2의 질문과는 다르다. 적격 기준이 충분히 상세하지만 정보 출처에 대한 제한이 명시적이지 않은 경우에는 부과된 것이 없다고 가정하고 질문에 '그렇다'라고 대답해야 한다.</p>

표 1-2. 연구 적격 기준에 대한 우려 정도

<p>낮은 우려</p>	<p>문헌고찰 질문과 목표를 명확히 명시하고, 문헌고찰 과정에서 적절하고 상세한 적격 기준을 사전에 구체화하고 정당화하기 위해 상당한 노력을 기울였다.</p>
<p>높은 우려</p>	<p>문헌고찰 질문에 대답하는데 중요하고 관련이 있는 연구들이 사전에 지정된 목표와 적격 기준이 부족하거나 부적절한 제한이 부과되었거나, 문헌고찰 질문을 다루는데 적절하지 않은 연구가 포함되었기 때문에 문헌고찰에서 제외되었을 가능성이 높다.</p>
<p>불확실한 우려</p>	<p>비독립위험에 대해 판단하기에 불충분한 정보가 보고되었다.</p>

(2) 등급부여의 예

a. 연구 적격 기준에서 높은 우려로 판단된 문헌고찰¹

본문 기술 내용:

Abstract - Aim

“To evaluate the accuracy of EUS in T staging of early rectal cancers.”

Study Selection Criteria

“Only EUS studies confirmed by surgical histology were selected. EUS criteria used for T0 was tumour confined to the mucosa. From this pool, only studies from which a 2×2 table could be constructed for true-positive, false-positive, false negative and true-negative values were included.”

표 1-3. 연구 적격 기준에 대한 평가 예시(‘높음’으로 평가)

신호질문	등급부여	이유
1.1 문헌고찰이 사전에 정의된 연구 목적과 적격 기준을 준수하였는가?	정보 없음	사전에 정의된 목표와 적격 기준이 존재한다는 근거가 없었다.
1.2 적격 기준이 문헌고찰의 질문에 부합하는가?	아마도 아니다	보고한 적격 기준은 적절하나 연구대상 인구집단 관련 일부 세부 사항이 부족했다. 따라서 적격 기준이 적절하다는 판단을 하는 것이 불가능하다.
1.3 적격 기준이 모호하지 않은가?	아마도 아니다	적격 기준에 대한 세부 사항이 불충분했다. 특히 어떤 연구 인구집단이나 어떤 연구설계가 적합한지에 대한 세부사항은 없었다.
1.4 연구 특성에 따른 적격 기준의 제한이 적절한가?	정보 없음	연구 특성에 따른 제한에 대한 세부사항은 제공되지 않았다. 전반적으로 적격 기준에 대한 세부 사항이 불충분하다.
1.5 정보의 출처에 따른 적격 기준의 제한이 적절한가?	아마도 그렇다	문헌고찰은 2×2표를 구성할 수 있는 연구로 제한되었다. 이에 대한 정당성이 제공되지 않았지만, 이는 진단 정확도 (diagnostic test accuracy, DTA) 문헌고찰에서 흔히 볼 수 있으므로 합리적일 수 있다.
연구 적격 기준의 규정에 대한 우려	높음	적절한 연구가 문헌고찰에 포함되었는지 여부를 판단할 수 있는 연구 적격 기준에 대한 세부 정보가 부족했으며, 특히 적격 참여자에 대한 세부 정보가 부족했다. 기준이 사전 지정되었는지, 사후 적용되었는지도 불확실했다. 따라서 적격 기준의 규정에 대한 우려가 높다.

1) Puli SR, Bechtold ML, Reddy JB, Choudhary A, Antillon MR. Can endoscopic ultrasound predict early rectal cancers that can be resected endoscopically? A meta-analysis and systematic review. Dig Dis Sci 2010;55(5):1221-9.

b. 연구 적격 기준에서 관련하여 낮은 우려로 판단된 문헌고찰²

본문 기술 내용:

Objective

“To evaluate the association and predictive value of ultrasound measurements of amniotic fluid volume for adverse pregnancy outcome.”

Methods

Study selection “1. Population. Pregnant women, with or without fetal growth restriction, no evidence of premature rupture of membranes, no evidence of congenital or structural abnormalities.

2. Index test. Any measure of amniotic fluid reported by the authors including AFI, amniotic fluid volume and maximum deepest pocket. Any threshold used to define low or high amniotic fluid as reported by the authors of the included studies was accepted.

3. Outcome. Any reference standard looking at compromise of fetal or neonatal wellbeing; including: abnormal cord pH at birth, Apgar scores, perinatal death and composite outcomes such as adverse perinatal outcome. Any reference standard for fetal growth restriction or small for gestational age: Birthweight <10th, <5th, <3 rd centile, absolute birthweight thresholds, ponderal index.

4. Study design. Observational studies in which the results of the test of interest are compared with the outcome findings as confirmed by a reference standard, allowing generation of a 2x2 table to compute indices of association and test accuracy for each available threshold. Case series of ten or fewer and case-control studies determined by outcome were excluded.”

Data Extraction “All studies had to state that they excluded rupture of membranes and congenital/structural anomalies due to the association of renal/urinary tract anomalies and karyotypic anomalies with abnormalities of liquor volume.”

2) Morris RK, Meller CH, Tamblin J, Malin GM, Riley RD, Kilby MD, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and metaanalysis. BJOG: An International Journal of Obstetrics & Gynaecology 2014;121(6):686-99.

표 1-4. 연구 적격 기준에 대한 평가 예시('낮음'으로 평가)

신호질문	등급부여	이유
1.1 문헌고찰이 사전에 정의된 연구 목적과 적격 기준을 준수하였는가?	그렇다	저자들은 논문의 초록과 연구배경에서 양수 부피의 초음파 측정과 부정적 임신결과 사이의 연관성과 예측값을 평가하는 것이 목적이라고 명확히 규정했다. 별도의 프로토콜에서는 “태아/신생아 위험과 태아 성장 제한을 예측하기 위한 양수 부피 측정의 정확성은 무엇인가?”라는 구체적인 문헌고찰 질문을 제공했다. 저자들은 인구집단, 증재 검사, 증재결과, 연구설계의 세부사항을 적격 기준으로 제공했다.
1.2 적격 기준이 문헌고찰의 질문에 부합하는가?	그렇다	문헌고찰 질문은 태아/신생아 위험과 태아 성장 제한을 예측하기 위해 양수 부피의 정확도를 조사하는 연구를 파악할 필요가 있음을 제시했다. 논문에서 제시한 연구 적격 기준의 세부 내용은 문헌고찰 질문에 적합한 것으로 보였다.
1.3 적격 기준이 모호하지 않은가?	아마도 그렇다	연구설계의 종류는 분명히 명시되어 있었다. 관찰 연구가 포함되었다. 저자들은(10례 이하의) 사례군 연구와 결과에 따라 결정된 환자 대조군 연구를 제외했다. 관심 인구집단의 명확한 세부 사항(태아 성장 제한이 있거나 없는 임신부, 조기양막 파열의 증거 없음, 선천적 또는 구조적 이상 징후 없음), 증재 검사(AFI, 양수 부피 및 최심부 포켓 등 사용된 임계값에 관계없이 보고된 양수 측정치), 증재결과(태아 또는 신생아 건강의 위험을 살피는 모든 참고 표준의 몇가지 구체적인 예가 열거되었다. 태아 성장 제한 또는 작은 임신 연령에 대한 모든 참조 표준의 몇가지 구체적인 예가 열거됨)이 제공되었다.
1.4 연구 특성에 따른 적격 기준의 제한이 적절한가?	아마도 그렇다	연구설계의 유형과 연구 규모 문제에 기초한 제한사항이 명확하게 설명되었다. 추가 제한사항은 자료 추출 섹션에서 제공되었다. 비록 정당성이 제공되지는 않았지만, 그 제한은 적절해 보였다.
1.5 정보의 출처에 따른 적격 기준의 제한이 적절한가?	아마도 그렇다	언어 제한은 적용되지 않았다. 비록 정당성이 제공되지는 않았지만 자료의 가용성에 기초했을 일부 제한이 있었다. 정보의 출처에 대한 다른 제한은 기술되지 않았다. 그러므로 이 질문은 “아마도 그렇다”라고 대답했다.
연구 적격 기준의 규정에 대한 우려	낮음	모든 신호 질문에 대해 “그렇다” 또는 “아마도 그렇다”로 대답되었기 때문에 적격 기준의 규정에 대한 잠재적 우려는 확인되지 않았다. 문헌고찰 질문과 목표를 명확히 명시하고, 적절하고 상세한 적격 기준을 사전에 명시하고 정당화하기 위해 상당한 노력을 기울였다.

2.2.2. 영역 2: 연구의 검색 및 선택

(1) 배경

이 영역은 포함 기준을 충족했을 일차 연구가 문헌고찰에 포함되지 않았는지 여부를 평가하는 것을 목적으로 한다. 가능한 한 많은 적격 연구를 파악하기 위한 민감한 검색은 모든 체계적 문헌고찰의 핵심 요소이다.

이상적으로 검색은 훈련된 정보 전문가의 도움하에 또는 함께 수행된다. 검색 결과에 기반한 연구의 편향되지 않은 선택은 검색으로 파악된 모든 관련 연구가 문헌고찰에 포함되고 부적합한 연구가 포함되지 않도록 하는 데 도움이 된다. 검색에는 출판되었거나 출판되지 않은 연구결과를 파악할 수 있는 적절한 데이터베이스와(논문, 컨퍼런스 및 임상시험기록 등을 색인한) 전자정보원이 포함되어야 하며, 적합한 연구 보고서를 파악하기 위한 데이터베이스 검색(예: 기존 고찰의 참조 확인, 인용 검색, 수기 검색) 및 사용 방법이 포함되어야 한다.

적절하고 민감한 검색 전략. 검색 전략에는 프리텍스트 용어(예: 제목과 초록)와 관련 연구를 파악할 수 있는 적합한 주제 색인(예: MeSH 또는 Emtree)이 포함되어야 한다. 또한 검색할 주제와 자원에 따라 CAS 레지스트리 번호와 같은 추가 필드, 의약품 또는 기기의 상품명 등을 검색하는 것도 도움이 될 수 있다.

전략은 일반적으로 인구집단, 중재/중재검사/노출 중 하나 이상의 조합이다. RCT의 문헌고찰과 같은 특정 유형의 문헌고찰의 경우, 검색된 데이터베이스의 일부 또는 전체에 방법론적 또는 연구설계 필터를 포함하는 것도 적절할 수 있다.

그러나, 진단 정확도 연구에 대한 것과 같은 다른 유형의 문헌고찰의 경우, 필터의 사용은 관련 연구의 누락가능성이 높아 피해야 한다. 문헌고찰과 관련된 검색 관행 및 내용 전문지식과 관련된 방법론적 지식이 없으면 검색 전략의 민감도를 평가하기 어려울 수 있다.

일반적으로, 평가자는 연구질문을 포착하기 위해 사용된 개념이 기술될 수 있는 가능한 모든 방법을 다루기 위해 적절한 범위의 용어가 포함되는지 여부를 고려해야 한다. 예를 들어 유방암에 대한 문헌고찰을 위해 '유방암'이라는 용어를 키워드로 검색하는 것만으로는 충분하지 않다. OvidSP 인터페이스를 통해 검색된 Medline 데이터베이스에 대해 이 인구집단에 대해 예상되는 전략의 예가 아래에 나와 있다.

- ① *exp breast neoplasms/ or Carcinoma, Ductal, Breast/ (214761)*
- ② *exp breast/ (31443)*
- ③ *exp neoplasms/ (2533252)*
- ④ *2 and 3 (17379)*
- ⑤ *(breast\$ adj5 (neoplasm\$ or cancer\$ or tumor\$ or metastas\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intraductal\$ or lobular or medullary)).ti,ab,ot,hw. (256928)*
- ⑥ *(mammar\$ adj5 (neoplas\$ or cancer\$ or tumor\$ or metastas\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intra-ductal\$ or lobular or medullary)).ti,ab,ot,hw. (34861)*
- ⑦ *1 or 4 or 5 or 6 (277790)*

표 1-5는 영역 2에 포함된 신호질문을 각 질문에 대한 등급 지침과 함께 요약한다. 표 1-6은 영역 2의 영역 수준 등급에 대한 개요를 제공한다. 이 영역에 대한 평가자의 높은 우려와 낮은 우려의 문헌고찰 예시는 표 2-7과 표 2-8에 제시되어 있다.

표 1-5. 영역 2에 대한 신호질문과 평가지침

신호질문	등급부여 지침
2.1 출판/미출판 보고서에 대한 적절한 범위의 데이터베이스/전자자료원인 검색에 포함되었는가?	평가자는 적절한 범위의 데이터베이스 조합을 판단할 필요가 있다. 이것은 문헌고찰 주제에 따라 달라질 것이다. 최소한 MEDLINE과 EMBASE 검색을 수행해야 한다. 학회발표 자료의 검색도 연구 레지스트리 검색과 함께 검토해야 한다. 적절한 데이터베이스 범위에 대한 지침은 코크란 핸드북 같은 체계적 문헌고찰 지침 또는 CRD(Center for Review and Dissemination) 웹 사이트에서 확인할 수 있다. (http://www.york.ac.uk/inst/crd/finding_studies_systematic_reviews.htm).
2.2 관련 보고서를 찾기 위해 데이터베이스 검색에 추가적인 방법이 사용되었는가?	인용한 문헌 검색, 전문가 접촉, 참고문헌 확인, 수기검색 등의 추가 방법이 수행되어야 한다.
2.3 검색 전략의 용어와 구조가 가능한 한 많은 적격 연구를 검색할 가능성을 높였는가?	검색을 복기하기에 충분하도록 세부적으로 사용된 모든 검색어를 보여주는 전체 검색 전략이 제시되어야 이 질문에 대한 완전한 파악이 가능하다. 검색어 조합 방법에 대한 표시가 없이 검색어 목록만 나열하는 식의 제한된 세부 사항만 제공되는 경우, 평가자는 “아마도 그렇다” 또는 “아마도 아니다”의 판단을 내릴 수 있다.

	<p>평가자는 검색 전략에 주제에 대한 적절한 범위의 용어가 포함되었는지 여부, MeSH(Medline용 Medical Subject Headings)와 제목 및 초록에 포함된 단어의 조합이 사용되었는지 여부, 적용된 필터가 적절했는지의 여부를 고려해야 한다.</p> <p>예를 들어, DTA 문헌고찰의 경우 필터의 사용은 관련 연구를 누락한 것으로 나타나므로 해당 필터를 포함하는 전략에 대해 이 질문은 “아니다”로 대답해야 한다. 검색 전략의 비판적 평가에 대한 지침은 PRESS 근거 기반 점검표에서 찾을 수 있다.</p> <p>(http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402).</p>
<p>2.4 날짜, 출판 형식 또는 언어에 따른 제한이 적절하였는가?</p>	<p>만일 검색 전략에 제한이 적용되지 않은 경우 이 질문에 “그렇다”로 대답해야 한다. 이는 선정기준에 대한 제한에 관한 영역 1(1.5)의 질문과는 다르다. 이 항목을 완전히 판단할 수 있도록 이 질문의 세 가지 구성 요소(즉, 날짜, 출판 형식, 언어) 모두에 대한 정보가 필요하다.</p> <p>언어에 기반한 논문 제한(예: 영어 문헌으로 제한)이나 출판 형식(예: 전문(full text)으로 출판된 연구로 제한)은 대부분 적절하지 않기 때문에, 만약 그러한 제한이 적용되었다면, 이 질문은 대개 “아니다”로 대답해야 한다. 날짜에 대한 제한은 적절할 수 있지만 이 질문에 “그렇다”로 대답하기 위해 명확하게 설명된 근거에 의해야 한다. 예를 들어, 특정 날짜 이전에는 해당 약물이나 검사가 존재하지 않는 경우, 약물이나 시험을 처음 사용할 수 있게 된 날짜로부터 검색을 시작하는 것이 합리적이다.</p>
<p>2.5 연구 선택의 오류를 최소화하기 위한 노력이 이루어졌는가?</p>	<p>선택을 위해 제목과 초록으로 선별하는 과정과 전문(full text)을 평가하는 과정 모두 이 질문에서 다룬다.</p> <p>두 가지 모두에 대한 정보는 이 항목을 완전히 판단할 수 있어야 한다. 평가에서 판단이 “그렇다”가 되기 위해서는 제목과 초록을 적어도 두 명의 검토자가 독립적으로 심사해야 하며, 전문 평가에도 적어도 두 명의 검토자가 참여해야 한다(각자 독립적으로 하거나 또는 한 명이 평가를 수행하고 두 번째 검토자가 결정을 확인해야 한다).</p>

표 1-6. 연구의 검색 및 선택에 대한 우려 정도

<p>낮은 우려</p>	<p>영역 1에서 평가한 문헌고찰 질문과 적격 기준을 고려할 때, 민감하고 적절한 검색 전략을 사용하여 다양한 검색 방법을 통해 가능한 한 많은 관련 연구를 파악하고 포함 연구를 선택할 때 비둘림과 오류를 최소화하기 위한 조치를 취했다.</p>
<p>높은 우려</p>	<p>일부 적격 연구가 문헌고찰에 제외되었을 가능성이 높다.</p>
<p>불확실한 우려</p>	<p>비둘림위험에 대한 판단을 할 수 있는 정보가 충분하지 않다.</p>

(2) 등급부여의 예

a. 연구의 검색 및 선택 방법에서 높은 우려로 판단된 문헌고찰³⁾

본문 기술 내용:

“A research librarian searched Ovid MEDLINE from 1947 to August 2012, the Cochrane Library Database (through the first quarter of 2012), Embase (1976 to August 2012), Scopus (1960 to August 2012), PsychINFO (1806 to August 2012), clinical trials registries, and grants databases. At least 2 reviewers independently evaluated studies for inclusion. Non-English language articles were excluded. Included studies published as conference abstracts were only used in sensitivity analyses.”

표 1-7. 연구 검색 및 선택에 대한 평가 예시(‘높음’으로 평가)

신호질문	등급부여	이유
2.1 출판/미출판 보고서에 대한 적절한 범위의 데이터베이스/전자자료원이 검색에 포함되었는가?	그렇다	MEDLINE, EMBASE, Scopus, PsycINfo 및 Cochrane 라이브러리를 검색하였다(Cochrane 라이브러리에 어떤 검색원이 있는지는 불확실). 이것은 적절한 범위라고 판단되었다. 임상 시험 레지스트리 및 연구비 데이터베이스도 검색하였다.
2.2 관련 보고서를 찾기 위해 데이터베이스 검색에 추가적인 방법이 사용되었는가?	아니다	아니다
2.3 검색 전략의 용어와 구조가 가능한 한 많은 적격 연구를 검색할 가능성을 높였는가?	정보 없음	전체 검색 전략을 보고하지 않았고 검색어에 대한 세부 정보가 없었다. 따라서 이 질문에 대한 평가의 기초가 되는 정보가 없었다.
2.4 날짜, 출판 형식 또는 언어에 따른 제한이 적절하였는가?	아니다	영어로만 제한한 문헌고찰로 출판비틀림의 가능성이 있다.
2.5 연구 선택의 오류를 최소화하기 위한 노력이 이루어졌는가?	아 마 도 그렇다	포함여부 평가는 최소 2명의 검토자가 독립적으로 실시한 것으로 알려졌다. 그러나 검색결과와 선별과 전문평가에 모두 이를 적용했다는 것은 명확하지 않았다. 따라서 이 항목은 “아마도 그렇다”로 평가되었다.
연구 파악과 선택방법에 대한 우려	높음	문헌고찰을 영어로만 제한한 것은 관련 연구가 이번 문헌고찰에 포함되지 않았을 위험이 크다는 것을 의미한다. 검색 전략이 목적에 부합하는지 판단할 수 없다.

3) Chou R, Hartung D, Rahman B, Wasson N, Cottrell EB, Fu R. Comparative effectiveness of antiviral treatment for hepatitis C virus infection in adults: a systematic review. Ann Intern Med 2013;158(2):114-23.

b. 연구의 파악 및 선택 방법에서 낮은 우려로 판단된 문헌고찰⁴**본문 기술 내용:**

“The following sources were searched from inception to October 2011: MEDLINE; EMBASE; Cumulative Index To Nursing And Allied Health Literature(CINAHL); The Cochrane Central Register of Systematic Reviews; The Cochrane Central Register of Controlled Trials; DARE; MEDION; SIGLE; Index of Scientific and Technical Proceedings, Web of Science and ClinicalTrials.gov database. The search consisted of keywords and MeSH terms relating to the tests under investigation combined with MeSH terms of ‘Prenatal Diagnosis’, ‘Ultrasonography’, ‘Amniotic Fluid’ and ‘Pregnancy Outcome’. The full search strategy is shown in the Appendix S2. The reference lists of all included primary and review articles were examined to identify cited articles not captured by electronic searches. No language restrictions were applied. The database was scrutinised by two reviewers(RKM, CHM) and full articles likely to meet the selection criteria were obtained. Translations were obtained for non-English articles. Three reviewers made the final inclusion/exclusion decisions according to adherence to the following criteria.”

4) Morris RK, Meller CH, Tamblyn J, Malin GM, Riley RD, Kilby MD, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and metaanalysis. BJOG: An International Journal of Obstetrics & Gynaecology 2014;121(6):686-99.

표 1-8. 연구 검색 및 선택 평가 예시('낮음'으로 평가)

신호질문	등급부여	이유
2.1 출판/미출판 보고서에 대한 적절한 범위의 데이터베이스/전자자료원이 검색에 포함되었는가?	그렇다	MEDLINE, EMBASE, CINAHL, SIGLE, Index of Scientific and Technical Proceedings, ClinicalTrials.gov 데이터베이스, Cochrane 라이브러리, DARE, MEDION을 검색하였다. Web of Science 또한 검색하였으나 이 플랫폼의 어떤 데이터베이스를 검색하였는지는 명확하지 않다. 적절한 범위의 자료원을 출판/미출판 보고서를 검색하려는 시도도 포함되었다. 그러나 Cochrane Central Register of Systematic Reviews는 기록오류일 가능성이 높으며 저자들은 CDSR을 검색했다.
2.2 관련 보고서를 찾기 위해 데이터베이스 검색에 추가적인 방법이 사용되었는가?	그렇다	포함된 모든 일차 연구와 고찰의 참고문헌 리스트를 조사하고 전자검색으로 파악되지 않은 인용문헌을 파악하였다.
2.3 검색 전략의 용어와 구조가 가능한 한 많은 적격 연구를 검색할 가능성을 높였는가?	그렇다	상세한 검색전략은 웹 부록을 통해 제공하였다. 인구집단(임산부)과 중재(초음파)의 조합은 적절하지 않은 제한(예: 연구설계 필터) 없이 민감해보였다.
2.4 날짜, 출판 형식 또는 언어에 따른 제한이 적절하였는가?	그렇다	데이터베이스의 시작일부터 검색하였고 언어제한을 두지 않았다. 회색문헌을 파악하는 단계도 포함되었다. 어떤 제한도 두지 않은 것으로 판단되어 "그렇다"라고 평가하였다.
2.5 연구 선택의 오류를 최소화하기 위한 노력이 이루어졌는가?	아 마 도 그렇다	제목과 초록을 선별하는 과정과 전문을 평가하는 과정은 다수의 평가자를 포함했다고 보고하였다. 그러나 평가자가 독립적으로 수행하였는지는 명확하지 않았다. 따라서 "아마도 그렇다"라고 평가하였다.
연구 파악과 선택방법에 대한 우려	낮음	모든 신호질문에 대해 "그렇다" 또는 "아마도 그렇다"로 평가되었고 따라서 잠재적 비플림의 영역도 파악되지 않았다. 이 문헌고찰은 관련연구를 포함하고 있을 가능성이 높다고 판단하였다.

2.2.3. 영역 3: 자료 수집 및 연구 평가

(1) 배경

세 번째 영역의 목적은 자료 수집이나 비뚫림위험 평가 과정 중에 비뚫림이 있었는지 여부를 평가하는 것이다. 엄격한 자료수집은 미리 프로토콜 단계에서 계획하고 파일럿한 구조화된 자료수집 양식 사용을 포함해야 한다. 연구결과 합성과 해석에 기여할 모든 자료는 수집되어야 한다. 이러한 자료에는 수치와 통계 자료 그리고 좀더 일반적인 일차 연구 특성들(연구설계, 연구비 출처, 세팅, 연구대상자 특성, 선택기준, 중재/노출/대상 진단검사의 세부 정보, 참여자 탈락 등)도 포함되어야 한다. 수치형 자료(numerical data) 유형은 문헌고찰에 포함된 연구설계에 따라 달라진다. 예를 들어, DTA 문헌고찰의 경우 검사 수행의 2×2 자료가 일반적으로 추출된다. 관찰연구의 문헌고찰의 경우, 95% 신뢰 구간과 보정된 효과 추정치, 보정 변수의 세부사항 및 분석 방법을 추출할 수 있다. 합성에 기여하는 데 요구되는 적절한 형식으로 자료를 이용할 수 없는 경우, 문헌고찰 저자는 이러한 자료를 어떻게 확보하였는지 보고해야 한다. 예를 들어, 일차 연구 저자에게 추가 자료를 문의할 수 있다. 필요한 자료를 도출하기 위해 적절한 통계 변환을 사용할 수 있다. 자료추출에 오류가 있을 수 있다. 데이터를 옮기는 과정에서나 연구 보고서에 나와있는 관련 정보를 수집하지 못한 경우 실수로 인해 오류가 발생할 수 있다. 자료추출은 본질적으로 주관적이고 해석에 따라 달라질 수 있기 때문에 그 과정에서 비뚫림이 발생할 수 있다. 중복(duplicate) 자료 추출(또는 엄격한 체크를 통한 단일 자료 추출)은 무작위 오류 및 잠재적 비뚫림을 방지하는 데 필수적이다.

포함된 연구의 타당성은 문헌고찰에 포함된 일차 연구의 설계를 고려한 적절한 기준을 사용하여 평가해야 한다. 이 평가는 평가중인 설계의 연구를 위해 특별히 개발된 검증된 도구를 사용하여 수행되거나, 단순히 중요한 잠재적 비뚫림의 원인이 될 수 있는 관련 기준의 목록일 수 있다. 출판된 도구 또는 특별히 이를 위한 기준을 사용하든지 간에, 평가자는 기준이 포함된 연구에서 모든 중요한 잠재적 비뚫림 출처(source)를 확인하기에 충분한지 여부를 고려해야 한다. 기준이 충분한지에 대한 결정은 특정 연구설계에서 비뚫림위험에 관한 경험적 증거에 근거해야 한다. 자료추출과 같이 비뚫림위험 평가 과정에서도 비뚫림 또는 오류가 발생할 수 있다. 따라서 비뚫림위험 평가는 두 명의 검토자가 독립적으로 작업하지만, 최소한 첫 번째 검토자의 결정을 두 번째 검토자가 확인해야 한다. 비뚫림위험 평가와 자료추출 모두에서 해결과 불일치 과정이 보고되어야 한다.

표 1-9는 각 질문의 평가지침과 함께 영역 3에 포함된 신호 질문을 요약한다.

표 1-10은 영역 3에 대한 등급(level) 평가의 개요를 제공한다. 표 2-11,2-12는 이 영역에 대한 평가자간의 우려사항이 높은 혹은 낮은 경우에 평가의 예이다.

표 1-9. 영역 3에 대한 신호 질문들과 평가지침

신호 질문	평가 지침
3.1 자료 수집 오류를 최소화하기 위해 노력하였는가?	자료수집 과정에서 비플림과 오류를 최소화하기 위해서는 최소한 두 명의 검토자가 필요하며 파일럿 과정을 거친 구조화된 자료 추출 양식을 사용해야 한다. 이상적으로는 독립적으로 수행해야 하지만 한 명의 검토자가 추출하고 두 번째 검토자의 검토도 허용된다. 두 번째 검토자가 논문을 자세히 읽고 추출된 자료의 정확성을 검사할 뿐만 아니라 관련 정보가 누락되지 않았는지 확인해야 한다.
3.2 연구결과를 해석할 수 있기에 충분한 연구 특성 정보가 문헌고찰 저자와 독자에게 제공되었는가?	연구 특성에 관한 정보는 이질성에 대한 적절한 조사와 연구결과와의 적용 가능성을 고려할 수 있으므로, 문헌고찰의 연구결과를 해석하는 데 필수적이다. 이 정보는 자료 수집에 대한 정보가 보고되는 방법 섹션에 보고되거나 포함된 연구표의 특성에서 사용 가능하거나 연구결과 본문에 요약될 수 있다. 문헌고찰의 일부로 수집된 모든 정보가 종종 공간 제한으로 인해 간행물에 모두 제시되는 것은 아니기 때문에, 이 질문은 판단하기 어려울 수 있다. 따라서 평가자는 웹부록과 같은 추가 출처에 접근해야 할 수도 있다.
3.3 합성에 사용되도록 타당한 모든 연구결과가 수집되었는가?	적절한 합성을 수행할 수 있도록 충분한 연구 결과를 추출해야 한다. 이상적으로, 문헌고찰 저자는 합성에 필요한 자료와 형식을 보고해야 한다. 예를 들어, DTA 문헌고찰을 위해 검사 수행의 2x2 자료가 일반적으로 추출된다. 관찰 연구의 문헌고찰을 위해, 관련 95% 신뢰 구간을 갖는 보정된 효과 추정치, 보정 변수의 세부 사항 및 분석 방법이 추출될 수 있다. 이러한 자료가 방법 섹션에 명시적으로 보고되지 않은 경우 연구결과 표, 그래픽(graphical) 요약(예: 숲 그림) 또는 본문에 보고된 자료에서 자료 추출을 할 수 있다. 사용 가능한 경우 웹부록 또는 문헌고찰 프로토콜과 같은 추가 출처에 접근해야 할 수도 있다. 문헌고찰에 포함된 모든 일차 연구가 합성에 기여하는 데 필요한 적절한 형식으로 자료를 보고하는 것은 매우 드물다. 예를 들어, 문헌고찰은 95% 신뢰 구간과 관련된 기저치(baseline) 변화에 대한 평균 차이 형식에서 연속형 자료가 필요할 수 있다. 변동 측정값을 보고하지 않은 연구에서 대신하여 보고될 수 있는 다양한 다른 측정치들이 있다. 이 질문에 대해 “그렇다”로 대답하기 위해서는 합성에 필요한 형식으로 보고되지 않은 연구결과 자료를 어떻게 얻었는지 설명하기 위한 자세한 정보가 방법 섹션에 포함되어야 한다. 예: 보고된 데이터에서 추정/변환 혹은 추가 정보를 위해 저자에게 문의

<p>3.4 비플립위험(또는 방법론적 질)이 적절한 기준을 사용하여 공식적으로 평가되었는가?</p>	<p>공식적인 비플립위험 평가는 모든 문헌고찰의 필수 구성 요소이다. 비플립위험이 공식적으로 평가되지 않은 경우 이 질문은 “아니다”로 대답해야 한다. 공식적인 평가를 수행한 경우 평가자는 사용된 기준이 적절한지 판단해야 한다. RCT용 Cochrane 비플립위험 도구 또는 DTA 연구용 QUADAS-2와 같이 출판된 도구가 적절한 설계에 사용된 경우, 매우 간단히 이 질문은 “그렇다”라고 대답할 수 있다.</p> <p>그러나 문헌고찰이 단순히 평가 질문을 나열하거나 출판되지 않은 도구 또는 더 이상 권장되지 않는 도구가 사용되는 경우에는 좀 더 복잡하다. 그런 경우 평가자는 연구설계에 따라 평가된 기준이 일차 연구에서 잠재적 비플립의 원인을 식별할 수 있는지 여부를 판단해야 한다. 예를 들어, Jadad 점수는 매우 널리 사용되었지만 RCT에서 주요 잠재적 비플립이 가능성을 나타내는 배정 은폐는 포함하지 않는다. 이 질문에 “그렇다”라고 대답하려면 Jadad 척도를 사용한 문헌고찰은 추가로 배정 은폐를 평가해야 한다.</p>
<p>3.5 비플립위험 평가의 오류를 최소화하기 위한 노력을 하였는가?</p>	<p>자료 수집과 마찬가지로, 비플립 평가의 위험에는 최소한 두 명의 검토자가 포함되어야 한다. 이상적으로는 독립적으로 수행해야하지만 한명 검토자의 평가와 두번째 검토자의 세부 검사도 허용된다.</p>

표 1-10. 자료수집과 연구평가에 대한 우려

<p>낮은 우려</p>	<p>영역 2에서 평가된 문헌고찰에 포함된 연구는 2명의 검토자가 적절한 기준, 자료 추출 및 비플립위험 평가를 사용하여 비플립위험을 평가했으며 관련 연구 특성과 결과를 추출했다.</p>
<p>높은 우려</p>	<p>자료 수집 또는 비플립위험 평가 과정을 통해 일부 비플립이 발생할 수 있다.</p>
<p>불확실한 우려</p>	<p>비플립위험에 대한 판단을 보고하기에는 정보가 충분하지 않다.</p>

(2) 등급부여의 예

a. 자료 수집 및 연구 평가에서 높은 우려로 판단된 문헌고찰

본문 기술 내용:

“Tables of 2×2 were constructed with the data extracted from each study. Two authors independently searched and extracted the data into an abstraction form. Any differences were resolved by mutual agreement... There is no consensus on how to assess studies designed without a control arm. Hence, these criteria do not apply to studies without a control arm. Therefore, for this meta-analysis and systematic review, studies were selected based on completeness of data and inclusion criteria.”

코멘트: 포함된 연구의 특성표에는 연구설계(후향적, 전향적), 압의 유형(모든 직장암) 및 확인적 수술(confirmatory procedure, 모든 수술)에 대한 세부 정보만 포함된다.

표 1-11. 우려가 높은 것으로 판단된 자료 수집 및 연구 평가에 대한 평가 예

신호질문	등급부여	이유
3.1 자료 수집 오류를 최소화 하기 위해 노력하였는가?	그렇다	추출 양식을 사용하여 두 명의 검토자가 독립적으로 데이터 추출을 수행했다. 불일치는 합의에 의해 해결되었다.
3.2 연구결과를 해석할 수 있 기에 충분한 연구 특성 정보 가 문헌고찰 저자와 독 자에게 제공되었는가?	아니다	요약표는 단지 연구설계(후향적, 전향적), 압의 유형(모든 직장 암) 및 확인적 수술(모든 수술)에 대한 정보만 포함한다. 본문에 연구설계나 인구집단에 대한 더 이상의 상세한 설명이 보고되 지 않았다.
3.3 합성에 사용되도록 타당 한 모든 연구결과가 수집 되었는가?	그렇다	각 연구에서 추출된 2×2 자료는 진단적 정확도 및 관련 신뢰 구간의 모든 측정값을 계산하고 메타 분석을 수행하기에 충분 하다.
3.4 비플림위험(또는 방법론 적 질)이 적절한 기준을 사용하여 공식적으로 평 가되었는가?	아니다	연구의 질은 공식적으로 평가되지 않았다. 저자들은 이것이 대 조군 없는 연구를 평가하는 방법에 대한 합의가 없기 때문이 며, 그러나 이용할 수 있는 DTA 연구 평가용 도구가 있었기 때 문이라고 진술했다.
3.5 비플림위험 평가의 오류 를 최소화하기위한 노력 을 하였는가?	아니다	연구의 질은 공식적으로 평가되지 않았다.
자료수집과 연구평가에 대한 우려	높음	공식적인 질 평가의 부재는 포함된 연구에서 비플림위험이 불 확실하다는 것을 의미한다. 독자가 결과를 해석할 수 있는 충 분한 연구 세부 사항이 없다. 따라서 이 문헌고찰에 대한 자료 수집 및 연구 평가 모두에서 비플림위험이 높다.

b. 자료 수집 및 연구 평가에서 낮은 우려로 판단된 문헌고찰⁵⁾

본문 기술 내용:

“One investigator abstracted details about the study design, population, setting, interventions, analysis, follow-up, and results. A second investigator reviewed data for accuracy. Two investigators independently applied predefined criteria to assess study quality as good, fair, or poor. Discrepancies were resolved through consensus.”

표 1-12. 우려가 낮은 것으로 판단된 자료 수집 및 연구 평가에 대한 평가 예

신호질문	등급부여	이유
3.1 자료 수집 오류를 최소화하기 위해 노력하였는가?	아 마 도 그렇다	한 명의 검토자가 자료 수집을 수행했으며 이는 두 번째 검토자가 정확성을 검사했다. 두 번째 검토자가 추출된 자료가 정확한지 여부만 확인했는지 또는 관련 정보를 놓치지 않도록 논문을 자세히 읽었는지 여부는 확실하지 않다. 따라서 이 질문은 “그렇다”가 아니라 “아마도 그렇다”로 대답했다.
3.2 연구결과를 해석할 수 있기에 충분한 연구 특성 정보가 문헌고찰 저자와 독자에게 제공되었는가?	그렇다	상세한 연구 특성 및 결과표는 문헌고찰 저자와 독자가 결과를 해석하기에 충분한 정보를 부록에 제공하였다.
3.3 합성에 사용되도록 타당한 모든 연구결과가 수집되었는가?	그렇다	이분형 자료는 각 증재군의 사건 수 및 총 환자 수로 추출되었다; 이는 상대 위험을 계산하는 데 사용되었다. 이것은 합성에 사용하기에 적합했다.
3.4 비둘림위험(또는 방법론적 질)이 적절한 기준을 사용하여 공식적으로 평가되었는가?	아 마 도 그렇다	비둘림위험 기준에 대한 상세내용은 제공되지 않았지만 문헌고찰 저자들은 적절한 기준을 포함하는 미국 USPSTF(US Preventive Services Task Force) 방법론 지침(2001)을 인용하였다.
3.5 비둘림위험 평가의 오류를 최소화하기위한 노력을 하였는가?	그렇다	두 명의 검토자가 독립적으로 연구의 질을 평가하였다.
자료수집과 연구평가에 대한 우려	낮음	모든 신호질문은 “그렇다” 또는 “아마도 그렇다”로 평가되었으므로 잠재적인 비둘림 영역은 확인되지 않았다. 따라서 자료 수집 및 연구 평가의 검토 과정은 문헌고찰에서 비둘림이 생기지 않을 것이다.

5) Chou R, Hartung D, Rahman B, Wasson N, Cottrell EB, Fu R. Comparative effectiveness of antiviral treatment for hepatitis C virus infection in adults: a systematic review. *Ann Intern Med* 2013;158(2):114-23.

2.2.4. 영역 4 : 합성 및 연구 결과

(1) 배경

이 영역은 포함된 일차 연구의 자료를 결합(질적 또는 양적 합성)하기로 결정한 경우, 검토자들이 적절한 방법을 사용했는지 평가하는 것을 목표로 한다. 합성을 위한 접근은 다루고 있는 검토 질문의 성격과 합성되는 일차 연구의 성격에 달려있다.

RCT의 경우, 일반적인 접근법은 치료 효과 추정치의 가중 평균(치료 효과의 비율 측정에 대한 로그 척도)을 사용하여 추정치의 정밀도에 의한 가중치를 구하는 것이다. 고정 효과 또는 변량 효과 모델을 가정할 수 있다. 그러나 여기에 다양한 변이와 확장이 있는데, 즉 결과 자료를 명시적으로 모델링 할 수 있는 옵션(예: 이분형 자료에 대한 로지스틱 회귀 접근, 두 개 이상의 결과를 동시에 모델링 하는 이변량 또는 다변량 메타 분석, 다중 치료 효과 모델링(네트워크 메타 분석) 혹은 치료 효과의 모델링 변화(메타 회귀), 그리고 이것들을 조합하여 합성을 매우 복잡하게 할 수 있다.

다른 유형의 문헌고찰 질문에 대해서도 유사한 옵션을 사용할 수 있다. 진단검사 정확도를 위해, 이변량 접근법이 표준이 되었으며, 민감도와 특이도는 상관관계를 고려하여 동시에 모델링된다. 일부 문헌고찰의 경우 통계적 합성이 적절하지 않을 수 있으며 대신 결과의 질적 또는 서술적 개요가 보고되어야 한다.

(질적 또는 양적) 합성에서 고려해야 할 가장 중요한 측면은 다음과 같다. (i) 분석 접근법이 제안된 연구질문에 적합한지; (ii) 연구 간 변이(이질성)가 고려되는지 여부; (iii) 일차 연구에서 비뚤림이 고려되는지 여부; (iv) 합성되는 일차 연구의 정보가 완전한지(특히, 누락 또는 누락된 자료가 가용한 자료와 체계적으로 다른 경우, 예로 출판 또는 보고 비뚤림); (v) 검토자가 그들이 찾은 결과를 보고하는 방식에 비뚤림이 있는지 여부.

추정 방법의 선택과 같은 메타 분석 방법의 기술적 측면은 중요하지 않을 것이다. 그러나 표준 오차를 표준 편차로 해석하거나 매칭되거나 클러스터 된 자료와 같은 설계 문제를 조정하지 못했거나, 표준 가중 평균 접근 방식을 로그 대신 위험비(risk ratio)로 적용하는 등의 실수가 중요할 수 있다.

표 1-13에는 각 질문에 대한 평가 지침과 함께 영역4에 포함된 신호 질문이 요약되어 있다.

표 1-14는 영역 4에 대한 영역수준 평가의 개요를 제공한다. 평가자의 영역에 대한 우려가 높거나 낮은 문헌고찰 평가 예가 표 2-15, 2-16에 제시되었다.

표 1-13. 영역 4에 대한 신호 질문들과 평가지침

신호질문	등급부여 지침
<p>4.1 포함되어야 하는 모든 연구가 자료합성에 포함되었는가?</p>	<p>합성은 일반적으로 연구질문과 관련된 자료가 있는 모든 연구를 포함하려고 해야 한다. (아마도 출판 비둘림 때문에) 문헌고찰 저자가 연구를 찾지 못해서 개별 연구의 연구결과가 합성에서 누락될 수 있다. 이렇게 누락된 연구는 영역 2와 신호질문 4.5에 의해 다루어진다.</p> <p>이 질문은 문헌고찰에 포함된 연구와 관련되어 세 가지 추가 상황을 다룬다. (i) 포함된 연구의 특정 연구결과를 문헌고찰 저자가 이용할 수 없다. (ii) 문헌고찰 저자가 이용할 수 있는 자료를 수집하거나 처리하지 못했다. (iii) 문헌고찰 저자가 의도적으로 연구결과를 배제하였다. (i)의 경우, 수집된 자료를 포함했지만 원하는 연구결과를 보고하지 않았을 수 있다. 예를 들어 임상시험에서 관심있는 임상 중재결과에 대한 자료를 수집할 수 있지만 그에 대한 연구결과는 보고하지 않을 수 있다. 여기에서 중요하게 구별해야할 점은 통계적 유의성과 같은 결과로 인해 연구결과가 제대로 표현되지 않았을 가능성이 높았는지 또는 보고서 길이 제한과 같은 실질적인 이유로 단순히 보고되지 않았을 가능성이 높았는지이다. 연구결과가 통계적으로 유의하지 않았기 때문에 하나 이상의 연구에서 연구결과를 사용할 수 없는 경우, 이러한 선택적 (비)보고 때문에 자료 합성에서 비둘림을 초래할 것이다. (ii)의 경우 (영역 3의 신호질문 3.3 참고), 실수로 연구가 생략되었거나, 문헌고찰 저자가 통계적 계산에 익숙하지 않아서 생략되었을 수도 있다. 이는 누락된 연구결과가 포함된 연구와 체계적으로 다른 경우 문제가 될 수 있다. (iii)에 대해 문헌고찰 저자는 일부 연구를 합성에서 제외하는 부적절한 결정을 내릴 수 있다. 예를 들어 통계적 고려사항에 기반하여 연구 간 이질성을 큰 것으로 간주되는 연구를 제외하는 것이다(민감도 분석으로서 합리적일 수도 있다).</p> <p>이러한 고려사항을 다루기 위해 평가자는 연구 수(예: 흐름도)와 합성 연구 수(예: 숲 그림 또는 표)를 검토하는 것이 좋다. 여기에서 불일치하게 되면 왜 연구가 분석에 포함되지 않았는지 고민하게 된다. 많은 수의 연구(예: 10개 또는 20개 이상의 합성)를 가진 문헌고찰에서 깔때기 그림과 같은 통계적 분석은 연구 내에서 선택적 (비)보고 비둘림 가능성에 대한 힌트를 제공할 수 있다. 신호질문 4.5 참조.</p>
<p>4.2 사전에 정의된 모든 분석을 보고하였거나 이탈(departure)에 대해 설명하였는가?</p>	<p>이 질문의 목적은 문헌고찰 저자가 선호하지 않은 연구결과를 보고하지 않거나 대체하는 방식으로 분석하거나 분석 방법을 선택하여 발생하는 비둘림을 확인하는 것이다. '그렇다'라고 대답하려면 문헌고찰이 공개되거나 접근가능한 프로토콜을 따라야 한다(예를 들어 PROSPERO, Cochrane 리뷰 프로토콜). 사전에 정의된 분석이 수행되었다고 보고한 경우, 예를 들어 방법 섹션이 엄격해 보이고 언급된 모든 분석을 연구결과에서 다루고 있으면 '아마도 그렇다'라고 대답할 수 있다. 사전에 정해진 프로토콜이 명확히 없는 경우 '아니다'라고 대답해야 한다. 그러나 문헌고찰에 프로토콜의 유무에 대한 언급이 없는 경우 '정보 없음'으로 대답할 것을 권장한다.</p>
<p>4.3 포함된 연구들의 연구질문, 연구설계, 연구결과와 특성과 유사성에 근거해보았을 때 합성이 적절하였는가?</p>	<p>이 질문은 주로 합성 방법론을 다룬다. 방법론은 연구의 본질과 질문의 본질에 의해 주도되어야 하므로 합성 방법의 평가의 일부로 이러한 측면을 고려하는 것이 중요하다. 양적 합성이 수행되는 경우(메타 분석), 이 질문은 이를 수행하는 것이 적절한지 여부와 사용된 통계적 방법을 다룬다. 서술적 접근 방식이 합성에 사용된다면, 이 질문은 양적 합성이 더 적절한지 여부와 특정 서술적 접근 방식이 적절한지 여부를 다룬다.</p>

	<p>양적 합성이 적절한지에 대한 결정은 주로 연구질문, 설계 및 증재결과(측정되는 경우)에서 연구들의 유사성에 의해 결정된다. 이를 종종 '임상적 이질성' 또는 '임상적 다양성'이라고 한다. 합성의 결과가 의미를 갖는 것, 즉 정책, 실무 또는 추가 연구에 정보를 제공할 수 있음을 평가자가 확신할 수 있는 것이 중요하다. 통계적 이질성 문제는 다음 질문에서 명시적으로 해결된다.</p> <p>각 연구에서 적절한 가중치를 포함하여 연구를 통계적으로 통합하는 '허용된' 방법이 사용된 경우 평가자는 이 질문에 "그렇다"라고 대답해야 한다. 예를 들어, 연구 간 변이가 존재하지만 치료 효과가 공동 분포로 나온 것으로 가정할 수 있는 경우, 무작위 효과 메타 분석 모델이 적절할 수 있지만 고정 효과 모델은 그렇지 않을 수 있다. 그러나 연구 크기와 효과 크기('작은 연구 효과(small study effect)' 또는 '비대칭적인 깔대기 그림') 사이에 강한 관계가 있는 경우, 단순 무작위 효과 메타 분석은 적합하지 않다. 연구 간 변이에 대한 이유를 탐색하기에 충분한 연구가 있는 경우, 회귀 접근 또는 하위 그룹화를 고려할 수 있다. 일반적으로 전체 합쳐진 추정치에 대한 연구 정밀도의 영향을 무시하는 연구를 결합하는 방법, 즉 모든 연구에 동일한 가중치를 부여하는 방법을 권장하지 않는다. 이러한 방법이 사용된 경우(예: 투표 수) 평가자가 이 질문에 "아니다"라고 대답하는 것을 제한한다.</p> <p>진단적 검사 정확도에 대한 연구에서 일반적인 방법은 민감도와 특이도 사이의 상관관계를 설명하기 위한 이변량 접근법이다. 다변량, 다단계, 베이지안(Bayesian) 또는 네트워크 메타 분석 방법을 사용하는 메타 분석에는 통계적 조인이 필요할 수 있다.</p>
<p>4.4 연구간 변이(이질성)가 작거나 자료의 합성 중 다루어졌는가?</p>	<p>이 질문은 연구 특성에서의 변이보다 연구결과에서의 변이에 대한 질문이다. 연구 간 변이는 시각적으로 평가될 수 있고(예: 숲 그림); 통계적 검사를 사용하고(예: Cochran의 x^2 검증 또는 Q); 연구 간 변이의 측정값을 사용하거나(종종 I^2라고 함) 또는 신뢰 구간 겹침에 대한 측정값을 사용한다(예: I^2 비일관성 통계량).</p> <p>메타 분석에서 실질적인 이질성이 무시된다면, 오해의 소지가 있는 결론과 또는 가짜 정밀도로 이어질 수 있다. 이질성의 존재하에 고정 효과 분석이 사용되는 경우, 분석에서 이질성을 무시된다는 것을 인정하는 것이 중요하다. 연구결과는 연구 간 변이에 관한 연구결과로 보충되어야 하고, 고정 효과 분석 연구결과는 적절하게 해석되어야 한다. 대부분의 경우, 변이의 원인을 탐색하기 위해 추가 분석없이 고정 효과 메타 분석을 사용하면 이 질문에 대해 "아니다"라고 답한다.</p> <p>이질성을 조사하기 위해, 그리고 하위그룹/메타 회귀 분석을 실행하기 위해 변량 효과 모델을 적절하게 사용되면 평가자는 "그렇다"라고 답한다. 그러나 특성이나 연구 결과가 매우 다양한 연구의 변량 효과 메타 분석, 특히 연구결과가 반대 방향으로 향하는 경우 의미없는 연구결과를 낳을 수 있다.</p> <p>임상적 이질성으로 인해 통계적 통합이 부적절하다는 것을 기초로 서술적 합성이 수행되었다면, 통합하지 않는 것으로 이질성이 처리되었기 때문에 이 질문은 "그렇다"로 대답한다.</p>
<p>4.5 연구결과가 견고하였는가? 예를 들어 깔대기 그림(funnel plot)이나 민감도 분석으로 연구결과의 견고성을 보여주었는가?</p>	<p>이 질문은 자료의 합성에 의한 연구결과의 불안정성과, 합성 방법을 변경함으로써 연구결과가 변할 수 있는지 여부를 다룬다. 깔대기 그림은 효과 크기와 연구 크기 사이의 관계를 조사하기 위해 사용된다(대개 정밀도를 사용하여 통계적으로 측정). 숲 그림의 비대칭성은 출판 비뮴림, 연구 내 (비)보고 비뮴림, 큰 연구와 작은 연구의 서로 다른 비뮴림의 크기 또는 다양한 크기의 연구에 기초한 효과의 진정한 차이 등 여러 가지 이유로 나타날 수 있다.</p>

	<p>깔때기 그림이 비대칭일 때 표준 고정 효과와 변량 효과 방법은 서로 다른 결과를 산출한다. 따라서 민감도 분석의 한 형태로 두 가지 방법을 수행하는 것이 종종 유익하다. 또는 절삭과 채움 방법(trim-and-fill method), 회귀 접근법 또는 선택 모델의 사용을 통해 다양한 유형의 출판 비뚤림이 연구결과에 미치는 잠재적 영향을 조사할 수 있다.</p> <p>민감도 분석의 다른 예로는 누락된 데이터를 대체(impute)하기 위해 다른 방법을 사용하거나, 합성한 연구결과를 강력하게 이끌어내는 것으로 생각되는 단일 연구를 배제하는 것이다. 또는 민감도 분석을 수행하여 질병에 대한 공식적인 진단 기준을 사용하지 않은 연구 또는 비뚤림위험이 높은 것으로 평가된 연구를 제거해도 결과가 확실한지 확인할 수 있다.</p> <p>민감도 분석을 했다는 것 자체만으로는 이 질문에 “그렇다”의 답하는 데 적절하지 않은 것으로 간주된다. 평가자는 민감도 분석으로 연구결과의 견고성을 입증했는지 여부로 결정해야 한다. 연구가 매우 적거나 매우 이질적인 연구가 있는 경우, 검토자가 민감도 분석을 하지 않았더라도 연구결과가 견고하지 않다는 것이 분명할 수 있다.</p> <p>서술적 합성이 수행된 경우, 평가자는 연구 요약에 대한 다양한 접근방식이 서로 다른 결론을 도출할 수 있는지와, 단일 연구가 결론을 이끌어내고 있는지 여부를 고려해야 한다.</p>
<p>4.6 일차 연구에서 비뚤림위험이 미미했거나, 비뚤림위험이 자료 합성 과정에서 고려되었는가?</p>	<p>문헌고찰 저자가 무시한 구성적 연구에 중요한 비뚤림이 있다고 판단하는 경우 이 질문에 “아니다”로 대답할 것을 권장한다. 예를 들어, 체계적 문헌고찰에서 비뚤림위험이 평가되지 않은 경우 ROBIS 평가자는 “아니다”라고 대답해야 한다. 마찬가지로, 비뚤림위험이 평가되었지만 문헌고찰 저자들이 그것을 연구결과/결론에 포함시키지 않은 경우에도 이 질문에 대해 “아니다”라고 대답해야 한다. 이 질문에 “그렇다”라고 대답할 수 있는 예로는 문헌고찰 저자가 “비뚤림위험 낮음” 등급을 받은 모든 연구를 포함하거나 비뚤림의 위험이 높은 연구가 사용된 경우 민감도 분석/보정 방법을 쓴 경우이다. 비뚤림이 연구결과의 토론의 일부로만 다루어지는 경우, 평가자는 “아니다”라고 대답해야 한다. 이러한 토론은 문헌고찰의 비뚤림위험에 대한 전반적인 판단에서 나중에 다루어진다(3단계).</p>

표 1-14. 합성 및 연구 결과에 대한 우려

<p>낮은 우려</p>	<p>합성은 비뚤린 연구결과를 낳지 않을 것 같다. 왜냐하면 자료의 한계를 극복했거나, 그 연구결과가 설득력이 있어 그 한계가 거의 영향을 미치지 않았다.</p>
<p>높은 우려</p>	<p>합성에 편향이 있을 수 있다. 왜냐하면 (i) (연구 내 그리고/또는 연구 간) 잠재적 비뚤림이 무시되었거나, (ii) 연구 간 중요한 변이가 설명되지 않았거나, (iii) 방법론에 중요한 부적절성이 있었거나, (iv) 연구결과가 우려스러운 정도로 불안전하게 보고되었다.</p>
<p>불확실한 우려</p>	<p>보고된 결과가 불충분하기 때문에 비뚤림위험을 평가할 수 없다.</p>

(2) 등급부여 예시

a. 합성 및 연구 결과에서 높은 우려로 판단된 문헌고찰⁶

표 1-15. 합성에 대한 등급부여 예시(‘높음’으로 평가)

신호질문	등급부여	이유
4.1 포함되어야 하는 모든 연구가 자료합성이 포함되었는가?	정보 없음	저자는 35개의 연구가 문헌고찰과 관련이 있으며, 그 중 29개는 메타 분석에서 사용할 수 있다고 말했다. 이를 확인하기 위한 다른 정보는 제공되지 않았다.
4.2 사전에 정의된 모든 분석을 보고하였거나 이탈(departure)에 대해 설명하였는가?	정보 없음	참고한 프로토콜에서는 어떤 분석도 미리 정의되지 않았다. 본문에서는 더 이상의 정보가 제공되지 않았다.
4.3 포함된 연구들의 연구질문, 연구설계, 연구결과와 특성과 유사성에 근거해보았을 때 합성이 적절하였는가?	아니다	저자는 “메타 분석이 수행되었다”고만 진술했다(p.11). 저자는 사건 및 인구집단을 요약하고 합계를 사용하여 요약 오즈비를 생성하였고, 연구 내 비교를 활용하지 않았다. 저자는 연구를 통합한다고 기술했지만, 연구 전반에 걸쳐 임상적 다양성이 있는 것으로 보인다.
4.4 연구간 변이(이질성)가 작거나 합성 중 다루어졌는가?	아마도 아니다	통계적 이질성에 대한 자세한 내용은 제공되지 않았다. 그러나 저자는 높은 비플림에 대한 하위군 분석, 파트너 연구 및 무작위 인구집단 연구에 대해 논의하였다. 또한 연구들은 지역사회의 포피 절제(circumcision) 유병률에 따라 분류되었다. 하위 그룹에 대한 자세한 내용은 제시되지 않았다.
4.5 연구결과가 견고하였는가? 예를 들어 깔대기 그림(funnel plot)이나 민감도 분석으로 연구결과의 견고성을 보여주었는가?	아마도 아니다	저자는 문제가 될 수 있는 몇몇 연구에 대해 논의했다. “이러한 가정은 잘못된 것일 수 있지만, 이러한 연구의 포함 또는 배제는 전체 결과에 영향을 미치지 않았다.”(P.11). 이 분석에 대한 오즈비 및 신뢰 구간은 보고되지 않았다.
4.6 일차 연구에서 비플림위험이 미미했거나, 비플림위험이 자료 합성 과정에서 고려되었는가?	아니오	그 연구는 질이나 비플림위험에 대해 명시적으로 평가되지 않았다. 비플림은 합성에서 명시적으로 다루어지지 않았다.
합성과 연구결과에 대한 우려	높음	개별 연구는 비플림위험에 대해 평가되지 않았으며, 합성에서 잠재적 비플림도 설명되지 않았다. 저자는 메타분석에서 위험한 접근 방식인 연구 내 비교를 분석하기보다는 연구 전반에 걸쳐 중재 집단을 요약하였다. 분석에서는 이질성에 대한 논의나 평가가 없었다.

6) Van Howe RS. Circumcision and HIV infection: review of the literature and meta-analysis. Int J STD AIDS 1999;10(1):8-16.

b. 합성에서 낮은 우려로 판단된 문헌고찰⁷

표 1-16. 낮은 우려로 판단된 합성의 등급부여 예시

신호질문	등급부여	이유
4.1 포함되어야 하는 모든 연구가 자료합성이 포함되었는가?	정보 없음	문헌선정 흐름도는 고찰에 포함된 43개의 연구를 보고했다. QUADAS-2의 등급을 보고한 별도의 수치에서도 43개의 연구 결과가 나왔다. 또한 43개의 연구를 보고한 연구 특성의 세부 사항이 포함된 부록이 포함되었다. 그러나 연구들이 하나 이상의 중재결과를 보고할 수 있으므로, 연구별로 연구결과를 제시하지 않는 요약 숲 그림을 보고하였다. 모든 연구가 이 합성에 포함되었는지 여부는 확인할 수 없다.
4.2 사전에 정의된 모든 분석을 보고하였거나 이탈(departure)에 대해 설명하였는가?	아 마 도 그렇다	프로토콜은 부록 정보로 이용 가능했다. 어떠한 이탈도 명시되거나 관찰되지 않았다. 프로토콜은 비둘림위험 평가나 분석을 위한 전략에 대한 세부사항을 명시하지 않았지만, 논문의 방법 섹션은 각각 엄격한 방식으로 다루어졌다. 방법 섹션의 모든 분석은 결과에서 다루어졌다.
4.3 포함된 연구들의 연구질문, 연구설계, 연구결과와 특성과 유사성에 근거해보았을 때 합성이 적절하였는가?	그렇다	저자들은 “양수(amniotic fluid)의 수치의 변량 효과 메타 분석 모델을 사용하여 결과를 합성했다”(688페이지). 예측 능력을 평가하기 위해 2×2 표의 데이터를 사용하여 요약 민감도, 특이도 및 우도비를 계산하고 이변량 변량 효과 메타 분석 모델을 사용하여 합성했다.
4.4 연구간 변이(이질성)가 작거나 합성 중 다루어졌는가?	그렇다	통계적 이질성은 이 고찰에서 예상되었으며 변량 효과 모델을 사용하여 다루었다. 잠재적 효과 추정치를 조사하기 위해 메타 회귀 또는 하위군 분석을 수행했다. I2 >0%일 때도 예측 구간이 사용되었다.
4.5 연구결과가 견고하였는가? 예를 들어 깔대기 그림(funnel plot)이나 민감도 분석으로 연구결과와 견고성을 보여주었는가.	아 마 도 아니다	저자들은 민감도 분석이 연구결과와 견고성(robustness)을 평가하는 데 사용되었는지 여부를 밝히지 않았다. 그러나 복합 중재결과(composite outcome)의 사용에 대한 논의가 있다.

7) Morris RK, Meller CH, Tamblin J, Malin GM, Riley RD, Kilby MD, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and metaanalysis. BJOG: An International Journal of Obstetrics & Gynaecology 2014;121(6):686-99.

<p>4.6 일차연구에서 비플립이 미미하거나 또는 비플립이 합성할 때 다루어졌는가?</p>	<p>아 마 도 그렇다</p>	<p>비플립은 QUADAS-2 툴을 사용하여 평가되었다. 이 평가는 참고 표준에 대한 적절한 설명과 기준 표준에 대한 눈가림에서 비플립위험이 높다는 것을 나타냈다. QUADAS-2의 다른 측면은 크게 우려되지 않았다. “질”은 하위군 분석으로 명시되었지만 하위군 분석은 두 가지 주요 결과 중 하나에 대해서만 보고되었다. 양수과소증과 태아 웰빙 증재결과에 대해 저자들은 “하위군 분석에서 차이가 없었다”(P.691)고 말했다.</p>
<p>합성과 연구결과에 대한 우려</p>	<p>낮음</p>	<p>저자는 그들의 분석에서 이질성을 충분히 다루고 하위군 분석을 사용하여 탐구했다. 개별 연구의 비플립위험을 다루었고 저자들은 그것이 하위군 분석으로 포함되었다고 언급했다.</p>

2.3. 3단계: 비플립 판단

최종 단계에서는 문헌고찰이 전체적으로 비플립위험이 있는지 검토한다. 이 평가는 2단계 영역과 동일한 구조를 사용하며, 여기에는 판단을 뒷받침하는 데 사용되는 신호 질문과 정보가 포함되지만, 비플립 우려에 대한 판단은 비플립위험의 전반적인 판단으로 대체된다. 이 단계의 첫 번째 신호 질문은 연구결과의 해석이 영역 1에서 4까지 확인한 모든 문제를 해결하는지 여부를 묻는다. 문제가 발견되지 않으면 “그렇다”로 대답할 수 있다. 이전 영역 중 하나 이상의 문제가 확인되었지만 결과를 해석하고 결론을 도출할 때 적절히 고려되었다면, 이 또한 “그렇다”로 평가될 수 있으며, 다른 신호 질문의 등급에 따라 문헌고찰은 여전히 “비플립위험 낮음”으로 평가될 수 있다.

이 단계에는 문헌고찰 연구결과의 해석과 관련된 추가적인 세 가지 신호 질문이 포함된다. 여기에는 문헌고찰 결론이 제시된 근거에 의해 뒷받침되었는지 여부, 포함된 연구의 관련성을 고려했는지 여부, 문헌고찰 저자가 통계적 유의성에 근거하여 연구결과를 강조하지 않았는지 여부가 포함된다. 이들은 문헌고찰 연구결과에 비플립이 초래될 수 있는 해석의 중요한 측면이다.

표 1-17에는 영역 4에 포함된 신호질문과 각 질문에 대한 등급 지침이 요약되어 있다. 표 1-18은 영역 4에 대한 영역 수준 등급의 개요를 제공한다.

표 1-17. 3단계에 대한 신호질문과 평가 방법

신호질문	등급부여 지침
a. 연구결과를 해석할 때 2단계에서 확인한 모든 우려 사항을 다루었는가?	이 질문은 2단계에서 수행되고 3단계 시작 시 요약된 문헌고찰 과정에 대한 우려의 평가를 다시 언급한다. 2단계에서 평가된 모든 영역이 “낮은 우려”로 평가되었다면, 이 질문은 “그렇다”로 대답할 수 있다. 만약 하나라도 “높은 우려” 또는 “불확실한 우려”로 평가되었다면, 검토자들은 문헌고찰 저자가 연구결과 해석에서 2단계 동안 확인된 우려를 적절히 다루었는지 여부를 고려해야 한다. 예를 들어, 관련 연구가 누락되었을 수 있다는 것을 의미하는, 검색에 연구설계 필터를 적용했기 때문에 비독립위험이 높다고 판단됨을 고려해 보라. 이 문헌고찰이 이를 잠재적 한계로 인정하고 문헌고찰의 토론과 결론에서 문헌고찰 결과에 어떤 영향을 미칠 수 있었는지를 논의한 경우, 이 신호 질문은 “그렇다”로 평가될 수 있다.
b. 문헌고찰의 연구질문에 대해 포함된 연구의 관련성이 적절히 고려되었는가?	문헌고찰 결과를 해석할 때 중요한 측면은 확인된(포함된) 연구와 문헌고찰의 연구질문에 대한 관련성(적용성/외부 타당성)을 고려하는 것이다. 문헌고찰에 포함된 연구가 연구질문에 직접 적용되지 않고 문헌고찰 연구결과를 해석할 때 이 점을 고려하지 않는 경우 비독립위험이 초래될 수 있다. 일부 문헌고찰에서는 공식적인 질 평가의 일부로 포함된 연구의 관련성을 고려할 수 있다. 이는 문헌고찰 질문에 대한 일차 연구의 적용 가능성에 대한 공식적인 평가와 일차 연구의 비독립위험 평가를 포함하는 QUADAS-2 도구를 사용한 진단 정확도 문헌고찰의 경우 특히 그렇다. 다른 문헌고찰은 토론 섹션에서 연구의 목적 적합성에 대한 논의와 함께 덜 공식적인 접근법을 사용할 수 있다. 문헌고찰의 결론이 포함된 연구의 관련성을 포함하여 근거를 적절하게 반영하는 한 어떤 방법이든 적절할 수 있다.
c. 문헌고찰 저자는 통계적인 유의성을 보이는 결과만을 강조하는 것을 피하였는가?	단일 문헌고찰에서 복수의 분석이 보고되는 경우 문헌고찰 저자는 모든 분석의 균형 잡힌 설명을 제시하는 것이 중요하다. 때때로 문헌고찰 저자는 통계적 유의성에 기초하여 연구결과를 강조하도록 선택할 수 있다. 예를 들어, 문헌고찰에서 10개의 중재결과를 평가하여 9개의 중재결과에 대한 효과가 없고 1개의 유의한 효과가 있는 경우, 그 중재결과를 지나치게 강조하고 싶을 수 있다. 그러나 이것은 문헌고찰 연구결과와 실제 연구결과에 대한 오해의 소지가 있는 그림을 줄 수 있으므로, 이렇게 해석는 안된다.

표 1-18. 연구를 확인하고 선택하기 위해 사용된 방법에 의해 개입된 비독립위험

비독립위험 낮음	이 문헌고찰의 연구결과는 믿을 만한 것 같다. 2단계에서 문헌고찰 과정에 대한 우려를 제기하지 않았거나 문헌고찰 결론에서 우려를 적절하게 고려하였다. 결론은 근거에 의해 뒷받침되었고 포함된 연구의 관련성에 대한 고려가 포함되었다.
비독립위험 높음	2단계 평가 중 제기된 우려 중 하나 이상이 문헌고찰 결론에서 다루지 않았거나, 문헌고찰 결론이 근거에 의해 뒷받침되지 않았거나, 결론이 문헌고찰 질문과 포함된 연구의 관련성을 고려하지 않았다.
비독립위험 불확실	보고된 정보가 불충분하기 때문에 비독립위험을 평가할 수 없다.

(1) 등급부여 예시

a. 비플립위험 높음으로 판단된 문헌고찰⁸

표 1-19. 2단계 평가 중 확인된 우려 사항 요약

영역	우려	우려에 대한 근거
1. 연구 적격 기준의 세부사항에 대한 우려	높음	적격기준이 명확하지 않고 애매모호하다. 적격 인구집단에 대한 세부사항은 보고되지 않았다.
2. 연구의 검색 그리고/또는 선택 방법에 대한 우려	높음	검색에서 관련 연구가 누락되었을 가능성이 높다. 세부 정보가 제한적으로 보고되었지만, 적절한 범위의 데이터베이스가 검색된 것으로 보이지 않고, 검색 전략에는 진단 연구설계 용어, 데이터베이스 검색에 추가되는 방법이 보고되지 않았으며, 1980년 이후로의 검색 제한은 정당화되지 않는다.
3. 자료수집과 연구평가 방법에 대한 우려	높음	공식적인 질 평가가 없다는 것은 포함된 연구에서 비플립의 위험이 불확실하다는 것을 의미한다. 독자가 결과를 해석할 수 있도록 하기에는 연구 세부 정보가 충분하지 않았다. 따라서 데이터 수집과 연구 평가 과정 모두에서 비플립위험이 높다.
4. 합성 및 연구 결과에 대한 우려	높음	이 문헌고찰에서 이용한 분석 방법이 통계적으로 가장 견고하지는 않지만 신뢰할 수 있는 것으로 보인다. 이 문헌고찰의 주요 관심사는 이질성, 특히 하나의 특이한(outlying) 연구결과는 조사되지 않았다는 것이다. 매우 제한된 연구 세부사항이 제시되고 비플립위험 평가 수행되지 않았기 때문에 독자는 이 연구가 다른 연구와 다른지 여부를 평가할 수 없다. 만약 이 연구가 문헌고찰에 포함된 다른 연구들보다 더 신뢰할 수 있을 것 같으면, 문헌고찰의 전반적인 연구결과는 편향될 것이다.

8) Puli SR, Bechtold ML, Reddy JB, Choudhary A, Antillon MR. Can endoscopic ultrasound predict early rectal cancers that can be resected endoscopically? A meta-analysis and systematic review. Dig Dis Sci 2010;55(5):1221-9.

본문 기술 내용:

“This meta-analysis and systematic review shows that the pooled sensitivity and specificity of EUS for T0 tumor invasion is very high(around 97%)… Heterogeneity among different studies was evaluated not only with test of heterogeneity but also by drawing SROC curves and finding the AUC, since different studies might use slightly different criteria for staging. An AUC of 1 for any diagnostic test indicates that the test is excellent. SROC curves for EUS showed that the value of AUC was very close to 1, indicating that EUS is an excellent diagnostic test for T0 stage of rectal cancers… EUS has excellent sensitivity and specificity, which helps to accurately diagnose T0 stage of rectal cancers..”

표 1-20. 비뚤림위험 평가 예시(‘높음’으로 평가)

신호질문	등급부여	이유
a. 연구결과를 해석할 때 2단계에서 확인한 모든 우려 사항을 다루었는가?	아니다	2단계 평가에서 확인된 한계 중 어느 것도 문헌고찰 저자에 의해 한계점으로 확인되지 않았으며, 따라서 연구결과 해석에 다루어지지 않았다.
b. 문헌고찰의 연구질문에 대해 포함된 연구의 관련성이 적절히 고려되었는가?	아니다	문헌고찰 질문에 포함된 연구의 관련성을 고려하지 않았으며, 특히 인구집단과 관련하여 독자가 이 평가를 하기에는 세부 사항이 불충분했다.
c. 평가자는 통계적인 유의성을 보이는 결과만을 강조하는 것을 피하였는가?	그렇다	토론에서 강조된 수치와 이로부터 얻어진 결론은 모두 1차 분석에서 도출한 요약 추정치들이다.
비뚤림위험	높음	2단계 평가에서 저자가 다루지 않은 문헌고찰 절차와 관련된 많은 우려가 있다. 여기에는 포함 기준의 명확성 부족, 연구 누락 가능성, 질 평가의 결여, 포함된 연구에 대한 불충분한 세부 사항, 합성에서 연구 간의 차이를 적절히 고려하지 않는 것 등이 포함된다. 또한 문헌고찰은 선택된 연구와 문헌고찰의 연구질문의 관련성을 고려하지 않았다. 따라서 이 문헌고찰은 비뚤림위험이 높다.

b. 비플림위험 낮음으로 판단된 문헌고찰

표 1-21. 2단계 평가 중 확인된 우려 사항 요약

영역	우려	우려의 근거
1. 연구 적격 기준의 세부사항에 대한 우려	낮음	모든 신호 질문은 “그렇다” 또는 “아마도 그렇다”로 답했기 때문에 적격 기준에 대한 잠재적 우려는 파악되지 않았다. 문헌고찰 질문과 목표를 명확히 명시하였고, 적절하고 상세한 적격 기준을 사전에 구체화하고 정당화하기 위해 상당한 노력을 기울였다.
2. 연구의 검색 그리고/또는 선택 방법에 대한 우려	낮음	제목과 초록을 통해 선별하고, 전문을 평가하는 과정에서 복수의 검토자를 포함했다. 그러나 검토자가 독립적으로 행동했는지는 명시되지 않아 이 항목은 “아마도 그렇다”로 평가되었다.
3. 자료수집과 연구평가 방법에 대한 우려	낮음	모든 논문은 최소 2명의 검토자가 독립적으로 평가하고 적절한 자료를 독립적으로 추출했다. 연구 질은 적절한 도구를 사용하여 공식적으로 평가되었다.
4. 합성 및 연구 결과에 대한 우려	낮음	저자는 분석에서 이질성을 만족스럽게 다루고 하위 그룹 분석을 사용하여 탐색하였다. 개별 연구의 비플림위험을 다루었고 저자들은 그것을 하위 그룹 분석으로 포함하였다고 기술하였다.

본문 기술 내용:

Conclusions

“Current evidence suggests that oligohydramnios is strongly associated with being small for gestational age and mortality, and polyhydramnios with birthweight >90th centile. Despite strong associations with poor outcome, they do not accurately predict outcome risk for individuals”

Discussion

“Our quality assessment revealed concerns regarding possibility of bias through patient selection, performance of the index test and reference standard. We were unable to perform subgroup analysis for preterm versus term pregnancies and some studies reported insufficient data to determine whether thresholds for amniotic fluid measurement were adjusted for gestation. Where possible we used the results obtained

closest to delivery and have performed subgroup analysis for those where the test was performed within 7 days of delivery. In particular, there was very poor reporting regarding the exact methods of the reference standards and whether there was any treatment used between the performance of the index and reference standard. A major concern therefore is in how many pregnancies was induction of labour performed due to the finding of oligohydramnios, which influences the results for pregnancy outcome, i.e. intervention bias. This bias can only truly be removed by performing an RCT, this would be impossible to perform as measurements of amniotic fluid volume have become the standard in fetal surveillance and management of high-risk pregnancies and so recruitment to such a trial would be very difficult. Finally, the outcome measures used in this review were those that were reported by the authors of the included studies, it is recognised that many of the outcome measures are subjective (e.g. admission to neonatal intensive care unit, need for resuscitation). The only real objective measure of poor fetal outcome is paired samples of cord pH and longer-term outcomes such as cerebral palsy, which were not reported.”

표 1-22. 비тол림위험이 낮은 것으로 판단된 합성에 대한 등급부여 예시

신호 질문	등급부여	이유
a. 연구결과를 해석할 때 2단계에서 확인한 모든 우려 사항을 다루었는가?	그렇다	2단계 평가 중에 확인된 우려는 없다.
b. 문헌고찰의 연구질문에 대해 포함된 연구의 관련성이 적절히 고려되었는가?	아 마 도 그렇다	명시적으로는 아니지만, 개별 수준에서 문헌고찰 결과의 함의가 구체적으로 논의되었다. 포함된 연구에서 이용할 수 있는 인구집단과 증재결과의 측면에서 비тол림의 잠재적 관련성에 대해 세부적으로 논의되었다.
c. 평가자는 통계적인 유의성을 보이는 결과만을 강조하는 것을 피하였는가?	그렇다	문헌고찰 결론은 통계적으로 유의하거나 유의하지 않은 문헌고찰의 연구결과를 모두 반영하였다.
비тол림위험	낮음	2단계 평가에서는 문헌고찰 과정과 관련하여 어떠한 문제도 발견되지 않았다. 비тол림위험 측면에서 문헌고찰에 포함된 연구의 잠재적 한계는 논의에서 자세히 논의되었다. 결론은 문헌고찰의 연구결과를 적절하게 반영하였다.

2.4. ROBIS 사용에 있어 현실적인 이슈

체계적 문헌고찰에서 비풀림위험을 판단하는 것이 항상 간단한 것은 아니다. ROBIS는 적용 가능성에 대한 평가를 통해 비풀림위험을 평가하기 위한 구조화된 접근방식을 제공하도록 설계되었다. ROBIS는 다른 배경을 가진 검토자가 사용할 수 있도록 설계되었으며, 일부 방법론적, 내용적 전문지식이 요구될 가능성이 높다. 이상적으로 ROBIS 평가는 독립적인 두 명의 검토자에 의해 수행될 것을 추천하지만 적어도 한 명의 검토자가 평가를 점검한 상태에서 다른 검토자에 의해 확인할 수 있다. ROBIS 평가를 위해 모든 신호 질문을 고려할 고려할 것을 권고한다. ROBIS는 다양한 주제를 다루는 문헌고찰에 적용하기 위해 일반적으로 개발되었으며, 따라서 모든 신호 질문은 모든 문헌고찰 유형에 관련되어야 한다. ROBIS는 의사결정을 내리는 것을 포함한다. 우리는 검토자들에게 판단을 뒷받침하는 정보, 신호 질문, 전반적인 우려 사항 판단에 대한 근거 등을 기록하도록 요구함으로써 가능한 한 이 과정을 투명하게 하는 것을 목표로 했다.

ROBIS 평가에서는 모든 영역에 대한 모든 신호 질문을 평가해야 한다. 그러나 특정 상황에서는 특정 등급이 달성되면 “정지”하는 것이 적절할 수 있다. 이것은 평가 목적에 따라 달라진다. 만약 ROBIS의 목적이 비풀림위험의 평가와 비풀림의 이유를 제공하는 것이라면 완전한 평가가 필요할 것이다. 그러나 전반적으로 비풀림위험이 높은지 또는 특정 영역에 대한 우려가 있는지를 단순히 확인하는 것이 목적이라면, 비풀림의 높은 우려/위험이 확인되면 중단하기로 선택할 수 있다. 예를 들어 적절한 범위의 데이터베이스(신호 질문 2.1)를 검색하지 않은 경우, 검토는 이 영역에 대한 높은 우려로 고려될 수 있다. 비풀림위험에 있는지 여부에만 관심이 있는 평가자는 단순히 편향의 “높은” 위험의 등급을 부여하면서 이 영역 또는 다른 영역의 나머지 신호 질문을 평가하지 않을 수 있다.

PART 2

Cochrane RoB 2 : 원본 매뉴얼 번역

1. 서론	54
2. RoB 2 실행에 있어서 이슈	63
3. 상세한 지침: 사전 고려사항	65
4. 무작위 배정 과정에서 발생하는 비틀림	67
5. 의도한 중재에서 이탈로 인한 비틀림	81
6. 중재결과 자료의 결측으로 인한 비틀림	102
7. 중재결과 측정의 비틀림	116
8. 보고된 연구결과 선택의 비틀림	127

1. 서론

RoB 2 도구는 무작위 임상시험 결과에서 비뚤림위험을 고려하는 틀을 제공한다. 평가는 두 가지 중재 또는 중재 전략이 특정 결과에 미치는 상대적 영향을 추정하는 단일 임상시험 결과에 한정된다. 여기서 중재라 함은 임상시험(실험) 중재와 대조 중재를 의미한다. 물론 때때로 두 가지 활성(active) 중재의 비교를 의미할 수도 있다.

이 도구는 연구 결과(results)에 비뚤림이 적용될 수 있는 5개의 영역으로 구성되어 있다. 이들은 경험적 근거(박스 1-1 참조)와 이론적 고려사항 모두를 바탕으로 확인되었다. 영역은 무작위 배정 임상시험 결과에 영향을 미칠 수 있는 모든 유형의 비뚤림을 포함하기 때문에, 각각은 필수이며, 더 이상의 영역을 추가해서는 안 된다. 개별 무작위 임상시험(교차 임상시험 포함)을 위한 5개 영역은 다음과 같다.

- ① 무작위 배정 과정에서 발생하는 비뚤림
- ② 의도한 중재에서 이탈로 인한 비뚤림
- ③ 중재결과 자료의 결측으로 인한 비뚤림
- ④ 중재결과 측정의 비뚤림
- ⑤ 보고된 연구결과 선택의 비뚤림

영역 이름은 영역에서 다루어지는 비뚤림의 원인을 직접 설명하는 방식이다. 우리는 현 도구 버전 1에서 사용되는 많은 용어(예: 선택 비뚤림, 수행 비뚤림, 탈락 비뚤림, 결과 확인 비뚤림)를 사용하지 않기로 했는데 그 이유는 이들 용어들이 혼란을 야기할 수 있기 때문이다(1).

다음 연구설계에 비뚤림 영역을 다루는 몇 가지 템플릿을 제공한다.

- ① 무작위 평행군 임상시험
- ② 군집 무작위 평행군 임상시험(단계적 위지 설계 포함)
- ③ 무작위 배정 교차 임상시험 및 기타 유사한 설계

군집 무작위 임상시험의 경우, 추가 영역이 포함된다(1b) 여기에는 군집 내 개별 연구대상자의 확인 또는 모집에서 발생하는 비뚤림을 포함한다.

이 문서에서는 RoB 2 도구의 주요 특징을 설명하고 개별적으로 무작위 배정된 평행군 임

상시험에 적용할 수 있는 지침을 제공한다. 추가 문서는 군집 무작위 평행군 임상시험과 개별 무작위 교차 임상시험의 추가 고려사항을 다룬다. 우리는 아직 군집 교차 임상시험에 적합한 버전을 개발하지 않았다.

Box 2-1. 무작위 임상시험에서 비플립의 경험적 근거: 메타 역학(meta-epidemiology)의 역할

무작위 임상 시험에서 비플립에 대한 경험적 근거는 메타 역학으로 알려진 분야에서 나온다 (2). 메타 역학 연구는 연구의 방법론적 특성이 그 결과와 어떻게 관련되는지 이해하기 위해 이전 연구의 결과를 수집하여 분석한다. 첫 번째 잘 알려진 메타 역학 연구는 250개의 임상 시험이 포함된 33개의 메타 분석을 조사했다(3). 각 임상시험은 순서 생성이 무작위 방식으로 이루어 졌는지 여부, 배정을 적절하게 은폐하는 것으로 보고되었는지 여부, 이중 눈가림 로 기술되었는지 여부, 그리고 임상시험에서 연구대상자를 분석에서 제외하는 것으로 보고 했는지 여부 등 네 가지 특성에 따라 분류되었다. 4가지 특징 각각에 대해, 각 메타 분석 내에서 임상시험을 비교하여 ‘더 나은’ 임상시험과 ‘더 나쁜’ 임상시험 사이의 오즈비를 추정 하였고, 이러한 오즈비는 33개의 메타 분석에서 결합되었다. 그 후 수많은 유사한 연구가 수행되었고, 결과의 비플립과 관련이 있는 많은 연구 특성을 조사하였다. 최근의 분석은 조사 대상 특성과 관련된 비플립의 평균과 변동성을 모두 다룬다. 구체적으로, 또한 특성이 임상 시험간 이질성 증가와 관련되는 정도와 메타 분석 간 평균 비플립이 변화하는 정도를 조사 하였다(4). 본 문서의 여러 곳에서 우리는 메타 역학 연구 또는 그에 대한 체계적 문헌고찰에서 얻은 경험적 근거를 참조하여 영역의 선택과 신호 질문 선택을 지원하였다.

1.1 신호 질문

비플립의 각 영역 내에 신호 질문을 포함하는 것은 RoB 2의 핵심적인 특징이다. 신호 질문은 비플립위험 평가와 관련된 정보를 도출하는 것을 목표로 한다. 이러한 질문에 대한 응답에 따라 우리가 개발한 알고리즘을 이용하여 도구 사용자들이 비플립위험에 대해 판단하도록 유도한다.

신호 질문에 대한 응답 옵션은 다음과 같다.

- ① 그렇다(Y)
- ② 아마도 그렇다(PY)
- ③ 아마도 아니다(PN)

- ④ 아니다(N)
- ⑤ 정보없음(NI)

신호 질문의 단순성과 명확성을 극대화하기 위해, 'Y'의 응답은 질문을 하는 가장 자연스러운 방법에 따라 비뚤림위험이 낮거나 높음을 나타낼 수 있도록 표현된다.

'Y'와 'PY'의 반응은 'N'와 'PN'의 반응과 마찬가지로 비뚤림위험에 대해서도 같은 의미를 갖는다. 확정 응답('Y' 및 'N')은 일반적으로 신호 질문에 대해 확실한 근거가 이용가능함을 의미하며, 가능성 응답은 일반적으로 판단이 내려졌음을 의미할 것이다. 문헌고찰 저자가 신호 질문에 대한 응답에 대한 합의 척도(예: 카파 통계)를 계산할 경우, 우리는 'Y'와 'PY'를 동일한 응답으로, 'N'와 'PN'를 동일한 응답으로 취급할 것을 권고한다.

'NI' 응답은 (i) 'PY' 또는 'PN'으로 응답하기에는 불충분한 내용이 보고된 경우, (ii) 해당 내용이 없고, 임상시험 상황에서 'PY' 또는 'PN'으로 응답하기는 것이 합리적이지 않은 경우에만 국한해야 한다.

예를 들어, 경험이 많은 임상시험 기관에서 수행한 대규모 임상시험이라면 단어수 제한이 엄격하게 적용되는 저널에 게재된 논문에서 무작위 배정 순서의 생성에 대한 특정 정보가 없으면 신호질문에 대한 응답이 'NI'보다는 'PY'가 적절할 것이다. 순서 생성에 대한 신호 질문에 대한 응답이 'NI'인 경우 그 의미는 질문의 목적에 따라 달라질 수 있다. 만약 질문이 문제의 근거를 확인하는 것이면 'NI'은 그 문제에 대한 어떤 근거에도 해당하지 않는다. 문제가 보고될 것으로 예상되는 항목(예: 연구대상자가 추적관찰에서 탈락되었는지 여부)에서는 정보가 없으면 문제가 있다는 우려를 낳는다.

이전 질문에 대한 응답이 필요한 경우에만 응답하는 신호 질문의 경우, 응답 옵션으로 '해당되지 않음(NA)'을 사용할 수 있다. 신호 질문은 독립적으로 응답해야 한다. 즉, 한 질문에 대한 응답은 후속 질문 답변에 대한 결정을 제외하고는 동일하거나 다른 영역의 다른 질문에 대한 응답에 영향을 미쳐서는 안 된다.

1.1.1. 신호 질문과 함께 프리 텍스트 박스

이 도구에는 신호 질문과 함께 프리 텍스트를 위한 공간이 있다. 동일한 정보가 둘 이상의 질문에 응답하는 데 사용될 가능성이 있는 경우, 하나의 텍스트 박스는 둘 이상의 질문을 다룬다. 이 박스들은 각각의 신호 질문에 대한 응답을 지원하는 데 사용된다. 가능한 경우 연구 보고서 본문의 간단한 직접 인용문을 사용하면 된다.

1.2. 비뚤림위험 평가

1.2.1. 비뚤림위험에 대한 영역 수준 판단

RoB 2는 계층적으로 구상되었다: 신호 질문에 대한 응답은 일어난 일을 이끌어내고 비뚤림위험에 대한 영역 수준의 판단의 기초를 제공한다. 다시 말해서, 이러한 영역 수준 판단은 평가되는 특정 임상시험 결과에 대한 전체적인 비뚤림위험 판단에 대한 근거를 제공한다.

이 도구에는 각 영역에 대해 제안된 비뚤림위험 판단에 신호 질문에 대한 응답을 매핑하는 알고리즘이 있다. 비뚤림위험 판단은 다음과 같다.

- ① 낮은 위험
- ② 일부 우려
- ③ 높은 위험

‘판단’이라는 단어를 사용하는 것은 비뚤림위험성 평가에 중요하다. 특히 알고리즘은 제안된 판단을 제공하지만, 사용자는 이를 검증하고 이것이 필요하다면 변경할 수도 있다. 최종 판단에 도달할 때 다음과 같은 고려사항을 적용한다.

- ‘비뚤림위험’은 ‘실질적인 비뚤림위험’(‘risk of material bias’)으로 해석한다. 즉, 연구에서 신뢰할 수 있는 결론을 도출하는 능력에 영향을 미칠 가능성이 있는 문제에 대해서만 우려를 표시해야 한다.
- 비뚤림위험에 대한 영역 수준 판단은 결과의 신뢰도에 대한 비뚤림이 미치는 영향에 대해 6개 영역은 각각 동일한 의미를 가져야 한다. 어떤 개별 영역에 대한 ‘높은’ 비뚤림위험으로 판단되면 결과에는 전반적으로 ‘높은’ 비뚤림위험을 가진 것으로 판단되며, 어떤 개별 영역에 대한 ‘일부 우려’로 판단되면 전반적으로 ‘일부 우려’ 혹은 ‘높은’ 비뚤림위험으로 판단하게 될 것이다.(1.2.3 참조).

1.2.2. 비뚤림 방향

이 도구에는 각 영역과 전체 영역의 비뚤림 방향에 대한 판단이 있다. 일부 영역의 경우, 비뚤림이 무효점(null)을 향하는지, 무효점에서 벗어나는 지 쉽게 판단될 수 있다. 예를 들

어, 연구대상자가 배정된 중재에서 다른 중재로 전환되는 비율이 높은 경우, 중재 준수의 추정효과는 무효점을 향하는 쪽으로 비뚤림 방향이 발생할 수 있다. 다른 영역의 경우, 비뚤림이 비교되는 중재 중 하나에 더 많이 나타날 수 있는 경우 어떤 중재에 더 비뚤림이 나타나는지에 따라 효과 추정치는 증가 또는 감소할 수 있다. 예를 들면 무작위 배정 과정의 조작, 받은 중재가 무엇인지 아는 것이 결과 평가나 결과의 선택적 보고 등에 영향을 미치는 경우가 그런 경우이다. 만약 문헌고찰 저자가 비뚤림의 가능한 방향을 판단하는 명확한 근거를 가지고 있지 않다면, 굳이 추측할 필요는 없다.

1.2.3. 비뚤림위험에 대한 전반적인 판단에 도달한다.

전체적인 비뚤림위험 판단에 대한 반응 옵션은 개별 영역과 동일하다. 표 2-1은 중재 결과에 대한 영역 내의 비뚤림위험 판단을 전체 영역에 걸쳐 판단하는 기본 접근방식을 보여준다.

표 2-1. 특정 중재 결과에 대한 전반적인 비뚤림위험 평가 방법

전반적인 비뚤림위험 판단	기준
낮은 비뚤림위험	임상 시험에서 현재 결과의 모든 영역에 대해 낮은 비뚤림위험으로 판단됨
일부 우려	임상시험에서 적어도 한 개의 영역에서 일부 우려가 있었지만 어떠한 영역도 높은 비뚤림위험이 없음
높은 비뚤림위험	임상시험에서 결과의 적어도 한 영역에서 높은 비뚤림위험을 보였거나, 다수의 영역에서 일부 우려사항이 있었고 이러한 우려는 결과의 신뢰도를 상당히 낮출 것으로 예상함

결과가 개별 영역에 대한 비뚤림위험의 특정 수준에 있다고 판단하는 것은 그 결과가 적어도 이 정도로 심각한 비뚤림위험을 가지고 있다는 의미이다. 따라서 어떤 영역에서 비뚤림위험이 높다는 판단은 평가되는 영역에 관계없이 결과 전체에 대해 유사한 의미를 가져야 한다. 복수의 영역에서 ‘일부 우려’는 문헌고찰 저자가 결과 또는 결과군에 대한 전반적 비뚤림위험이 높다고 판단하게 할 수 있다.

1.2.4. 위험 평가와 함께 프리 텍스트 박스

판단 근거를 제시하기 위해 각 비뚫림위험 판단과 함께 프리 텍스트 공간이 있다. 제안된 알고리즘을 따르지 않는 판단을 할 경우 이에 대한 이유를 제시하는 것이 특히 중요하다.

1.3. 관심 효과 지정

‘의도한 중재로부터 이탈로 인한 비뚫림’ 영역에 대한 평가는 문헌고찰 저자가 관심을 기울이는 효과가 다음 중 어떤 것인가에 따라 이루어진다.

- ① 기저상태에서 이루어진 중재 배정 효과(중재가 추적관찰 기간 중 시행되었는지 여부에 관계없음, ‘intention-to-treat effect’ ITT 효과로 알려져 있음)
- ② 임상시험 프로토콜에 명시된대로 중재가 이루어졌는지에 대한 중재 준수 효과(‘per-protocol effect’ PP 효과라고도 한다)(5)

이러한 두 효과는 일부 환자가 배정된 중재를 받지 못하거나 기저상태 이후 배정된 중재에서 이탈되는 경우 달라질 것이다.

이 두 가지 효과 중 한가지가 관심 사항이 될 수 있다. 예를 들어, 중재 배정에 따른 추정 효과는 특정 보건의로 시스템에서 중재를 권고할 것인지 여부(예: 선별검사 프로그램을 도입할 것인지 또는 새로운 콜레스테롤 강하제를 처방할 것인지)에 대한 보건 정책 의사결정에 정보를 제공하는데 가장 적절할 수 있다. 임상시험 프로토콜 명시된 중재에 대한 준수 효과는 개별 환자의 치료 의사결정(예: 선별검사 수검 여부 또는 신약 복용 여부)에 더욱 직접적으로 영향을 미칠 수 있다.

문헌고찰 저자는 관심 있는 중재 효과를 정의하고, 이 효과에 대해 비뚫림위험 도구를 적절히 적용해야 한다. 중재 준수 효과를 평가할 때, 문헌고찰 저자는 어떤 유형의 의도한 중재로부터 이탈을 조사할 것인지 지정해야 한다. 이러한 이탈에는 다음과 같은 것이 있다 : (i) 비프로토콜 중재(non-protocol intervention)의 발생, (ii) 결과에 영향을 미칠 수 있는 중재 시행 실패(failures in implementing the intervention), (iii) 임상시험 배정 중재에 대한 비준수(non-adherence)(제5.3절 참조). 예를 들어, START 무작위 임상시험은 HIV 양성자의 항레트로바이러스 치료(ART)의 즉각 투여와 연기 투여를 비교하였다. 연기 투여

배정자 중 30%가 지정된 프로토콜보다 일찍 ART를 시작했다(6). Lodi 등은 이러한 프로토콜 이탈에 대해 보정된 중재 준수 효과를 추정했지만, 연구대상자들이 임상시험 추적관찰 내내 항레트로바이러스 치료를 계속했는지 여부는 추정하지 않았다(7). 그러한 예에서, 문헌고찰 저자는 배정 중재에 대한 비준수는 발생하지 않았지만 비프로토콜 중재가 발생했고 이를 비풀림위험 평가에서 다루었다라고 명시할 수 있다.

주요 관심 효과는 문헌고찰 프로토콜에 명시되어야 한다. 때때로, 문헌고찰 저자는 두 가지 관심 효과 모두에 관심이 있을 수 있다.

RoB 2에 대한 ‘관심 효과’를 지정하는 것이 치료 효과 측정지표(odds ratio, risk difference 등)의 선택과 관련이 없다는 점에 유의해야 한다.

1.3.1 관심 효과 추정

무작위 임상시험 저자는 중재 효과를 추정하기 위해 몇 가지 분석적 접근법을 사용할 수 있다. 그들이 왜 그런 분석 접근법을 선택하였는지, 그들의 목적이 배정효과인지 중재 준수 효과인지 설명하지 않을 수 있다. 우리는 분석에 대한 다른 접근법과 그것이 가지는 비풀림에 대한 영향에 대해 논한다. 복수의 분석이 보고될 수 있기 때문에, 문헌고찰 저자가 중재에 대한 배정 효과에 관심이 있을 때 중재의 추정 효과를 선호도에 따라 선택할 것을 제안한다.

중재에 대한 배정 효과는 모든 무작위 연구대상자를 포함하는 ‘배정된 대로(ITT) 분석’에 의해 추정되어야 한다(8). 배정된 대로 분석의 원칙은 아래와 같다(9, 10):

- ① 실제 받은 중재에 관계없이 무작위로 배정된 연구대상자를 분석한다. 그리고
- ② 분석에는 모든 무작위 배정된 연구대상자를 포함하며, 모든 연구대상자의 결과를 측정해야 한다.

‘배정된 대로 분석’은 측정 또는 측정되지 않은 예측 요인과 관련하여 평균적으로 중재군이 기저상태에서 다르지 않은 무작위 배정의 이점을 유지한다. ‘배정된 대로 분석’이라는 용어는 일관된 정의를 가지고 있지 않으며, 연구 보고서에서 일관되지 않게 사용된다(11-13).

배정된 중재를 준수하지 않는 눈가림 위약 대조 임상시험에서, ‘배정된 대로 분석’은 대상자가 프로토콜을 준수했을 경우만 분석하는 ‘계획서 순응군 분석’(PP 분석)에 비해 중재 효과를 과소평가할 것으로 예상되는데 이는 동등성 임상시험이나 비열등 임상시험에서 특히

문제가 된다. 그러나 중재 배정 효과 추정치는 두 개 이상의 활성 중재를 비교하는 경우나 다른 연구대상자에 대해 서로 다른 방향을 비교하는 경우 과대평가하게 되며(14, 15), 이것도 동등성 임상시험이나 비열등 임상시험에서 문제가 된다(14). 모든 연구대상자가 중재에 준수하는 경우 나타나는 효과 과소평가는 중재군의 위해(부작용)를 조사할 때 특히 문제가 될 수 있다. 배정된 중재에 대한 비준수의 변이도 중재 효과의 ITT 추정에서 이질성의 원천이 될 수 있다. 즉, 연구대상자의 상당 부분이 준수하지 않은 경우에 비해 완벽한 준수를 보이는 경우 효과성이 더 커진다(5).

환자와 다른 이해당사자들은 종종 임상시험 프로토콜에 기술된 대로 중재에 준수하는 효과(중재 준수 효과)에 관심이 있다. 무작위 배정 임상시험의 자료를 사용하여 중재에 따른 효과에 대한 편향적이지 않은 추정치를 도출하는 것은 가능하지만, 적절한 방법은 강력한 가정을 필요로 하며 실제로 적용된 경우는 현재까지 비교적 드물다(5). 중요한 것은 일반적으로 준수 정도에 대한 보정에 사용되는 접근방식은 개인에게 배정된 중재를 받을지 여부에 영향을 미칠 수 있는 예측 요인에 대한 보정을 포함하지 않기 때문에 중재 준수 효과의 추정에 편향을 초래할 수 있다(14). 여기에는 다음과 같은 것이 포함된다.

- naïve ‘per protocol’ analyses : 배정된 중재에 준수한 중재군내 연구대상자만 분석
- ‘as-treated’ analyses : 배정군이 다른 경우에도, 연구대상자가 실제로 받은 중재에 따라 분석

임상시험 연구자들은 종종 두 가지 이상의 접근법을 사용하여 중재의 효과를 추정한다. 우리는 관심 효과가 중재에 대한 배정 효과일 경우 메타 분석에 포함하고, 비뚤림위험 평가 결과는 다음 우선순위에 따라 선택하도록 권고한다.

- ① 위에서 정의한 완전한 ‘배정된 대로 분석’에 해당하는 결과
- ② 결과 자료가 누락된 연구대상자를 제외한 것 이외에는 ITT 원칙을 준수하는 ‘수정된 ITT’ 분석(mITT)에 해당하는 결과(섹션 5.3.1 참조). 이러한 분석은 결과자료 누락으로 인한 비뚤림을 방지하지 못하며, 이는 「결측치로 인한 비뚤림위험 영역」에서 다루어진다.
- ③ ‘치료받은 대로’ 혹은 ‘계획서 순응군’ 분석에 해당하는 결과나 적격한 임상시험 연구 대상자가 제외된 분석

중재 준수 효과를 타당하게 추정하기 위해서는 의도한 중재로부터 이탈이 어떤 것인지에 대한 자료가 필요하고, 의도한 중재로부터 이탈을 예측하는 예측 요인에 대한 보정이 필요하다. 추적관찰 중 이러한 요인을 측정하는 데 집중하는 것이 향후 프로토콜 효과의 더 빈번한 추정을 용이하게 할 수 있다.

시간 경과에 따라 지속되는 중재를 비교하는 시험에서 유효한 추정은 일반적으로 예측 요인의 무작위 사전, 사후 값에 대한 적절한 보정이 필요할 것이다. 표준 회귀 모형과 같은 기존 통계 방법은 사전 중재에 의해 무작위 사후 예측 변수가 영향을 받는 경우 유효하지 않기 때문에 역확률 가중치(inverse probability weighting)와 g-formula와 같은 ‘g-methods’를 일반적으로 사용해야 한다(5, 14, 15).

기저상태에만 시행되고 완전 준수 혹은 비준수 중재가 있는 임상 시험의 경우 무작위 상태를 도구 변수(instrumental variable)로 활용하는 방법을 사용하여 프로토콜 효과를 추정할 수 있다. 도구변수 방법에는 준수에 대한 자료가 필요하고 강력한 가정에 기초해야 하지만(5), 이러한 가정이 충족될 경우 중재를 받았는지를 예측하는 예측 요인에 대한 보정 필요성을 우회할 수 있다. 예를 들어 유연 S자형 내시경 선별 검사가 대장암 발병률과 사망률에 미치는 영향에 대한 무작위 임상시험에서 1차로 배정된 대로 분석을 시행하고 2차적으로 도구 변수 분석을 보고하고 비준수에 대해 보정한 효과 측정치를 보고하였다(16).

각각의 관심 각 효과에 대해, ‘의도한 중재로부터 이탈로 인한 비뚤림’ 영역의 신호 질문은 해당 효과를 추정하기 위해 적절한 통계적 방법을 사용했는지를 묻는다.

2. RoB 2 실행에 있어서 이슈

2.1. 다중 평가

일반적으로 임상시험은 여러 결과를 만들어 내며, 여러 중재결과에 기여함으로써 체계적인 문헌고찰이 가능해 진다. 따라서 각 연구에 대해 몇 가지 비돌림위험 평가가 필요할 수 있다. 우리는 어떤 결과를 평가의 대상으로 삼아야 하는지, 얼마나 많은 연구 결과(results)를 평가해야 하는지에 대한 권고안을 아직 공식화하지 않았다. 그러나 이러한 결정은 결과 요약 표(Summary of Findings)에 포함된 중재 결과와 일치할 가능성이 있다.

2.2. 자료수집 프로세스

비돌림위험 평가는 특정 시점에 측정된 특정 중재 결과에 대한 특정 연구 결과에 대해 수행된다. 그러나 비돌림의 일부 영역(예: 무작위 과정에서 발생하는 비돌림)은 전체 연구에 적용되며, 어떤 영역(예: 의도한 중재로부터 이탈로 인한 비돌림)은 주로 측정하는 중재 결과에 적용되며, 어떤 영역(예: 결과 측정 비돌림)은 주로 사용된 중재 결과 측정 방법에 적용된다. 또한 어떤 영역(예: 선택적 결과 보고 비돌림)은 특정 연구결과에 적용한다. 이는 문헌고찰 저자가 연구 보고서로부터 비돌림위험에 관련된 정보를 가장 효율적으로 추출할 수 있는 방법이다.

2.3. 비돌림위험 평가의 제시

우리는 RoB 2 평가 결과를 다음과 같이 제시할 것을 제안한다. 이 부분에 대해서는 더 많은 작업이 필요하다.

- 과정의 완전한 투명성을 위해 문헌고찰 저자는 각 평가자별로 응답, 프리 텍스트 내용, 판단을 별도로 제시하기를 원할 수 있다. 하지만 이렇게 하면 독자들에게 혼란을 줄 수 있기 때문에, 우리는 이를 부록이나 보충 문서에 포함할 것을 권고한다.
- 주요 문헌고찰 문서에 영역 수준 판단(예: 표 또는 그림 또는 결과의 숲 그림으로)을 제

시한다. 복수의 평가자인 경우 합의된 판단만 제시해야 한다. 지면이 허용하는 경우, 각 판단에 대해 요약된 프리 텍스트를 통한 정당화는 주요 문헌고찰 문서 내에서 이에 대한 매력적인 보완이 될 것이다.

- 부록 또는 보충 문서에서 각 신호 질문에 대한 응답과 각 응답에 대한 프리 텍스트 지원을 제공한다. 복수의 평가자에 대한 합의된 응답만 제시하도록 한다.

2.4. 신속 평가

영역 중 하나 이상에 대해 '높은' 비뚤림위험으로 판단될 경우, 해당 결과에 대한 전반적 비뚤림위험은 '높음'이기 때문에, 도구 사용자는 한 영역이 '높음' 로 판단되면, 평가를 중지하고 싶을 수 있다. 하지만 우리는 다음과 같은 이유로 표준적 체계적 문헌고찰의 맥락에서 사용될 경우 그렇게 하는 것에 대해 반대한다.

첫째, 체계적인 문헌고찰의 많은 독자들은 투명성을 위해 포함된 근거에 대한 완전하고 일관된 평가를 보는 것을 선호한다. 둘째, 기존의 무작위 임상시험의 한계에 대한 완전한 평가를 통해 해당 증재에 대한 향후 임상시험 설계와 수행에 유용할 가능성이 있다. 셋째, 문헌고찰 저자에 의해 생산된 임상 시험에 대한 비뚤림위험 평가를 임상시험 결과와 함께 공개적으로 이용할 수 있도록 하려는 연구 커뮤니티의 움직임이 있다. 이를 위해서는 완전한 형태로 문서화된 영역 수준의 평가가 필요하다. 조금 사소한 고려사항이기는 하지만 메타역학 연구의 경우 다양한 메타 분석을 재분석하여 임상시험 디자인 특성의 영향에 대해 들여다 본 후 연구의 제한점 때문에 생기는 비뚤림의 크기와 방향에 대한 정보의 귀중한 출처가 되며 이 연구는 각 영역에 대한 완전한 평가를 필요로 한다.

평가의 목적이 현 논문이 비뚤림위험이 '높음'인지만을 신속하게 확인하는 것이라면 평가에 '정지 규칙'을 도입할 필요가 있을 수 있다. 비뚤림위험 '높음'으로 판단되는 임상시험은 근거의 종합에 아무런 역할을 하지 않는다는 것이 프로토콜에 미리 명시되어 있을 때에만 이 조치를 취할 것을 권고한다. 민감도 분석 또는 하위집단 분석에 임상시험을 포함하려면 연구가 적절히 특징지어질 수 있도록 전체 평가를 수행할 것을 권고한다.

3. 상세한 지침: 사전 고려사항

비뚤림위험 평가를 완료하기 전에, 평가 대상의 설계, 평가할 중재 결과(평가되는 특정 연구 결과뿐만 아니라), 관심 효과가 배정효과인지 중재 준수효과 같은 평가의 중요한 특성을 문서화하는 것이 도움이 된다. 문헌고찰 저자는 평가를 완료하는 데 사용되는 출처를 문서화해야 한다(실무적으로 가능한 많은 출처를 사용해야 한다). RoB 2 표준 템플릿에는 이러한 세부 사항을 기록하고(box 2-2), 어떤 중재가 '임상시험 중재'인지, 어떤 중재가 비교 중재인지를 명확하게 하기 위한 질문이 포함되어 있다.

Box 2-2. ROB 2 도구(제 1부). 사전 고려사항

연구설계

- 개별 무작위배정 평행군 임상시험
- 군집 무작위배정 평행군 임상시험
- 개별 무작위배정 교차(혹은 other matched) 임상시험

평가 목적의 관점에서 중재와 비교중재의 정의는?

중재 비교중재

비뚤림위험 평가를 시행하는 중재결과는 구체적으로 무엇인가?

평가할 수치형 연구 결과는 구체적으로 무엇인가? 여러 분석이 시행된 경우 수치형 연구 결과가 구체적으로 무엇인지(예 : RR = 1.52 (95% CI 0.83 to 2.77)), 평가 결과를 어디에서 얻었는지 명시할 것(표, 그림, 문장 등)

현 결과에 대한 평가 목적은 무엇인가?

- 중재 배정 효과 평가(ITT 효과)
- 중재 준수 효과 평가(PP 효과)

평가 목적이 중재 준수 효과라면, 다루어야 할 중재 이탈의 종류는 무엇인가

(최소한 한 개는 체크해야 함)

- 비프로토콜 중재의 발생
- 중재결과에 영향을 미칠 수 있는 중재 실행 실패
- 연구대상자가 배정된 중재 비준수

비독립위험 평가 정보는 어디에서 얻었나? (모두 체크하십시오)

- 학술지 논문
- 임상시험 프로토콜
- 통계분석 계획(Statistical analysis plan, SAP)
- 비상업적 임상시험 등록처 자료(예: ClinicalTrials.gov 자료)
- 회사 소유 임상시험 등록처 자료(예: GSK Clinical Study Register record)
- 회색 문헌 (예: 출판되지 않은 학위 논문)
- 학회 초록
- 규제 문서(예: Clinical Study Report, Drug Approval Package)
- IRB 지원 서류
- 연구비 지원 자료베이스 요약(예: NIH RePORTER 혹은 Research Councils UK Gateway to Research)
- 임상시험 시행자와 사적 교신
- 스폰서와 사적 교신

4. 무작위 배정 과정에서 발생하는 비뚤림

4.1. 배경

성공적으로 달성된 경우, 무작위 배정은 중재군 배정에 대한 알려진 또는 알려지지 않은 예측 요인(질환의 심각성 또는 동반질환과 같은 결과를 예측하는 요인)의 영향을 없앨 수 있다. 이는 평균적으로 중재 집단이 중재가 시작되기 전에 동일한 예후를 갖는다는 것을 의미한다. 예측 요인이 연구대상자가 배정된 중재군에 영향을 미치는 경우, 중재의 추정 효과는 ‘교란’에 의해 편향되며, 교란은 중재군 배정과 결과에 공통적인 원인이 될 때 발생한다. 통상 진료에서 치료 결정은 종종 예측 요인에 의해 영향을 받기 때문에 교란 요인은 관찰 연구의 중재 효과 추정치에 비뚤림을 초래하는 중요한 잠재적 원인이다.

연구대상자를 무작위로 배정하기 위해 연구대상자가 중재에 배정되는 방법을 지정하는 배정 순서는 우연의 요소를 포함하는 프로세스에 기반하여 생성된다. 우리는 이 프로세스를 **배정 순서 생성**이라고 부른다. 그 후, 모집 확정 후까지 연구대상자나 임상시험 담당자가 향후 배정을 알지 못하도록 조치를 취해야 한다. 이 과정을 흔히 **배정 순서 은폐**이라고 한다.

다음 배정에 대한 지식(예: 순서가 게시판에 공개적으로 게시된 경우)이 있으면 예측 요인에 기초하여 연구대상자를 선별적으로 등록할 수 있다. ‘부적절한’ 것으로 간주되는 중재에 배정된 연구대상자는 참여를 거부할 수 있다. 이것은 선택 비뚤림의 한 유형이다. 다른 연구대상자는 ‘적절한’ 중재를 원할 수 있으며, 이는 원하는 배분이 나타날 때까지 임상시험 진입을 지연시킴으로써 이루어질 수 있다. 역학적으로 볼 때, 배정된 중재에 대한 그러한 조작용은 교란 요인을 초래할 수 있다. 이러한 이유로, 성공적인 배정 순서는 무작위의 필수적인 부분이다.

배정은폐와 임상시험 과정에서 이루어지는 눈가림을 혼동해서는 안 된다. 배정 은폐는 임상시험 담당자와 연구대상자가 배정 전과 배정 시에 배정 순서를 알지 못하도록 하여 중재 배정에 의한 비뚤림을 방지하는 과정이다. 연구설계나 임상 영역에 관계없이 항상 성공적으로 구현할 수 있다(18, 19). 이와는 대조적으로 눈가림(연구대상자, 중재 제공자 또는 결과 평가자)은 무작위 이후에도 배정된 중재를 계속 은폐함으로써 무작위 배정에 따른 비뚤림을 방지하고자 하며, 항상 구현될 수 있는 것은 아니다(19, 20). 예를 들어 수술과 비수술적 중재를 비교하는 임상시험에서 흔히 이런 상황이 발생한다. 중재의 배정 시점까지와 그 시점 이후의 눈가림(blinding)은 서로 다른 비뚤림의 원천을 다루고 그 실현가능성에 차이가 있

지만, 그럼에도 불구하고, 배정 시점의 연구대상자와 직원에게 배정이 은폐되지 않으면 이후에도 눈가림은 이루어지지 않았다고 볼 수 있다.

4.1.1. 순서생성에 대한 접근 방법

계약조건이 없는 무작위를 단순 무작위 배정 또는 무제한 무작위 배정이라고 한다. 때로는 블록 무작위 배정(제한적 무작위 배정)을 사용하여 임상시험 중재 및 대조 중재군에서 원하는 연구대상자의 비율(예: 1:1)을 달성하도록 순서를 생성한다(21, 22). 이는 각 중재군에 배정된 연구대상자 수가 지정된 크기의 블록 내에서 균형을 이루도록 함으로써 이루어진다(예: 연속적으로 참여하는 10명). 중재군과 대조군에 대한 지정된 배정 수는 각 블록 내에서 무작위 순서로 배정된다. 블록 크기가 임상시험 담당자에게 알려지면 각 블록 내의 마지막 배정을 항상 예측할 수 있다. 이 문제를 피하기 위해 블록 크기를 다양하게 하고 무작위로 다양한 블록을 사용할 수 있다.

층화 무작위는 제한적 무작위 배정으로 질병 심각도 및 연구 센터와 같은 잠재적으로 중요한 예측 요인에 의해 정의된 연구대상자의 하위 집합 내에서 별도로 수행되는 무작위법으로 일반적으로 많이 사용된다. 각 층(strata)에서 단순(제한된 것이 아니라) 무작위 배정을 사용하는 경우, 계층화는 아무런 이점도 제공하지 않지만, 무작위 배정은 여전히 유효하다.

최소화법은 계층화와 제한된 무작위 배정을 모두 포함하는 특정 예측 인자와 관련하여 중재군을 밀접하게 유사하게 만드는 데 사용될 수 있다. 최소화에는 일반적으로 무작위 요소(최소 알고리즘에 포함된 예측 요인과 관련하여 군이 균형을 이루도록 연구대상자를 참여시킴)를 포함한다.

때때로 사용되는 다른 유형의 무작위법으로 편중된 동전 또는 urn 무작위 배정, 대체 무작위 배정, 혼합 무작위 배정, 최대 무작위 배정 등이 있다(21, 23, 24). 이러한 접근법이나 다른 접근법을 접할 경우 방법론 전문가와 협의가 필요할 수 있다.

4.1.2. 배정 은폐 및 무작위 배정 실패

배정 순서가 적절하게 생성되는 경우에도 다음 배정에 대해 알게 되면 예측 요인에 근거하여 연구대상자를 선별적으로 등록할 수 있다. 연구대상자가 부적절한 것으로 생각하는 중재에 배정되면 참여를 거부하거나 원하는 배정이 될 때까지 임상시험 진입을 늦춤으로써 '적절한' 중재를 받도록 조작할 수 있다. 이러한 이유로, 성공적인 배정 순서 은폐는 무작위

의 필수적인 부분이다.

향후 배정을 예상하여 배정 은폐의 실패로 이어질 수 있는 방법에는 다음과 같은 것이 있다.

- ① 교대, 생일 또는 내원일과 같은 결정론적 배정 규칙을 알게 되는 것
- ② 배정 순서에 대한 지식(예: 무작위 배정 순서가 게시판에 공개적으로 게시된 경우)
- ③ 이전 배정을 기준으로 배정을 성공적으로 예측할 수 있는 것

이 중 ③은 블록 무작위법에서 발생할 수 있는데 각 연구대상자가 임상시험에 등록한 후 임상시험 모집자에게 배정이 알려진 경우 발생할 수 있다. 특히 블록의 크기가 고정되어 있고 다기관이 아닌 경우, 향후 배정을 예측할 수 있을 것이다(25).

최소화를 사용하게 되면 배정을 예측할 수 있는 위험 때문에 일부 방법론자들은 이 접근 방식을 허용하는 것에 대해 조심스러워 하는 반면, 다른 이들은 특히 단순한 무작위 배정을 사용할 경우 기준 특성의 상당한 불균형이 우연히 발생할 수 있는 소규모 임상시험의 경우 매력적이라고 생각한다(26, 27). 이러한 위험을 완화하기 위해 최소화 접근법을 단순한 무작위 배정과 결합하는 경우가 많으므로, (예를 들어) 전체 배정 중 80%는 최소화에 의해 나머지 20%는 단순 무작위법으로 할 수 있다. 최소화 기반 전략을 사용할 때 배정 은폐는 최소화가 각 센터를 통틀어 시행된다면 다기관에서 시행될 때 더 많이 보호된다.

배정 은폐 노력이 실제로 훼손될 수 있다. 예를 들어, 밀봉되지 않은 배정 봉투는 열 수 있고, 반투명 봉투는 밝은 빛에 비추어 내용물을 알아낼 수 있다(3, 28, 29). 실제로 은폐 방법이 불충분했기 때문에 많은 배정 체계가 연구자들에 의해 해독되었다고 한다(28).

순서 생성 및 배정 은폐 방법에 대한 정보는 대개 임상시험 프로토콜에서 확인할 수 있지만, 불행히도 출판물에서는 충분히 보고되지 않는 경우가 많다(27). 예를 들어, 무작위 임상시험 보고의 완전성에 대한 코크란 리뷰에서 CONSORT를 채택하는 학술지의 경우 무작위 임상시험의 45%(393/876)에서 채택하지 않는 학술지의 경우 무작위 임상시험의 22%(329/1520)에서만 배정 은폐가 적절히 보고되었다. 이것은 학술지의 제한된 단어 수 때문일 수 있으며 다시 한번 정보원의 중요성을 강조할 수 있다. 학술지 논문에서 무작위 배정 및 배정 은폐 방법에 대한 설명이 부족하다고 해서 사용된 방법이 반드시 부적절한 것은 아니다. 즉, 단어 수 제한, 심사 또는 편집 권고 사항 때문에 세부 사항이 누락되었을 수 있다(31)

비교 가능한 군을 생성하는 데 있어서 무작위법의 성공은 종종 중재군들 사이의 중요한

예측 요인의 기저상태 값을 비교함으로써 조사할 수 있다. 무작위 방법이 과소 보고되는 성향과는 대조적으로, 기준 특성은 CONSORT를 채용하는 학술지에 발표된 RCT의 95%, 비채용 학술지의 87%에 보고되었다(30). Corbett 등은 연구대상자 특성이 중재군 간에 균형을 이루는지 여부를 고려해야 한다고 주장했다(32). RoB 2는 기저상태 불균형이 무작위 프로세스에 문제가 있는지를 판단하는 신호 질문을 포함한다(4.3.3 참조).

4.2. 무작위 배정 과정에서 발생하는 비뚤림의 경험적 근거

7개 메타 역학 연구의 최근 메타 분석 결과, 순서 생성이 부적절하거나 불분명한 경우 적절한 방법인 경우에 비해 중재 효과 추정치를 미미한 정도로 과다 추정(7%)하였다(33). 예외로, 주관적인 결과를 보고하는 임상시험에서 비뚤림이 더 컸다. 즉, 모든 원인의 사망률과 기타 객관적인 결과를 평가하는 임상시험에서 비뚤림에 대한 근거는 거의 없었다.

마찬가지로, 배정은폐가 부적절하게 혹은 불확실하게 이루어진 임상시험의 경우 중재 효과 추정치에 대한 경미한 과장(10%)이 관찰되었다. 부적절한 배정 은폐와 관련된 평균 비뚤림은 주관적 결과를 보고하는 임상시험과 대체요법 임상시험에서 가장 컸으며 사망률이나 객관적인 중재결과를 보고하는 임상시험에서는 비뚤림에 대한 근거는 거의 없었다.

기저상태 불균형에 대한 근거는 더 이상하다. 세 가지 경험적 연구에서 기저상태 불균형이 중재 효과 추정치를 부풀린다는 근거는 발견되지 않았지만(34-36), 모든 추정치는 정밀하지 않았으며 무작위 배정 방법이 과장된 중재 효과 추정치와 관련이 있다는 근거는 발견되지 않았다. 교란 변수를 보정하지 않은 임상시험(34)이나 블릭 무작위법을 사용한 비눈가림 임상시험에서 중재 효과 추정치가 과장되었다는 근거도 거의 없었다(35). 그러나 각각의 특성은 작은 연구에서만 조사되었다.

4.3. 현재 영역 사용의 실제

4.3.1. 무작위 순서 생성 평가

적절한 순서생성을 위해서는 무작위 구성요소의 사용이 충분해야 한다.

반복적인 동전 던지기, 주사위 던지기, 이전에 뽑았던 카드를 잘 처리하는 것, 난수 목록

출판물(21, 22)을 참조함으로써, 원칙적으로 간단한 무작위 배정을 달성할 수 있다. 일반적으로 무작위 배정 목록은 컴퓨터에 의해 생성된다. 비뿔림위험은 임상시험에서 무작위 배정을 계층화했다고 주장하든 그렇지 않은 같은 방식으로 판단할 수 있다.

무작위 순서생성 예: “우리는 단순 무작위법으로, 동일한 배정 비율 사용하고, 난수표를 써서 두 개의 비교군을 생성했다.”

무작위 순서생성 예: “블록 무작위법을 사용하여 두 개의 비교군에 대한 배정 목록을 구성했다. 컴퓨터 무작위 번호 발생기를 이용해 블록 크기가 8이고 배정 비율이 같은 임의의 *permuted blocks*를 사용하였다”

교대, 생년월일에 기초한 배정, 사례 기록 번호 및 발표 날짜와 같은 체계적인 방법은, 때로는 “준 무작위”라고 일컬어지며, 순서생성의 부적절한 방법이다. 교대(또는 2개 이상의 중재군에 대해 회전)는 원칙적으로 유사한 군을 만들 수 있지만, 다른 많은 체계적인 순서생성 방법은 그렇지 않을 수 있다. 예를 들어, 환자입원일은 단지 우연의 문제가 아니다. 모든 체계적인 방법의 중요한 약점은 배정 일정을 숨기는 것이 대개 불가능하다는 것이다. 이는 연구에 참여하는 사람들 사이에서 중재 배정을 미리 알아 배정에 비뿔림이 생길 수 있다.

비무작위 순서생성 예: “환자들은 성씨의 첫 글자에 의해 무작위로 배정되었다.”(37)

비무작위 순서생성 예: “작수날 날짜에 태어난 사람은 실내 공기(룸 에어군)로 배정했고, 홀수 날짜에 태어난 사람은 100% 산소군(산소군)에 배정하였다.”

4.3.1.1 사용된 방법에 대한 정보가 부족할 경우 순서생성 평가

“무작위로 배정했다” 또는 “무작위법 사용”과 같은 간단한 문구는 종종 배정 순서가 진정으로 무작위 배정되었다는 것을 확신하기에 불충분하다. 실제로, 저자들은 그것이 정당화되지 않은 경우에도 “무작위”라는 용어를 사용하는 것이 일반적이다. 체계적 배정이 된 임상시험에서조차 일부 저자들은 “무작위”라는 용어를 사용한다. 어떤 상황에서는 무작위법을 사용했는지에 대해 합리적인 판단이 내려질 수 있다. 예를 들어, 숙련된 임상시험 기관에서 운영하는 대규모 임상시험의 경우, 단어 수 제한이 엄격하게 적용되는 학술지에 게재된 논문에서 무작위 배정 순서생성에 대한 특정 정보가 없으면 ‘정보 없음’이 아니라 ‘아마도 예’라는 응답이 더 적절할 수 있다. 또는, 동일한 조사팀의 다른 임상시험에서 명백히 무작위법이 사용되지 않은 경우, 현재의 연구가 유사한 방법을 사용하여 이루어졌다고 가정하고, 신호 질문에 ‘아마도 아니다’라고 대답하는 것이 합리적일 수 있다. 도구 사용자가 이러한 판단을 할 수 없는 경우, ‘정보 없음’에 체크해야 한다.

임상시험 연구자들은 무작위 구성요소가 있다는 것을 확인하지 않고 순서생성에 대한 그

들의 접근방식을 불완전하게 설명할 수 있다. 예를 들어, 저자들은 블록 배정 순서가 무작위라는 명시적 진술 없이 블록 배정을 사용했다고 진술할 수 있다. 이러한 경우 일반적으로 '정보 없음'으로 응답해야 한다.

4.3.2. 배정 순서의 은폐 평가

배정 은폐를 위해 사용되는 방법 중 제3자에 의한 중앙 무작위 배정이 가장 바람직하다. 봉투를 사용하는 방법은 다른 접근법보다 조작에 더 취약하다(18, 27). 연구자들이 봉투를 사용할 경우 은폐를 보존하기 위해 배정 과정을 개발하고 감시해야 한다. 순차적으로 번호가 매겨진 불투명한 봉인 봉투를 사용하는 것 외에 봉투가 순차적으로 열렸는지, 그리고 봉투가 연구대상자에게 돌이킬 수 없을 정도로 배정된 후에만 열렸는지 확인해야 한다. 블록법이 사용된 경우 각 블록 내의 마지막 중재 배정을 예측할 수 있다. 이것은 모집 연구대상자가 각 블록의 시작과 끝을 알고 배정 내용을 배정 후 공개할 때 문제가 될 것이다. 블록 크기가 작고 크기가 같다면 문제가 더 심각할 가능성이 있다. 그러한 경우, 배정을 은폐했는지 여부에 관한 신호 질문에 대해 '아니오' 또는 '아마도 아니오'로 응답해야 한다.

표 2-2. 배정순서 은폐에 대해 판단하는 최소 기준, 확장 기준

최소 기준	확장 기준
중앙 무작위법	중앙의 무작위 배정 센터는 환자 모집 센터에서 멀리 떨어져 있었다. 연구대상자의 세부사항은 전화(상호작용하는 음성 응답 시스템 포함), 이메일 또는 상호작용하는 온라인 시스템으로 제공되었고, 연구대상자가 최종 등록될 때까지 배정 순서를 숨겼다.
연속번호가 부여된 약용기	독립적인 약사가 준비한 약용기에 연속 번호를 매기고 순차적으로 개봉했다. 약용기는 외관이 같았고, 조작이 불가능했으며(tamper-proof) 무게도 같았다.
연속번호가 부여된 밀봉, 불투명 봉투	봉투에 연구대상자 세부사항을 적은 후에만 봉투에 연속 번호를 부여하고 순차적으로 개봉했다. 봉투 안에 압력을 감지하는 또는 탄소 용지가 연구대상자의 세부 정보를 배정 카드로 전송했다. 봉투 안의 판지나 알루미늄 포일이 강렬한 빛에 비칠 수 없게 만들었다. 봉투는 변조 방지 보안 테이프로 봉인되었다.

Box 2-3. 적절한 배정 순서 은폐의 예(Schulz and Grimes(39)에서 편집)

“...코드가 부여된 숫자와 약물 배정을 결합하였다. 열 개의 숫자로 이루어진 각 블록은 중앙 센터에서 각 센터의 무작위 담당자에게 전송되었다. 이 사람(대상 환자 관리에 관여하지 않고 연구자와 독립적인 약사 또는 간호사)은 약물 주입 배정, 준비 및 회계를 담당했다. 약물 주입은 별도의 장소에서 준비한 후 24시간마다 병상 간호사에게 이송되었다. 간호사는 적절한 비율로 환자에게 주입했다. 이에 따라 무작위 배정 일정은 모든 간호 제공자, 병동의 사 및 기타 연구 인력으로부터 은폐되었다.”(40)

“...연속 번호 부여, 밀봉, 불투명한 봉투에 의해 은폐되었고, 두 센터 병원 약사가 보관하고 있다.”(41)

“포함기준의 정확성에 대한 전화 검증을 받고 중앙에서 처방을 배정하였다...”(42)

“글렌펠드 병원 약국은 무작위 배정을 하고, 약제를 배포하고, 시험코드를 보유했는데, 이것은 연구 후에 공개되었다.”(43)

4.3.3 기저 상태 불균형을 사용하여 무작위 배정 프로세스 문제 확인

RoB 2 도구는 기저상태 특성을 비교하여 무작위 배정 프로세스에 문제가 있었는지를 고려한다. 단, 우연으로 인해 예상되는 범위를 명백히 벗어난 경우에만 무작위 과정에 문제가 있는 것으로 해석해야 한다(섹션 4.3.3.1 참조).

무작위 배정 과정이 잘못되어 심각한 기저상태 불균형이 발생할 수 있다(44). 이러한 오류는 예를 들어 ‘마이너스’ 대신 ‘플러스’를 작성하는 것과 같은 최소화 프로그램의 오류같이 안전장치가 불충분하여 발생한 의도하지 않은 오류 때문에 발생할 수 있다. 이로 인해, 하나 이상의 예측 요인의 구간 차이를 최소화하는 대신 극대화할 수 있다.

기저상태 불균형 평가는 모든 무작위 배정된 연구대상자 자료에 기초해야 한다. 기저 상태 자료가 임상시험을 완료한 연구대상자(또는 무작위 배정된 연구대상자 중 일부 다른 하위 집합)에게만 제공되는 경우, 기저상태 불균형을 평가하기가 더 어려우며, 자료누락의 비율과 이유를 고려할 필요가 있다. 분석된 연구대상자의 기저상태 특성만을 보고하는 관행은 보건의료 임상시험에서는 일반적인 것은 아니지만 사회적 돌봄(social care)과 같은 다른 영역의 임상시험에서는 일반적일 수 있다.

4.3.3.1 기저상태의 균형 불균형

표본수가 큰 임상시험에서, 단순 무작위법은 비교적 유사한 크기의 중재군을 생성한다

(21~23). 표본 수가 작은 임상시험의 경우, 단순한 무작위 배정은 때때로 확률적으로 또는 예측 인자의 분포에서 실질적으로 다른 군을 생성할 수 있다(45). 예를 들어, 군당 250명의 연구대상자와 각각 20% 유병률인 5가지 중요한 예측 요인을 가진 경우, 두 군 사이에 적어도 7% 이상 차이를 가질 확률은 23%이다(46).

우연에 의한 불균형은 체계적인 비뚤림의 원천이 아니며, RoB 2 도구는 **우연으로 인해 발생한 기저상태 변수의 불균형을 확인하는 것을 목표로 하지 않는다**. 기존 0.05 임계값에서 '통계적으로 유의함'으로 확인된 약간의 차이는 대개 우연의 결과로 간주해야 한다.

중재 효과에 대한 95% 신뢰 구간은 중재군 간 예측 요인의 불균형에서 발생하는 불확실성을 포함한다(47). 그럼에도 불구하고, 예측 요인에서 우연에 의해 기저상태 불균형이 발생하는 경우에는 보정이 필요하다(48). 이는 임상시험 분석 계획에 명시된 규칙에 따라 시행되어야 하고 연구자에게 자료의 눈가림을 풀기 전에 시행하여야 한다(47).

메타 분석에 포함된 임상시험 전체를 보면, 우연에 의한 불균형의 평균 효과는 0이 되며 메타 분석 결과의 신뢰 구간은 그 효과를 포함한다. 연구 내부가 아닌 전체에서 중요한 우연에 의한 불균형이 있는 연구가 합성에 미칠 수 있는 영향은 연구 특정 비뚤림 평가 위험 외부에서 고려될 필요가 있다.

4.3.3.2 무작위 과정에 문제가 있었음을 암시하는 기저상태 불균형

① 의도된 배정 비율과 비교하여 중재군 크기 간의 상당한 차이

한 예로 1948년에 시행된 심근경색에 대한 항응고제 효과를 기존 치료법과 비교한 임상시험이 있다(49). 흡수 입원일(n=589)환자에게는 항응고제를 짝수 입원일(n=442)에 입원한 환자에게 기존 치료법을 투여했다. 대상자 수가 매우 다르게 나왔는데(P=0.001) 이 정도는 예상 배정 비율(1:1)로 볼 때 이 같은 큰 수치 차이는 가능하지 않았다. 이에 연구자들이 흡수 날짜에 임상시험에 더 많은 환자를 모집하도록 배정을 조작했다는 의혹이 제기되고 있다(49).

② 중재군 간 기저상태 특성 차이가 통계적으로 유의한 차이가 나고 이것이 우연으로 예상되는 범위를 벗어난 경우

기저상태 특성 차이에 대한 통계적 검정은 진정한 무작위 임상시험에 사용해서는 안 되는 것이 널리 알려져 있는데 그 이유는 중재 집단 사이에 체계적인 차이가 없다는 귀무가설이 사실로 알려져 있기 때문이다. 그러나 이러한 기저상태 검정은 원칙적으로 무작위 배정이 성공적으로 실행되었는지 여부를 검사하는 데 사용될 수 있다. 그러한 근거가 적절히 해

석되는 것이 중요하다. 무작위 배정이 이루어졌으면, 20개 검사 중 1개가 5% 수준에서 통계적으로 유의한 기저상태 불균형이 있을 것으로 예상된다. 그러나 기저상태 불균형이 이보다 상당히 큰 비율로 나타나거나 P 값이 극히 작은 경우에는 무작위 과정에 문제가 있음을 시사할 수 있다.

그러나, 임상시험 연구자들이 보고된 기준 상태 불균형에 대한 검사 결과 중 일부를 통계적으로 유의미하거나 통계적으로 유의하지 않기 때문에 선택할 수 있다. 또한, 서로 다른 예측 요인 사이에 상관관계가 있을 수 있다(예를 들어, 연령의 불균형이 연령에 영향을 받는 다른 예측 요인의 불균형으로 이어질 수 있음). 따라서 문헌고찰 저자는 기저상태 특성 간에 통계적으로 유의한 차이 이상의 차이가 있다고 결론내리는데 신중해야 한다.

③ 확률로 인해 발생할 가능성이 낮은 주요 예측 요인 또는 결과 변수의 기저상태 불균형 일부 요소들은 연구대상자 모집에 영향을 미칠 수 있으며, 따라서 (의식적이든 무의식적이든) 이들 요소는 시험 결과에 영향을 미치기를 원하는 연구자들에 의해 조작될 가능성이 높다. 문헌고찰 팀은 가능하면 관심 결과에 영향을 미칠 수 있는 주요 예측 요인을 사전에 파악해야 한다. 즉 문헌고찰 팀의 구성원인 주제 전문가의 지식, 초기문헌 문헌고찰(주제범위 문헌고찰) 등을 통해 이를 파악해야 한다. 이 지식을 바탕으로, 하나 이상의 주요 예측 요인에서의 불균형 여부를 고려하여 P 값이 우연으로 인해 발생할 것 같지 않을 정도로 작고(예: <0.001), 그 차이가 결과적으로 비뚤림에 대한 교란요인될 가능성이 충분히 클 경우 연구를 비뚤림위험 ‘높음’으로 간주해야 한다.

숲 그림을 통해 중재군별 기저상태 특성 차이를 도표화하면 여러 연구간에 중재군간의 기저상태 차이를 시각화하는 데 도움이 될 수 있다. 방법론적 사례 연구를 통해 명백한 치료 효과의 차이가 실제로 중재군 간의 기저상태 불균형 때문에 나타난다는 것을 입증했다(50).

④ 우연으로 설명할 수 없는 기저상태 특성의 과도한 유사성

중재군간에 과도한 유사성이 우연에 의해 발생하는 차이와 양립할 수 없는 경우, 무작위 배정 방법의 결점이 있거나 시행하지 않았다는 근거를 제공할 수 있다. 5,087건의 무작위 임상시험의 기저상태 자료를 조사하여, Carlisle은 우연으로 설명할 수 없는 기저상태 유사성을 관찰했으며, 이는 여러 이유 중 자료 위조의 가능성이 크다고 하였다(51). Carlisle은 또한 이후 철회된 임상시험들에서 지나치게 유사한 임상시험의 비율이 높은 것도 관찰하였다. 제한 무작위 배정 방법(섹션 4.3.1 참조)은 단순한 무작위 배정 방법보다 기저상태 값이 더 유사할 가능성이 큰 점에 유의할 필요가 있다.

4.3.4. 기저상태 불균형을 보정하는 분석

임상시험 연구자가 중재군 간의 기저상태 불균형을 관찰한 경우, 예측 변수의 기저상태 값을 보정하는 분석을 수행할 수 있다. 그러나 이러한 불균형이 우연에 의한 것이 아니라 무작위 배정 과정의 문제로 인해 발생한 경우, 비뚤림위험을 제거하기 위해서는 중재군 배정에 영향을 미치는 모든 예측 요인에 대해 보정이 필요할 것이다. 하지만 그런 작업은 가능하지 않기 때문에 그러한 분석은 기껏해야 비뚤림위험을 줄이는 정도이다. 문헌고찰 저자가 무작위 배정 실패를 완화하기 위해 기저상태 불균형에 대해 보정하는 임상 시험에 대해 비뚤림위험을 평가하고자 하는 경우, 연구는 ROBINS-I 도구(52)를 사용하여 평가하여야 한다.

4.4. 신호 질문 및 비뚤림위험 판단 기준

이 영역에 대한 신호 질문은 box 2-4와 같다. 하나의 신호 질문에 대한 응답이 다른 질문에 대한 응답에 영향을 미쳐서는 안 된다는 점에 유의해야 한다. 예를 들어, 임상 시험의 기저상태 불균형이 크지만 저자가 적절한 무작위 배정 방법을 보고하는 경우, 보고된 적절한 방법에 근거하여 순서 생성과 배정 은폐를 평가해야 한다. 불균형에 대한 우려는 기저상태 불균형에 대한 질문에 대한 응답에 반영되어야 하며 영역 수준 판단에 반영되어야 한다.

비뚤림위험 판단에 도달하기 위한 기준은 표 2-3에 제시되어 있으며, 이를 구현하기 위한 알고리즘은 표 2-4와 그림 2-1에 제시되어 있다. 비뚤림위험이 낮다고 판단하려면, 임상 시험에서 연구대상자의 배정 순서를 은폐하는 적절한 방법을 필요로 하며, 배정 순서 생성에 대한 우려가 없어야 한다.

Box 2-4. 무작위 배정 과정에서 발생하는 비뚫림위험

신호질문	세부 설명	응답 옵션
<p>1.1 배정 순서는 무작위였나?</p>	<p>순서 생성 과정에서 무작위 구성요소가 사용된 경우 '그렇다(Y)'에 답한다. 예를 들어 컴퓨터 난수 생성, 난수표 사용, 동전 던지기, 카드나 봉투 뽑기, 주사위 던지기, 추첨 등이 해당한다. 최소화는 일반적으로 무작위 요소(적어도 점수가 같을 때)로 구현되므로 최소화를 사용하여 생성되는 배정 순서는 무작위로 간주해야 한다.</p> <p>배정 순서를 생성하는 데 무작위 요소를 사용하지 않았거나 순서가 예측 가능한 경우 '아니다(N)'로 응답한다. 예를 들어, 교대 날짜(생년월일 또는 입원일)에 기초한 방법, 환자 의무기록 번호, 임상 의사 또는 연구대상자에 의한 배정 결정, 중재의 가용성에 기초한 배정 또는 그 밖의 체계적 또는 우발적 방법 등이 해당한다.</p> <p>무작위 배정 방법에 대한 유일한 정보가 "연구가 무작위로 진행되었다"라는 문구일 경우 '정보 없음(NI)'에 해당한다.</p> <p>어떤 상황에서는 '아마도 아니다(PN)' 또는 '아마도 그렇다(PY)'가 적절한 응답일 수 있다. 예를 들어, 숙련된 임상시험팀에서 시행한 대규모 임상시험 논문이 논문 분량 제한이 엄격하게 적용되는 학술지에 게재되었다면, 무작위 순서 생성에 대한 특정 정보가 없는 경우 '정보 없음'이 아니라 '아마도 그렇다'라고 응답할 수 있다. 또한 유사한 시기에 동일한 연구팀에서 수행한 다른 임상시험이 비무작위 연구로 진행되었다고 되어 있으면 현재의 연구는 유사한 방법을 사용하여 수행되었다고 가정하는 것이 합리적일 수 있다.</p>	<p>Y/PY/ PN/N/ NI</p>
<p>1.2 연구대상자가 모집되어 중재에 배정될 때까지 배정순서가 은폐되었나?</p>	<p>임상시험에서 연구대상자들에게 중재를 배정하기 위해 어떤 형태의 원격 또는 중앙 관리 방법을 사용한 경우, 배정 과정은 모집 담당자와 무관하게 외부 조직에서 통제하는 경우(예: 독립적 중앙 약국, 전화 또는 인터넷 기반 무작위 배정 서비스 제공자) '그렇다'라고 응답한다. 봉투나 약용기가 적절하게 사용된 경우 '그렇다'라고 대답한다. 봉투는 불투명하고 순차적으로 번호가 부여되어야 하며 변조 방지 봉인으로 밀봉되어야 한다. 연구대상자들이 비가역적으로 배정된 후에만 봉투가 개봉되어야 한다. 약물 용기는 순차적으로 번호가 매겨져야 하며 동일한 외관을 가지며, 연구대상자에게 비가역적으로 배정된 후에만 폐기 또는 투여해야 한다. 이 수준의 세부사항은 보고서에 거의 제공되지 않으며, 'PY' 또는 'PN'의 대답을 정당화하기 위한 판단이 필요할 수 있다. 모집담당자 또는 연구대상자가 향후 배정에 대한 지식을 가지고 있다고 의심할 만한 이유가 있는 경우 '아니다'에 응답한다.</p>	<p>Y/PY/ PN/N/ NI</p>
<p>1.3 중재기간 기저상태 차이가 무작위 과정의 문제를 시사하나?</p>	<p>우연에 의한 차이로 간주되는 차이는 비뚫림위험으로 이어지지 않는다는 점에 유의한다. 기준 0.05 임계값에서 '통계적으로 유의함'으로 확인된 적은 차이는 우연에 의한 차이로 간주해야 한다.</p> <p>불균형이 뚜렷하지 않거나 관찰된 불균형이 우연에 의한 차이로 간주해야 하는 경우 '아니다'에 답한다. 무작위 배정 과정의 문제를 시사하는 다음과 같은 불균형이 있는 경우 '그렇다'에 답한다.</p>	<p>Y/PY/ PN/N/ NI</p>

	<p>(1) 의도된 배정 비율과 비교하여 중재군 크기 간의 상당한 차이 혹은</p> <p>(2) 중재군 간 기저상태 특성의 통계적으로 유의한 차이에 대한 우연이라고 할 수 없는 상당한 초과 혹은</p> <p>(3) 하나 이상의 핵심 예측 요인 또는 예후 변수에 대한 기저 상태 불균형이 우연으로 인해 발생할 가능성이 매우 낮고, 군간 차이가 중재 효과 추정치의 비뚤림을 초래할 정도로 큰 경우 또한 무작위 배정 과정에 문제가 있었다고 의심하는 다른 이유가 있는 경우 혹은</p> <p>(4) 우연이라고 보기 힘들 정도로 기저상태 특성이 과도하게 유사한 경우</p> <p>이용 가능한 유용한 기준 정보가 없을 때(예: 최종 분석 참여자의 기저 상태 특성만 보고한 연구) '정보 없음'으로 답한다. 이 질문에 대한 응답이 1.1 또는 1.2 질문에 대한 응답에 영향을 미치지 않아야 한다. 예를 들어, 임상시험 결과에 있어서 기저상태 불균형을 가지고 있지만 저자들이 적절한 무작위 배정 방법을 보고하는 경우, 질문 1.1과 1.2에 대한 응답은 보고된 적절한 방법에 근거하여 답변해야 하며, 불균형에 대한 우려는 질문 1.3에 대한 답변에서 제기되어야 하며, 영역 수준 비뚤림위험 판단에 반영되어야 한다. 임상시험 시행자는 예측 인자의 불균형을 기저상태에서 통제함으로써 결함있는 무작위 배정을 해결하려는 분석을 수행할 수 있다. 무작위 배정 과정의 문제로 인한 비뚤림위험을 제거하기 위해서는 기저상태에서 불균형이 있는 모든 예측 요인을 알고 측정할 필요가 있을 것이다. 모든 중요한 예측 요인을 알고 측정할 가능성은 낮으므로 그러한 분석은 기껏해야 비뚤림의 위험을 줄이는 정도이다. 문헌고찰 저자가 무작위 배정 실패를 완화하기 위해 기저상태 불균형을 보정하는 임상 시험의 비뚤림위험을 평가하고자 하는 경우, ROBINS-I 도구를 사용하여 평가해야 한다.</p>	
<p>비뚤림위험 평가</p>	<p>알고리즘 참조</p>	<p>낮음/높음/일부 우려</p>
<p>선택사항: 무작위 배정 과정에서 발생하는 비뚤림의 예측 방향은 어떠한가?</p>	<p>비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향함, 혹은 임상시험 중재선호, 비교중재 선호로 기술한다.</p>	<p>적용불가 / 임상 시험 중재선호 / 비교중재선호 / 무효점을 향함 / 무효점에서 멀어짐/예측불가</p>

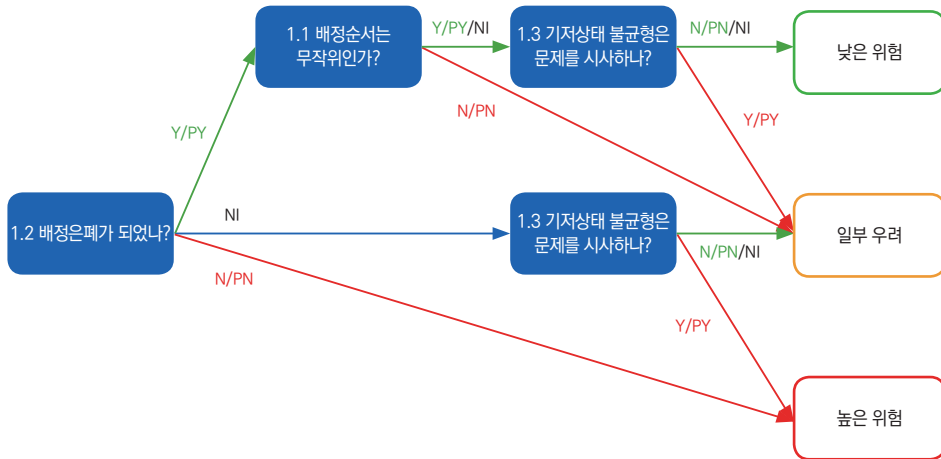
표 2-3. 무작위 배정 과정에서 발생하는 비돌림위험 판단 기준

<p>비돌림위험 낮음</p>	<p>(i) 배정 은폐가 적절히 수행되었다. 그리고 (ii 1) 증재군간 기저상태 차이는 우연에 의해 발생하는 정도이다. 혹은 (ii 2) 기저상태 불균형에 대한 정보가 없다. 그리고 (iii 1) 배정 순서가 무작위이다. 혹은 (iii 2) 배정순서가 무작위라는 정보가 없다.</p>
<p>비돌림위험 일부 우려</p>	<p>(i.1) 배정 은폐가 적절히 수행되었다. 그리고 (i.2.1) 배정 순서가 무작위가 아니다. 혹은 (i.2.2) 증재군간 기저상태의 차이가 무작위 과정의 문제를 시사한다. 혹은 (ii.1) 배정은폐에 대한 정보가 없다. 그리고 (ii.2) 증재군간 기저상태 차이는 우연에 의해 발생하는 정도이다. 혹은 (iii) 신호질문에 응답할 수 있는 어떠한 정보도 없다.</p>
<p>비돌림위험 높음</p>	<p>(i) 배정 은폐가 부적절하게 수행되었다. 혹은 (ii.1) 배정 은폐에 대한 어떠한 정보도 없다. 그리고 (ii.2) 증재군간 기저상태의 차이가 무작위 과정의 문제를 시사한다.</p>

표 2-4 무작위 배정 과정에서 발생하는 비뚤림위험 판단에 대한 신호질문 맵핑

신호질문			영역 수준 판단	
1.1 순서 무작위?	1.2 배정 은폐?	1.3 불균형 문제시사?	비뚤림위험	설명
Y/PY/NI	Y/PY	NI/N/PN	낮음	판단에 상당한 융통성이 있다. 무작위 방법은 적절하지만 상당한 기저상태 불균형이 있는 경우 잘 조사해 보아야 한다. 이런 경우 비뚤림위험 '낮음' 혹은 '높음'에 도달할 수 있다. 기저상태 불균형이 있는 경우 특히 순서 생성 방법이 부적절한 경우 비뚤림위험 '높음'으로 판단될 수 있다.
Y/PY	Y/PY	Y/PY	일부 우려	
N/PN/NI	Y/PY	Y/PY	일부 우려	
Any response	NI	N/PN/NI	일부 우려	
Any response	NI	Y/PY	높음	
Any response	N/PN	Any response	높음	

그림 2-1. 무작위 배정과정에서 발생하는 비뚤림위험 판단을 위한 알고리즘



5. 의도한 증재에서 이탈로 인한 비뚤림

5.1. 배경

이 영역은 의도한 증재에서 이탈이 있을 때 발생하는 비뚤림에 대한 것이다. 이러한 이탈은 임상시험 프로토콜과 일치하지 않는 추가 증재, 의도한 프로토콜 증재 이행의 실패, 임상시험 연구대상자가 배정된 증재에 대해 비준수하는 것 등이 포함된다. 의도한 증재에서 이탈로 인한 비뚤림은 실행 비뚤림의 일종이다.

의도한 증재는 임상시험 프로토콜에 충분히 명시되어야 하지만, 특히 증재가 연구대상자의 건강 또는 사건의 경험에 따라 바뀌도록 의도된 경우에는 명시되지 않는 경우도 있다. 예를 들어, 연구자는 다음과 같이 의도할 수 있다:

- 류마티스 관절염 증상 조절을 위한 신약 임상시험에서 심각한 독성을 경험한 연구대상자는 추가 치료를 받거나 대체 약물로 전환해야 한다.
- 특정 항암제 임상시험에서 암이 진행되는 연구대상자는 2차 증재로 전환해야 한다.
- 안정 협심증의 보존적 관리와 외과적 증재를 비교한 임상시험에서 불안정 협심증으로 진행되는 연구대상자는 외과적 증재를 받게 한다.

임상시험 프로토콜과 일치하는 증재에 대한 이러한 변경은 비뚤림을 유발하지 않으며, 의도한 증재에서 이탈로 간주되지 않아야 한다.

불행히도, 임상시험 프로토콜에는 초기 증재에서 이탈되는 상황을 완전하게 명시하거나 명확하게 설명하지 못할 수 있으며, 프로토콜에 일치하지 않은 증재의 변경(의도한 증재에서 이탈된 것임)과 연구자의 의도와 일치하는 증재의 변화를 구별 할 수 없는 경우가 많다. 예를 들어, 중앙임상시험에서 진행을 정의하지 않거나 질병이 진행되고 있는 환자에게 사용해야 할 2차 약물을 구체적으로 명시하지 않을 수 있다. 따라서 RoB 2 사용자는 그들이 프로토콜과 일치한다고 고려하거나 일치하지 않는다고 고려하는 증재 변경 사항을 기록하는 것이 필수적일 수 있다. 유사하게, 비교증재가 “일상 진료”인 임상시험의 경우, 프로토콜에는 일상진료가 무엇인지 특정되어 있지 않거나 임상시험 증재와 함께 사용될 것으로 예상되는지 여부를 명시하지 않을 수 있다. 따라서 RoB 2 도구 사용자는 일상적인 진료와 일치하는 증재를 기술할 필요가 있다.

5.1.1. 비프로토콜 중재

임상시험 연구대상자가 임상시험 추적관찰 기간 동안 받을 수 있고 관심 결과에 영향을 줄 수 있는 비프로토콜 중재는 추정된 중재 효과에 비뚤림을 유발할 수 있다. 가능한 경우, 문헌고찰 저자는 잠재적인 비프로토콜 중재를 사전에 지정해야한다(문헌고찰 프로토콜 작성 단계에서). 그것은 문헌고찰팀 구성원의 전문 지식, 초기(주제범위) 문헌고찰 및 보건의료 전문가와의 토론을 통해 확인될 수 있다.

5.1.2. 관심 효과의 역할

섹션 1.3에 설명된 바와 같이, 이 영역에 대한 평가는 문헌고찰 저자가 관심을 갖는 중재 효과에 따라 달라진다.

- ① 기저상태에서 중재에 대한 배정의 효과(추적관찰 기간 동안 중재를 받거나 준수했는지 여부에 관계없이, 때로는 '중재 배정 효과'라고도 함) 또는
- ② 임상시험 프로토콜에 명시된 대로 중재를 준수한 효과('중재 준수 효과'라고도 함).

이러한 효과는 일부 환자가 배정된 중재를 받지 않거나 기저시점 후에 배정된 중재에서 이탈이 생기면 달라진다.

특정 임상시험에서 중재 배정의 순 효과는 세 가지 요소, 즉 중재의 실제 효과, 중재에 대한 준수 정도 및 유형, 결과에 영향을 미칠 수 있는 임상시험에 특정된 모집방법이나 연구대상자의 참여(engagement) 활동에 따라 달라진다. 세 번째 요소의 예로, 사전 동의를 확보하는 과정에서 제공되는 정보가 행동 변화를 유발하여 예후가 개선되거나, 연구대상자가 건강 관리에 참여 혹은 중재 준수 성향이 강화될 수 있다. 대조적으로, 중재 준수의 순 영향은 중재의 실제 영향과 연구대상자의 결과에 영향을 미치는 임상시험에 특정된 모집방법이나 연구대상자의 참여활동에만 달려 있다.

의도한 중재에서 이탈의 정도는 중재 배정 치료효과에 대한 메타분석에서 이질성을 유발하는 비슷한 중재를 비교하는 임상시험마다 다를 수 있다. 예를 들어, 배정된 중재에 대한 비준수비율은 연구대상자 모집 중에 무작위 배정 과정에서 설명된 치료에 따라 다를 수 있다.

임상시험에 모집된 연구대상자는 비교될 중재들로부터 혜택을 볼 수 있는 모든 개인을 대표하는 것은 아니다. 예를 들면, 프로토콜에는 최대 연령을 정하거나 동반 질환이 있는 환자를 배제함으로써 자격을 제한하는 내용이 있을 수 있다. 모집에 동의한 환자는 동의하지 않는 환자보다 특정 중재에 대한 선호도가 낮거나 중재에 대한 동기 부여가 다를 수 있다. 또한, 보건의료의 질은 임상시험 환경과 비임상시험 환경 사이에 차이가 있을 수 있다. 이러한 문제로 인해 임상시험 결과의 일반화가 임상시험 이외의 환경에는 제한적일 수 있으며(외적 타당도), 제한적인 자격 기준을 피하거나 일상적인 진료의 전형적인 상황에서 임상시험을 시행함으로써 임상시험 결과를 보다 대표적인 것으로 만드는 이니셔티브가 생겨났다. 그러나 이러한 것들은 임상 시험이 수행되는 의료 환경에서 치료받는 임상시험 대상자들에서 비뚫림 없는 중재 효과를 추정하는 우리의 능력에 영향을 미치지 않는다. 따라서 일반화가능성 또는 이동성(transportability)의 이러한 측면은 RoB 2 도구에서 다루지 않는다.

5.1.2.1 중재 배정 효과에서 비뚫림

일부 임상 시험, 특히 눈가림이 되지 않은 시험에서는 임상시험 모집 및 참가 활동으로 인해 비프로토콜 중재가 발생할 수 있다. 중재 제공자가 비 프로토콜 중재를 시행하거나 프로토콜 중재를 시행하지 않음으로써 임상 시험을 통한 중재의 비교성이 훼손될 수도 있다. 이것은 무의식적 과정(예를 들어, 임상적 동등성이 명확하지 않아 한 그룹에서 비 프로토콜 중재를 시행하고 다른 군에는 프로토콜 중재를 제공하는 등)일 수도 있고 의식적 과정(예 : 이해 상충으로 인한)일 수도 있다

우리는 연구대상자 모집 및 참여 활동(engagement activities)의 영향과 임상시험 인력이 중재를 전달하는 과정에서 나타나는 무의식적 또는 의식적 과정의 영향을 임상시험 맥락이라는 용어로 표현한다. RoB 2에서 중재배정의 효과와 관련하여 다루는 의도한 중재에서 이탈은 다음과 같다.

- ① 임상시험 맥락에서 발생했다.
- ② 임상시험 프로토콜과 일치하지 않았다.
- ③ 중재결과에 영향을 미친다.

예를 들어, 눈가림되지 않은 연구에서 동의를 확보하는 과정은 추후 비교 그룹에 배정된 연구대상자가 운이 좋지 않다고 느껴 임상시험 중재 또는 예후를 향상시킬 수 있는 다른 중재를 받으려고 할 수 있다. 강력한 모집(참여) 효과가 있는 임상시험의 효과 추정치는 그런 효과가 없는 환경에는 나타나지 않을 수 있기 때문에 그러한 점을 ROB2에서 평가되어야 한다.

유사하게, 식도암에 대한 최소 침습 수술과 개복 수술을 비교하는 공개 연구에서, 프로토콜은 one-lung ventilation을 두군 모두에서 사용하도록 명시했다. 그러나 one-lung ventilation 은 호흡기 감염과 같은 합병증 발생률을 높였기 때문에, 최소 침습군에서는 통상적으로 two-lung ventilation을 사용하는 경우가 많았다. 이러한 비프로토콜 중재는 중재의 안전성에 대한 외과 의사의 신념(임상적 동등성에 차이가 있다)으로 인해 일어나며, 중재에 대한 배정 추정 효과는 비뚤림위험이 높아진다. 하지만 환자가 배정 약물을 복용 중단하는 것과 같은 임상시험 맥락 때문에 발생한 이탈은 중재 배정 효과에 비뚤림을 초래하지 않는다.

5.1.2.2 중재 준수 효과의 비뚤림

시험 프로토콜과 일치하지 않고 중재 결과에 영향을 미치는 의도한 중재에서 이탈은 임상시험 맥락에서 발생했는지 여부에 관계없이 중재 준수 효과에 관련하여 다루어져야 한다. 의도한 중재에서 이탈한 것을 보정할 수 없는 경우가 있다. 그러므로, 임상시험 프로토콜에 명시된 대로 중재 준수 효과를 다룰 때 문헌고찰 저자는 사전 고려사항(섹션 3 참조)에서 의도한 중재(임상시험 프로토콜에서 출발)에서 어떤 유형의 이탈이 조사되는지를 명시하여야 한다. 그러한 것에는, (1) 중재 결과에 영향을 줄 수 있는 비프로토콜 중재의 발생; (2) 중재 결과에 영향을 줄 수 있는 프로토콜 중재 실행 실패; 그리고 (3) 시험 연구대상자가 배정된 중재 비준수가 있다. 예를 들어, START 무작위 배정 임상시험은 HIV 양성자에서 항 레트로 바이러스 요법(ART) 즉시 투여와 지연 투여를 비교하였는데, 지연 개시에 배정된 연구 참여자 중 30%는 프로토콜보다 일찍 ART를 시작했다(6). Lodi와 동료들은 이러한 프로토콜 이탈을 보정하여 중재 준수 효과를 추정했지만 연구대상자가 추적관찰 동안 ART를 계속했는지 여부에 대해서는 시행하지 않았다(7). 그러한 이탈이 있는 경우 문헌고찰 저자는 적절한 통계적 방법을 사용하여 효과를 보정했는지 여부를 고려해야 한다. 의도한 중재에서 이탈이 있는 연구의 예는 박스 5에 있다.

5.1.3. 눈가림의 역할

연구대상자, 중재 제공자가 제공받는 중재를 인지하지 못하도록 하는 것을 눈가림이라고 한다. 일부 영역(예: 안과)에서는 '마스킹'이라는 용어가 선호된다. 성공적인 경우, 눈가림은 중재 배정에 대한 지식 때문에 오염(특정 중재에 배정된 연구대상자가 나머지 다른 중재에 적용을 받는 것)이 늘어나는 것을 막고, 프로토콜에 없는 중재를 받거나 임상시험 연구대상

자가 배정된 중재를 받지 않는 것을 방지할 수 있다. 눈가림이 시행되지 않으면, 임상시험적 맥락과 관련된 의도한 중재에서 이탈이 발생시키거나, 중재준수 효과를 추정하기 위해 중재 구간 이탈의 정도가 달라진다면 비뚫림을 야기할 수 있다.

눈가림은 자신의 치료를 알고 있는 개인의 중재 전략을 비교하는 것이 목표인 실용적 임상시험에는 적합하지 않다. 그러나 위약 효과를 제거하고 프로토콜 중재의 특정 효과를 분리하는 것이 목표인 임상시험에서는 필수적이다. 예를 들어, 통증을 치료하기 위한 침 치료는 치료가 없는 것과 비교할 때 이점을 발견하는 경향이 있지만, 가짜 침 치료와 비교하거나 눈가림이 연구대상자와 치료자(중재 제공자 이외)에게 시행된 경우 중요한 이점은 없다. 유사하게, 약물과 위약의 눈가림 비교는 관련 화합물의 약리학적 효과를 추정 할 수 있는 반면, 중재를 시행하지 않은 것과 약물을 비교한 경우의 추정 효과는 화합물의 효과와 전체 치료 과정의 효과를 결합한 것이다.

적절성에 관계없이, 예를 들어 수술을 비수술적 중재와 비교하는 임상시험에서, 임상시험 연구대상자에 대한 눈가림은 어렵거나 일부 상황에서는 불가능하다. 눈가림 되지 않은(‘공개된’) 연구는 프로토콜에 없는 중재의 시행을 막기 위한 엄격한 기준에 따라 환자를 치료하는 것과 같이 의도한 중재에서 이탈을 피하기 위해 다른 조치를 취할 수 있다.

RoB 2 도구는 임상시험 동안 연구대상자와 임상시험 담당자가 눈가림되었을 때, 의도한 중재에서 이탈이 임상시험 맥락에 따라 영향을 받지 않는다고 가정한다. 따라서 관심 영역이 중재 배정 효과인 경우 이러한 이탈을 다루는 신호 질문은 눈가림되지 않았을 경우에만 하게 된다.

중재결과 측정의 비뚫림을 피하기 위한 결과 평가자의 눈가림은 ‘중재결과 측정의 비뚫림’ 영역(섹션 7)에서 별도로 고려된다. 탈락(연구에서 탈퇴)으로 인한 비뚫림은 ‘결과 자료 누락으로 인한 비뚫림’ 영역(섹션 6)에서 고려된다.

연구대상자, 보호자, 중재제공자에 대한 눈가림 시도는 성공적인 눈가림을 보장하지 않는다. 많은 눈가림 약물 임상시험의 경우, 약물의 부작용으로 인해 부작용이 비슷한 약물 또는 활성 위약과 같은 유사한 중재를 비교하지 않는 한 일부 연구대상자가 받고 있는 중재를 감지할 수 있다. 또한 무작위배정 시점에 중재 배정이 은폐되지 않은 경우(섹션 4.1.2 참조), 임상시험을 진행하는 동안 배정에 대한 지식을 이용할 수 있어 보호자와 중재 제공자에게 완전히 눈가림되지 않을 수 있다.

몇몇 그룹은 임상시험이 끝날 때 임상시험 연구대상자에게 그들이 어떤 중재를 받고 있는지 물어 보는 것이 합리적이라고 제안했으며, 그러한 보고서에 대한 문헌고찰이 발표되기도 하였다(61, 62), 그러나 성공적인 추측을 한다는 것은 중재결과가 좋은 것 혹은 심각한 부작용이 적극적인 중재로 인한 것일 수 있다는 것을 단순히 반영할 수 있다. 따라서 효능이나 부작용에 차이가 있을 때 성공적인 “추측”이 가능하지만, 눈가림이 유지 된 경우에도 중재가 매우 유사한 효과를 가질 때 아무런 영향도 미치지 않을 것으로 예상된다.

받은 중재를 추론하는 것만으로는 비뚤림위험을 초래하지 않는다. 위에서 논의한 바와 같이, 독성으로 인한 약물 중재의 중단은 일반적으로 의도한 중재에서 이탈로 간주되지 않는다. 이 영역에 대한 신호 질문과 함께 제공되는 세부 사항은 이 문제를 보다 자세히 설명한다.

연구 출판물들은 자주 “이중 눈가림”과 같은 광범위한 용어로 눈가림을 기술한다. 이 용어는 누가 눈가림 되었는지 알기 어렵다. 이러한 용어는 매우 일관적이지 않게 사용되며, 명확히 하기 위한 CONSORT 성명서의 권장 사항에도 불구하고 연구대상자 및 임상시험 담당자의 눈가림 상태에 대한 명확한 보고 빈도는 최고 저널에 출판된 임상시험에서도 낮게 유지된다. 눈가림에 사용된 방법에 대한 문헌고찰은 실제로 사용되는 다양한 방법을 강조한다.

Box 2-5. 의도한 중재에서 이탈이 있는 연구의 예

예 1 : 무작위 배정된대로 치료 받지 않은 상당수의 연구대상자

요추 추간판 탈출증 수술의 효과를 결정하기 위해, SPORT 임상시험 (Spine Patient Outcomes Research Trial)은 요추 추간판 탈출증을 가진 환자를 무작위로 배정하여 수술 치료(절제술) 또는 비수술 치료(진통제, 교육, 물리 치료 및 침술 등을 포함한 다양한 중재 포함)을 받았다.(70).

ITT 분석에서는 무작위 배정 후 2 년 동안 1차 결과 (SF -36 신체 통증 및 신체 기능 척도)에서 중재기간에 차이가 있다는 근거를 발견하지 못했다. 그러나 외과 치료에 배정 된 환자의 60% 만이 수술을 받았고 비수술군의 45%가 외과적으로 치료되었다. 저자는 ‘치료된 대로’ 분석을 수행하여 수술의 이점을 보여주었다. 이는 수술 받은 연구대상자(지정된 중재에 관계없이)의 기본 특성이 그렇지 않은 사람들의 특성과 실질적으로 다르기 때문에 중재에 따른 영향(중재준수 효과)에 비뚤림이 있었다.

예 2 : 비교군에서 임상시험군으로 교차

경피 관상동맥 중재술의 효과를 결정하기 위해, FAME 2 임상시험(Fractional Flow Reserve vs Angiography for Multivessel Evaluation 2 trial)은 안정 협심증 환자를 약물 방출 관상 동맥 중재술(PCI)+최적 내과치료(optimal medical therapy, OMT) vs OMT로 무작위 배정했다(71). 이 임상시험은 조기종료되었는데 PCI군에서 모든 원인, 비치명적 심근경색 또는 긴급 재관류술로 인한 사망의 일차 복합 중재결과의 위험이 감소하였기 때문이다(HR 0.39; 95% CI 0.26-0.57; $p < 0.001$). 이는 PCI군에서 긴급 재관류술이 77% 감소(HR 0.23; 95% CI 0.14-0.38; $P < 0.001$)하였기 때문이다. 반면 사망 또는 심근경색 복합 중재결과는 21% 감소(HR 0.79; 95% CI(0.49-1.29)에 그쳤다. 그러나 OMT에 배정된 환자 41%가 PCI와 '교차' 했다. PCI 수행의 40%는 임상적 적응증이 되었지만(심근 경색이나 불안정 협심증 발생) 대부분은 프로토콜 이탈이었다. 따라서 사망 또는 심근경색에 대한 PCI 대 OMT (ITT 효과)에 대한 배정의 영향은 중재에 따른 영향(PP 효과)보다 무효점(null)에 가깝다. 중재준수효과를 추정하려면 OMT 배정군에서 비 임상적으로 시행된 PCI가 시행되었을 때 후속 조치를 중도절단하고 이를 설명하기 위한 분석(역확률 가중 분석 등을 통해)을 보정해야 한다.

예 3 : 비프로토콜 중재가 중재기간 다르면 중재결과에 영향을 줄 수 있음

공개 라벨 연구를 통해 식도암에 대한 최소 침습 수술과 개복 수술 후 호흡기 감염 비율을 비교했 (72). 비프로토콜 중재 제공에서 중재군 간에 두 가지 중요한 차이점이 있었다. 첫째 one-lung mechanical ventilation(RTI 등 호흡기 합병증을 증가시키는 것으로 생각됨)는 개복군에만 사용되었으며, 최소 침습군에는 two-lung mechanical ventilation를 수행했다(두 유형의 기계 환기는 두 군에서 모두 사용 가능). 둘째, 경막 외 진통제가 개복 수술군에서 더 빈번하게 사용되었다. 경막 외 진통제 투여 환자는 일반적으로 이동성이 떨어지기 때문에 RTI 발병 위험이 높아진다. 이러한 비프로토콜 중재 중 일부가 임상시험 맥락으로 인해 관리 된 경우, 공동 중재가 중재기간에 균형을 이루지 않으면 중재 결과에 영향을 줄 수 있기 때문에 중재에 대한 배정의 추정 효과에는 비뚤림위험이 있다.

예 4 : 중재가 성공적으로 실행되지 않았다

급성 충수염 환자를 표준 복강경 충수절제술 또는 단공 절개 복강경 충수절제술을 받도록 무작위 배정하였다(73). 표준 복강경 수술은 3개의 구멍(2 개는 복강경기구 1개는 복강경 카메라)이 있는 반면, 단공 절제술은 카메라와 1개의 기구를 배치하는 단 하나의 구멍만 있다. 그러나 단공 절제술로 무작위 배정된 환자의 50% 미만이 의도한 대로 시술을 받았다. 추가기구와 구멍이 각각 51%와 10%에서 추가로 필요했다. ITT 분석에 따르면 상처 감염 위험(1차 결과), 정상복귀 시간에서 중재군 간 차이가 없었다. 이런 분석은 중재 배정효과를 추정하는 데는 적합하지만, 중재준수에 대한 효과는 무효점에 가깝다.

5.2. 의도한 중재에서 이탈로 인한 비뚤림의 경험적 근거

의도한 중재에서 이탈로 인한 비뚤림의 경험적 근거는 대체로 “이중 눈가림”에 대한 보고가 중재 효과와 관련이 있는지를 탐색하는 연구에서 나오게 된다. 이러한 연구는 적어도 하나의 임상시험이 눈가림이 이루어진 연구에 대한 메타분석으로 제한되므로, 눈가림이 실행 가능하지 않거나 적절하지 않은 상황에는 적용되지 않는다. 현재까지 진행된 가장 큰 메타역학 연구에서, 주관적으로 평가된 결과를 가진 임상시험에서 이중 눈가림되지 않았거나 불명확한 이중 눈가림(이중눈가림 대비)은 23% 정도 오즈비를 과대 추정하였다. 또한, 이중 눈가림이 되지 않았거나 불명확한 경우 임상시험 간 이질성이 커졌다. 반면 사망률 임상시험이나 기타 객관적으로 평가된 결과에서 그러한 비뚤림의 근거는 거의 없었다. 주관적으로 측정된 연속 중재결과(예: 환자 평가 설문지)를 조사하는 두 개의 다른 연구에서는 표준화 평균 차이가(연구대상자의 눈가림이 시행된 것과 비교하여) 연구대상자의 눈가림이 부족하거나 불명확한 임상시험에서 과장된 경향이 있었다. 기존의 경험적 근거는 연구대상자와 임상시험 담당자의 눈가림을 결과 평가자의 눈가림과 구별하지 않기 때문에 이러한 눈가림의 측면 중 어느 것이 비뚤림을 방지하는 데 가장 중요한지는 명확하지 않다.

‘계획서 순응군’ 분석(Naïve ‘per-protocol’ analyses)은 ‘배정된 대로’ 분석과 비교하여 중재 배정의 효과 추정치를 과장하는 것으로 밝혀졌다. (ITT에 비해) ‘mITT’ 분석을 사용하였으면 효과가 과대추정 되었다. Tierney 등은 모든 임상시험 연구대상자를 포함한 분석과 비교하여 연구대상자를 배제하고 분석한 임상시험에서 실험중재를 선호하는 경향을 관찰하였다(80).

특히 1996년 이전, 즉 CONSORT 이전 시대에는 연구대상자 제외가 빈약하게 보고되었기 때문에 실증적 연구의 해석이 어렵다. 예를 들어 Schulz는 연구대상자 제외에 대한 명백한 보고가 결여되어 “유익한” 효과 크기를 키울 뿐 아니라 적절한 배정은폐의 가능성을 더 낮추게 되는 것을 관찰하였다. 따라서 Schulz 연구에 따르면 임상시험에서 연구대상자 제외 보고에 대한 실패는 제외가 실제로 시행되지 않았다고 보다 임상시험 수행이 열악하다는 표시일 수 있다(18).

5.3. 도구의 이 영역 사용하기

5.3.1. 중재 배정의 효과

중재 배정의 효과를 평가할 때 신호 질문은 다음과 같다:

- ① 연구대상자, 보호자 및 중재 제공자에게 눈가림되었는지 여부
- ② 이들 그룹 중 일부가 눈가림되지 않은 경우, 의도한 중재에서 이탈이 임상시험 맥락으로 인해 발생했고 중재 효과를 편향시킬 가능성이 있었는지
- ③ 중재 배정 효과를 추정하기 위해 적절한 분석이 사용되었는지 여부

적절한 분석은 연구대상자가 무작위 배정된 중재군대로(실제로 받은 중재에 관계없이) 분석되고 (중재가 의도한대로 실행되었는지 여부와 연구대상자의 준수 여부에 관계없이) 무작위 배정된 모든 연구대상자가 분석에 포함되어야 한다는 ITT 원칙을 따라야 한다. 일부 저자는 결측 중재결과 자료가 있는 연구대상자가 제외된 ‘수정된 ITT’(mITT) 분석을 보고 할 수 있다. 이러한 분석은 이 영역의 목적에 적합한 것으로 간주되어야 한다. 결측된 중재결과 자료로 인한 비뚤림은 “중재결과 자료의 결측으로 인한 비뚤림”(섹션 6 참조) 영역에서 개별적으로 다룬다. ‘mITT’라는 문구는 다른 방식으로 사용되기도 하며, 적어도 한 번의 치료 용량을 받은 연구대상자를 포함하는 것을 나타낼 수 있다. 우리는 이 용어를 중재의 준수보다는 결측 자료와 관련하여 사용한다.

부적절한 분석에는 프로토콜 대로(PP) 분석, ‘치료받은 대로’ 분석 및 무작위 배정 후 적격한 임상시험 연구대상자의 제외(1.3.1 참조)가 포함된 기타 분석이 포함된다. 무작위 배정 후 적격한 임상시험 연구대상자를 배제하는 이유로 부적절한 것은 “효능 부족” 또는 독성 경험에 의해 배제하는 것이다. 그러나 부적격 연구대상자의 무작위 배정 후의 제외(무작위 배정 이후까지 자격이 확인되지 않고 중재군 배정에 의해 영향을 받지 못한 경우)는 적절한 것으로 간주 될 수 있다.

연구 저자에 의해 제외된 연구대상자를 포함(“재 포함”)하는 개별 연구대상자 자료 메타 분석(individual participant data meta-analyses)을 수행하는 것이 가능할 수 있다. 문헌 고찰 저자는 가능하면 이 작업을 수행하는 것이 좋다. 이러한 상황에서, 비뚤림 평가의 위험은 임상시험자가 보고한 연구결과보다는 합성에 포함된 임상시험 연구 결과에 적용되어야

한다.

임상시험이 CONSORT 성명문에 따라 보고되고 CONSORT 흐름도를 포함하는 경우 분석에 대한 신호 질문에 대답하는 것은 간단해야 한다. ‘치료받은 대로’ 분석으로 보고한 경우는 흔하지 않다. 프로토콜대로 분석을 하였다고 보고한 경우는 그보다는 흔하기는 하지만 이들은 ITT 또는 mITT를 대신하기보다는 추가적인 분석을 시행하였다고 보고한 경우가 더 흔하다. 선택이 가능한 경우, ‘배정된 대로’ 분석 결과를 비뚤림위험 평가 및 메타 분석에 포함되는 것이 선호되어야 한다(1.3.1 절 참조). 평가된 결과가 프로토콜대로 분석 또는 “치료 받은 대로” 분석을 기반으로 하는 경우 문헌고찰 저자는 중재의 추정 효과에 상당한 영향을 미칠 가능성이 있는지 평가해야 한다. 이들 중 ‘치료받은 대로’ 분석보다 프로토콜대로 분석을 선택해야 한다.

이 영역에 대한 비뚤림위험은 동일한 사람들이 임상시험 동안 중재 배정을 알고 있더라도 중재 결과마다 다를 수 있다. 예를 들어, 배정된 중재에 대한 지식은 행동(예: 진료소 방문 횟수)에 영향을 줄 수 있지만 생리학적 중재결과(예: 사망 위험)에는 중요하게 영향을 미치지 않는다.

5.3.2. 중재 준수의 효과(‘PP 효과’)

중재 준수의 영향을 평가할 때 신호 질문은 다음과 같다.

- ① 연구대상자, 보호자 및 중재 제공자에게 눈가림 되었는지 여부
- ② 연구대상자, 보호자 또는 중재 제공자가 눈가림되지 않은 경우, 중요한 공동 중재가 중재군간 균형을 이루었는지 여부
- ③ 중재가 성공적으로 이행되었는지 여부와 연구대상자가 배정된 중재를 준수했는지 여부;
- ④ 위의 2번과 3번 질문에서 발생하는 의도한 중재에서 이탈이 발생한 경우, 적절한 분석이 수행되었는지 여부

적절한 분석 방법에 대해서는 Hernán과 Robins가 잘 설명하였다(5). 예를 들어, 일부 상황에서는 도구 변수 접근법을 사용하여 배정된 중재를 받은 연구대상자들 사이의 중재 효과를 추정할 수 있다(1.3.1 절 참조).

5.4. 신호 질문 및 비독립위험 판단 기준

중재 배정의 효과에 대한 신호 문제는 box 2-6에 제공되어있다. 비독립위험 판단에 도달하기위한 기준은 표 2-5에 제시되어 있으며, 이를 구현하기 위한 알고리즘은 표 2-6과 그림 2-2에 제시되어있다. 비독립위험 판단에 도달하기 위한 기준은 표 2-7에 나와 있으며 이를 구현하기 위한 알고리즘이 표 2-8과 그림 2-3에 제공되어있다.

Box 2-6. The RoB 2 tool: 의도한 중재에서 이탈로 인한 비독립(중재 배정 효과)

신호질문	세부 설명	응답 옵션
2.1 연구대상자는 임상 시험 중에 배정된 중재에 대해 알고 있는가?	연구대상자가 배정된 중재를 알고 있다면, 건강 관련 행동이 중재군에 따라 다를 가능성이 있다. 연구대상자에 대한 눈가림은 위약이나 가짜 중재를 통해, 그러한 차이를 방지할 수 있다. 연구대상자가 중재 중 하나에 특정한 것으로 알려진 부작용이나 독성을 경험한 경우, 'Y' 또는 'PY'라고 응답한다.	Y/PY/PN/N/ NI
2.2 보호자와 중재 제공자는 임상시험 중에 연구대상자에 배정된 중재에 대해 알고 있는가?	중재 제공자가 배정된 중재를 알고 있다면, 그 중재를 실행했는지, 비프로토콜 중재가 발생했는지 여부가 중재군 간에 다를 수 있다. 눈가림은 그러한 차이를 막을 수 있다. 중재제공자가 중재 중 하나에 특정한 것으로 알고 있는 부작용이나 독성을 연구대상자들이 경험한 경우, 'Y' 또는 'PY'라고 대답한다. 무작위 배정 은폐가 되지 않는 경우, 중재제공자는 임상시험 중에 연구대상자들에 배정된 중재가 무엇인지 알고 있었을 가능성이 높다.	Y/PY/PN/N/ NI
2.3 2.1과 2.2이 Y/PY//NI 인 경우 : 임상 시험적 맥락 때문에 의도한 중재에서 이탈이 있었나?	중재 배정 효과의 경우, 임상시험 프로토콜과 일치하지 않는 배정 중재의 변경이 임상시험적인 맥락 때문에 발생하여 결과에 영향을 미칠 가능성이 있는 경우에만 이 영역의 비독립이 발생한다. 예를 들어, 동의서를 받는 과정에서 비교군에 배정된 대상자가 불운한 것이라고 느끼게 되어, 임상시험 중재나 더 나은 예후를 가지는 중재군이 배정되려고 노력할 수 있다. 임상시험적 맥락으로 인해 임상시험이 진행되지 않은 상황과 비교했을 때, 프로토콜에 명시된 것 이외의 추가 중재가 제공되었거나, 중재제공자가 의도한 대로 중재 실행을 실패하였거나, 연구대상자가 배정된 중재에 대해 비준수했다는 근거가 있는 경우, 'Y' 또는 'PY'로 응답한다. 임상시험 프로토콜과 일치하는 중재 변경에 대해서는 'N' 또는 'PN'으로 대답한다. 예를 들어, 급성 독성 때문에 약물 중재가 중단되거나 의도된 중재 중에 발생한 중재 결과를 처리하는 목적으로 추가 중재를 사용하는 것과 같은 경우이다. 임상시험적 맥락과 관련이 없는 경우, 중재에 대한 비준수에 대해 'N' 또는 'PN'로 응답한다.	NA/Y/PY/ PN/N/NI

	<p>연구대상자가 중재 중 하나에 특이한 부작용이나 독성이 생겨서 눈가림이 깨진 경우, 임상시험 프로토콜과 일치되지 않고 임상시험적인 맥락 때문에 발생한 중재 변경인 경우에만 'Y' 또는 'PY'라고 응답한다. '정보 없음'이라는 응답은 임상시험 연구자가 임상시험적 맥락 때문에 이탈이 발생했는지 여부를 항상 보고하지 않는 경우에 부여한다. 그러나 그러한 이탈이 발생했을 가능성이 있는 경우 응답은 'PY'여야 한다.</p>	
<p>2.4. 2.3이 Y/PY 인 경우 : 이러한 이탈로 중재 결과에 영향을 미칠 것 같은가?</p>	<p>임상시험 프로토콜과 일관되지 않고 임상시험적 맥락 때문에 발생한 배정된 중재의 변경은 중재 결과에 영향을 미치는 경우에는 중요할 것이고 그렇지 않은 경우에는 중요하지 않을 것이다.</p>	<p>NA/Y/PY/ PN/N/NI</p>
<p>2.5. 2.4이 Y/PY/NI 인 경우 : 의도한 중재에서 이탈이 두 군에서 유사한가?</p>	<p>임상시험 프로토콜과 일관되지 않고 임상시험 맥락 때문에 발생한 배정된 중재의 변경은 중재군 간에 균형이 맞지 않는 경우 비뚤림을 초래할 가능성이 더 크다.</p>	<p>NA/Y/PY/ PN/N/NI</p>
<p>2.6. 중재 배정효과 추정에 적절한 분석을 사용하였나?</p>	<p>중재결과 자료가 결측된 연구대상자를 제외한 '배정된 대로 분석(ITT)'과 '수정된 ITT(mITT)분석' 모두 적절한 것으로 간주해야 한다. '계획서 순응군 분석'(naïve 'per-protocol', 배정된 중재를 받지 않은 연구대상자는 제외)과 '치료한대로 분석(as-treated)'(배정된 중재가 아닌, 받은 중재에 따라 연구대상자가 그룹화되는 경우) 모두 부적절한 것으로 간주해야 한다. 적격 연구대상자를 제외한 분석 역시 부적절하다고 보아야 하지만, 부적격 연구대상자의 무작위 배정 후 배제(무작위 배정 이후까지 자격이 확인되지 않았고, 중재군 배정에 의해 영향을 받을 수 없었을 때)는 적절한 것으로 간주될 수 있다.</p>	<p>Y/PY/PN/N/ NI</p>
<p>2.7. 2.6이 N/PN/NI 인 경우 : 무작위 배정군에 따라 연구대상자를 분석하지 못한 경우(연구 결과에) 상당한 영향을 미칠 가능성이 있었는가?</p>	<p>이 질문은 잘못된 중재군에서 분석되었거나 분석에서 제외된 연구대상자의 수가 결과에 상당한 영향을 미칠 수 있을 정도로 충분한지 여부를 다룬다. 정확한 규칙을 지정할 수 없다. 연구대상자의 5% 미만 이 잘못된 그룹에서 분석되거나 배제되었더라도, 중재 결과가 드물거나 예측 요인과 강하게 관련되는 경우라면 상당한 영향이 있을 수 있다.</p>	<p>NA/Y/PY/ PN/N/NI</p>
<p>비뚤림위험 평가</p>	<p>알고리즘 참조</p>	<p>낮음/높음/일부 우려</p>
<p>선택사항: 의도한 중재에서 이탈로 인한 비뚤림의 예측 방향은 어떠한가?</p>	<p>비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향함, 혹은 임상시험 중재번호, 비교중재 선호로 기술한다.</p>	<p>적용불가 / 임상시험 중재번호 / 비교중재 선호/무효점을 향함/ 무효점에서 멀어짐/ 예측불가</p>

표 2-5. 의도한 중재에서 이탈로 인한 비밀림(중재배정효과) 판단 기준

	제 1부 : 2.1에서 2.5까지 질문에 대한 기준	제 2부 : 2.6에서 2.6까지 질문에 대한 기준	영역 판단
비밀림 위험 낮음	(i) 연구대상자, 보호자, 중재제공자가 임상시험 도중 중재군에 대해 모른다. 혹은 (ii 1) 연구대상자, 보호자, 중재제공자가 임상시험 도중 중재군에 대해 알고 있다. 그리고 (ii 2) 임상시험 맥락에서 의도한 중재에서 이탈이 없다.	중재배정 효과 추정에 적절한 분석을 사용하였다.	제 1부 비밀림 위험 낮음 그리고 제 2부 비밀림 위험 낮음
비밀림 위험 일부 우려	(i) 연구대상자, 보호자, 중재제공자가 임상시험 도중 중재군에 대해 알고 있다. 그리고 (ii 1) 임상시험 맥락에서 의도한 중재에서 이탈이 있는지에 대한 정보가 없다. 혹은 임상시험 맥락에서 의도한 중재에서 이탈이 있다 그리고 (ii 1.1.1) 이 이탈이 결과에 영향을 끼치지 않는 가능성이 있다. 혹은 (ii.i.1.2) 이들 이탈이 두 군에 균형이 있다	(i)중재배정 효과 추정에 적절한 분석을 사용하지 않았다. 그리고 (ii) 실패가 분석에 대한 잠재적 영향(중재 효과 추정에 대한)이 그다지 크지 않다.	제 1부 일부 우려 혹은 제 2부 일부 우려 그리고 제 1부가 비밀림위험 높음이 아님 그리고 제 2부가 비밀림위험 높음이 아님
비밀림 위험 높음	(i) 연구대상자, 보호자, 중재제공자가 임상시험 도중 중재군에 대해 알고 있다. 그리고 (ii) 임상시험의 맥락 때문에 의도한 중재에서 이탈이 있다. 그리고 (iii) 이들 이탈이 중재군간 불균형이 있다.	i)중재배정 효과 추정에 적절한 분석을 사용하지 않았다. 그리고 (ii) 실패가 분석에 대한 잠재적 영향(중재 효과 추정에 대한)이 상당히 크다.	제 1부의 비밀림위험 높음 혹은 제 2부의 비밀림위험 높음

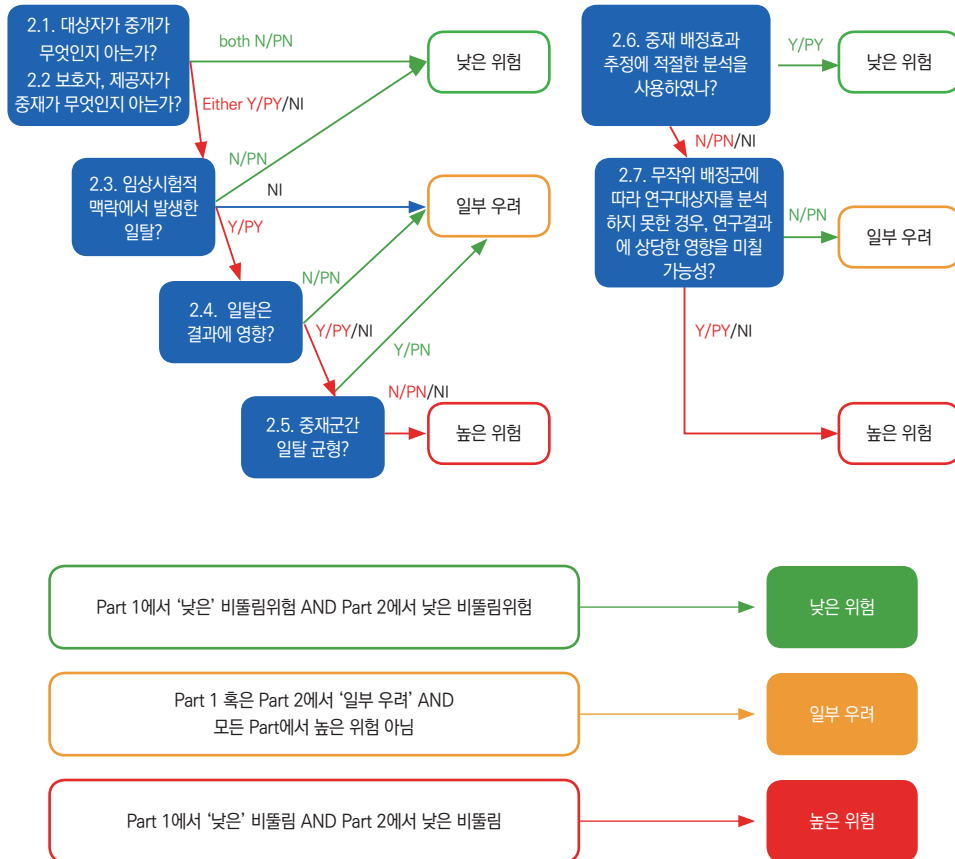
표 2-6. 의도한 중재에서 이탈로 인한 비뚤림(중재 배정 효과) 위험 판단에 대한 신호질문 맵핑

신호질문					영역 수준 판단
제 1부 : 신호질문 2.1 ~ 2.5					
2.1 연구대상자 얇?	2.2 중재제공자 얇?	2.3 이탈?	2.4 중재결과에 영향?	2.5 이탈 균형?	1부에서의 비뚤림위험
Both 2.1 & 2.2 N/PN		NA	NA	NA	낮음
Either 2.1 or 2.2 Y/PY/NI		N/PN	NA	NA	낮음
Either 2.1 or 2.2 Y/PY/NI		NI	NA	NA	일부우려
Either 2.1 or 2.2 Y/PY/NI		Y/PY	N/PN	NA	일부우려
Either 2.1 or 2.2 Y/PY/NI		Y/PY	Y/PY/NI	Y/PY	일부우려
Either 2.1 or 2.2 Y/PY/NI		Y/PY	Y/PY/NI	N/PN/NI	높음
제 2부 : 신호질문 2.6 ~ 2.7					
2.6 적절한 분석?		2.7 분석에서 균의 변경으로 인한 중재결과에 미치는 잠재적인 영향?			2부에서의 비뚤림위험
Y/PY		NA			낮음
N/PN/NI		N/PN			일부 우려
N/PN/NI		Y/PY/NI			높음
영역수준 판단					
제 1부 '낮은 비뚤림위험' AND 제 2부 '낮은 비뚤림위험'					낮음
제 1부 혹은 제 2부에서 '일부 우려' AND 제 1 혹은 2부에서 '높은 비뚤림위험' 아님					일부 우려
제 1 혹은 2부에서 '높은 비뚤림위험'					높음

그림 2-2. 의도한 중재에서 이탈로 인한 비돌림위험 판단 알고리즘(중재 배정 효과)

Part 1 : 문항 2.1에서 2.5까지

Part 2



Box 2-7. ROB 2 도구: 의도한 중재에서 이탈로 인한 비뚤림위험(중재 준수 효과)

신호질문	세부 설명	응답 옵션
2.1 연구대상자는 임상시험 도중 배정된 중재에 대해 알고 있는가?	연구대상자가 배정된 중재를 알고 있다면, 건강 관련 행동이 중재군에 따라 다를 가능성이 크다. 연구대상자에 대한 눈가림은 위약이나 가짜 중재를 통해, 그러한 차이를 방지할 수 있다. 연구대상자가 중재 중 하나에 특정한 것으로 알려진 부작용이나 독성을 경험한 경우, 'Y' 또는 'PY'라고 응답한다.	Y/PY/PN/N/NI
2.2 중재 제공자는 임상시험 도중 연구대상자에 대한 배정된 중재에 대해 알고 있는가?	중재 제공자가 배정된 중재를 알고 있다면, 그 중재를 시행했는지 비프로토콜 중재를 시행했는지 여부가 중재군간에 다를 수 있다. 눈가림은 그러한 차이를 막을 수 있다. 중재제공자가 중재 중 하나에 특정한 것으로 알고 있는 부작용이나 독성을 연구대상자들이 경험한 경우, 'Y' 또는 'PY'라고 응답한다. 무작위 배정 은폐가 되지 않는 경우, 중재제공자는 임상시험 중에 연구대상자에 배정된 중재가 무엇인지 알고 있었을 가능성이 크다.	Y/PY/PN/N/NI
2.3 [해당되는 경우] 2.1과 2.2에 대해 Y/PY/NI 인 경우 : 중재군간 중요한 비프로토콜 중재가 유사한가?	이 질문은 사전 고려사항에서 평가 과정에 '중재군 간의 중요한 비프로토콜 중재의 불균형'을 다룰 것이라고 표시한 경우에만 하도록 한다. 중요한 비프로토콜 중재는 (1) 임상시험 프로토콜과 일치하지 않는 추가 중재 또는 노출, (2) 연구대상자가 배정된 중재와 같이 혹은 그 이후에 받을 수 있는 추가 중재 또는 노출, (3) 중재결과에 대해 예후요인인 추가 중재 또는 노출이다. 중재군간에 중재에 불균형이 있을 경우 비뚤림위험이 더 높아질 것이다.	NA/Y/PY/PN/N/NI
2.4. [해당되는 경우] 중재결과에 영향을 미칠 수 있는 중재 실행의 실패가 있는가?	이 질문은 사전 고려사항에서 평가과정에 '중재 결과에 영향을 미칠 수 있는 중재의 실행 실패'를 다룰 것이라고 표시한 경우에만 하도록 한다. 예를 들어, 치료를 제공하는 보건의료인에 의해 의도한 대로 중재가 실행되지 않은 경우 비뚤림위험이 더 높아질 것이다. 모든 또는 대부분의 연구대상자가 중재 실행에 성공했다면 'N' 또는 'PN'라고 응답한다.	NA/Y/PY/PN/N/NI
2.5. [해당되는 경우] 연구대상자의 중재 결과에 영향을 미칠만한 중재에 대한 비준수가 있는가?	이 질문은 사전 고려사항에서 평가 과정에 '연구대상자의 결과에 영향을 미칠 수 있는 비준수'를 다룰 것이라고 표시한 경우에만 하도록 한다. 비준수에는 지속적인 중재에 대한 불완전한 이행, 중재 중단, 대조군 중재로 교차 이동, 다른 활성 중재로 전환 등이 포함된다. 추적관찰 기간 동안 배정된 중재를 지속적으로 받고 있는 연구대상자 비율에 대해 이용할 수 있는 정보를 고려하고, 비준수 비율이 문제를 제기할 정도로 높으면 'Y' 또는 'PY'라고 응답한다. 불완전한 준수가 불가능한 1회 투여 중재 연구나 모든 혹은 대부분의 연구대상자가 배정된 중재를 받은 경우 'N'라고 응답한다.	NA/Y/PY/PN/N/NI

<p>2.6. 2.3에 대해 N/PN/NI 혹은 2.4나 2.5에 대해 Y/PY/NI 인 경우 : 중재 준수 효과 추정에 적절한 분석 방법을 사용하였나?</p>	<p>‘계획서 순응군 분석’(배정된 중재를 받지 않은 연구대상자는 제외)과 ‘치료한대로 분석’(배정된 중재가 아닌, 받은 중재에 따라 연구대상자가 그룹화되는 경우) 모두 중재 준수 효과를 추정하는데 적합하지 않다. 그러나 무작위 임상 시험 자료를 사용하여 중재에 따르는 효과에 대한 비뚤림없는 추정치를 도출할 수 있다. 적절한 방법의 예로는 (1) 표준치료와 비교된 단일중재 임상시험에서 중재의 영향을 추정하기 위한 도구변수분석(instrumental variable analysis), (2) •지속적인 치료 전략에 대한 임상시험에서 중도절단된 연구대상자에 대해 보정해 주는 역확률 가중치를 사용하고, 의도한 중재로부터 이탈된 대상자에 대한 정보가 있는 경우이다. 이러한 방법은 강력한 가정에 의존할 수밖에 없으므로 응답을 ‘Y’ 혹은 ‘PY’로 하려면 방법 자체가 적절해야 하며 정당화되어야 한다. 중요한 비프로토콜 중재가 한 중재군의 모든 연구대상자에게 시행된 경우, 보정으로 이를 극복할 수 없다. 중재 준수 효과를 추정하는 데 적절하지 않은 분석 전략의 몇 가지 예는 (i) 배정된 대로 분석(ITT), (ii) 계획서 순응군 분석(PP), (iii) 치료한 대로 분석(as-treated analysis) (iv) 받은 치료에 의한 분석(‘analysis by treatment received’) 등이다.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p>비뚤림위험 평가</p>	<p>알고리즘 참조</p>	<p>낮음/높음/일부 우려</p>
<p>선택사항: 의도한 중재에서 이탈로 인한 비뚤림의 예측 방향은 어떠한가?</p>	<p>비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향함, 혹은 임상시험 중재번호, 비교중재 번호로 기술한다.</p>	<p>적용불가 / 임상 시험 중재번호 / 비교중재번호/ 무효점을 향함/ 무효점에서 멀어짐/예측불가</p>

표 2-7. 의도한 중재에서 이탈로 인한 비뚤림(중재준수효과) 판단 기준

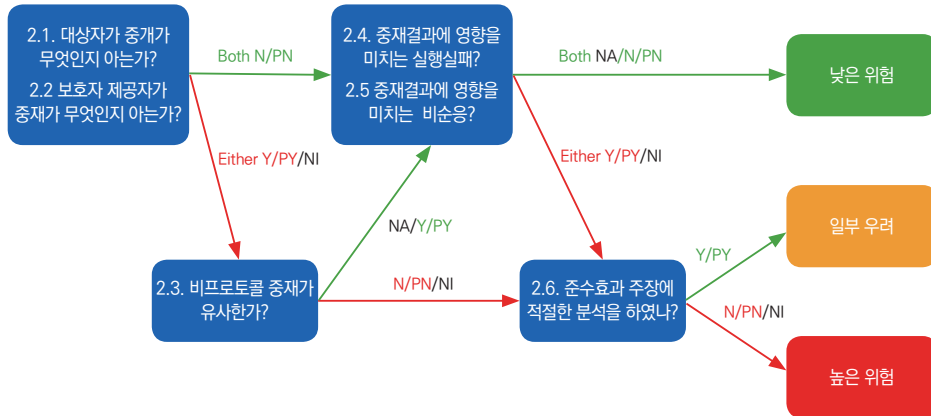
비뚤림위험 낮음	<p>(i.1) 연구대상자, 보호자, 중재제공자가 임상시험 중에 중재군에 대해 모른다. 혹은 (i.2.1) 연구대상자, 보호자, 중재제공자가 임상시험 중에 중재군에 대해 알고 있다. 그리고 (ii.2.2) (해당되는 경우) 중재군간 중요한 비프로토콜 중재가 유사하다. 그리고 (ii) [해당되는 경우] 중재 실행의 실패는 중재결과에 영향을 미칠 수 없다. 그리고 (iii)[해당되는 경우] 연구대상자는 배정된 중재를 준수하였다.</p>
비뚤림위험 일부 우려	<p>(i.1.1) 연구대상자, 보호자, 중재제공자가 임상시험 중에 중재군에 대해 모른다. 그리고 (i.1.2.1) 연구대상자, 보호자, 중재제공자가 중재군을 대해 알고 있었다. 그리고 (ii.2.2) (해당되는 경우) 중재군간 중요한 비프로토콜 중재가 유사하다. 그리고 (ii) [해당되는 경우] 중재 실행의 실패는 중재결과에 영향을 미칠 수 없다. 그리고 (iii)[해당되는 경우] 연구대상자는 배정된 중재를 준수하였다. (i.1.1) 연구대상자, 보호자, 중재제공자가 임상시험 중에 중재군에 대해 모른다. 그리고 (i.1.2.1) [해당되는 경우] 중재 시행 실패가 중재결과에 영향을 미칠 수 있다. 혹은 (i.1.2.2) [해당되는 경우] 연구대상자는 배정된 중재를 준수하지 않았다. 혹은 (i.2.1) 연구대상자, 보호자, 중재제공자가 중재군에 대해 알고 있었다. 그리고 (i.2.2) (해당되는 경우) 중재군간 중요한 비프로토콜 중재가 유사하다</p>
	<p>그리고 (i.2.3.1) [해당되는 경우] 중재 시행 실패가 중재결과에 영향을 미칠 수 있다. 혹은 (i.2.3.2) [해당되는 경우] 연구대상자는 배정된 중재를 준수하지 않았다. 혹은 (i.3.1) 연구대상자, 보호자 또는 중재제공자는 중재군을 알고 있었다. 그리고 (i.3.2) [해당되는 경우] 중요한 비프로토콜 중재가 중재군간 균형을 이루지 않았다. 그리고 (ii) 적절한 분석을 통해 중재의 영향을 추정했다.</p>

<p>비돌림위험 높음</p>	<p>(i.1.1) 연구대상자, 보호자, 중재제공자가 임상시험 도중 중재군에 대해 모른다. 그리고 (i.1.2.1) 중재 실행의 실패는 중재결과에 영향을 미칠 수 있다. 혹은 (ii.1.2.2) [해당되는 경우] 연구대상자는 배정된 중재를 준수하지 않았다. 혹은 (i.2.1) 연구대상자, 보호자, 중재제공자가 중재군에 대해 알고 있었다. 그리고 (i.2.2) (해당되는 경우) 중재군간 중요한 비프로토콜 중재가 유사하다. 그리고 (i.2.3.1) [해당되는 경우] 중재 시행 실패가 중재결과에 영향을 미칠 수 있다. 혹은 (i.2.3.2) [해당되는 경우] 연구대상자는 배정된 중재를 준수하지 않았다. 혹은 (i.3.1) 참가자, 보호자 또는 중재제공자는 중재군을 알고 있었다 그리고 (i.3.2) [해당되는 경우] 중요한 비프로토콜 중재가 중재군간 균형을 이루지 않았다. 그리고 (ii) 적절한 분석을 통해 중재의 영향을 추정하지 않았다.</p>
-----------------	---

표 2-8. 의도한 중재에서 이탈로 인한 비풀림(중재 준수 효과) 위험 판단에 대한 신호질문 맵핑

신호질문						영역 수준 판단
2.1 연구대상자 얇?	2.2 중재제공자 얇 ?	2.3 비프로토콜중 재균형?	2.4 실행실패가 중재결과에 영향?	2.5 비준수?	2.6 적절한 분석?	비풀림위험
Both 2.1 & 2.2 N/PN		NA	N/PN or NA	N/PN or NA	NA	낮음
Either 2.1 or 2.2 Y/PY/NI		Y/PY or NA	N/PN or NA	N/PN or NA	NA	낮음
Both 2.1 & 2.2 N/PN		NA	Y/PY/NI	N/PN or NA	Y/PY	일부우려
Both 2.1 & 2.2 N/PN		NA	N/PN or NA	Y/PY/NI	Y/PY	일부우려
Either 2.1 or 2.2 Y/PY/NI		Y/PY or NA	Y/PY/NI	N/PN or NA	Y/PY	일부우려
Either 2.1 or 2.2 Y/PY/NI		Y/PY or NA	N/PN or NA	Y/PY/NI	Y/PY	일부우려
Either 2.1 or 2.2 Y/PY/NI		N/PN/NI	Any response		Y/PY	일부우려
Both 2.1 & 2.2 N/PN		NA	Y/PY/NI	N/PN or NA	N/PN/NI	높음
Both 2.1 & 2.2 N/PN		NA	N/PN or NA	Y/PY/NI	N/PN/NI	높음
Either 2.1 or 2.2 Y/PY/NI		Y/PY or NA	Y/PY/NI	N/PN or NA	N/PN/NI	높음
Either 2.1 or 2.2 Y/PY/NI		Y/PY or NA	N/PN or NA	Y/PY/NI	N/PN/NI	높음
Either 2.1 or 2.2 Y/PY/NI		N/PN/NI	Any response		N/PN/NI	높음

그림 2-3. 의도한 중재에서 이탈로 인한 비돌림위험 판단 알고리즘(중재 준수 효과)



PART 2

6. 중재결과 자료의 결측으로 인한 비뚤림

6.1. 배경

무작위 배정은 두 개 이상의 중재군간에 알려진 예측 요인과 알려지지 않은 예측 요인의 분포를 기저상태에서 균형 있게 보정하여 공정한 비교를 하게 해준다. 하지만 연구 중 탈락으로 인한 중재 중재결과와 결측은 그런 비교성을 깨트려 중재 효과 추정치에 비뚤림을 초래할 수 있다. 중재결과 자료가 결측되는 이유는 다음과 같다(83).

- 연구대상자가 연구참여를 철회하거나 추적불가능함(‘추적관찰 소실(loss to follow up)’ 또는 ‘탈락(dropout)’)
 - 연구대상자가 결과를 측정해야 하는 연구 방문에 오지 않음(추적관찰 실패, 탈락)
 - 연구대상자가 연구 방문하였지만 관련 자료는 제공하지 않음
 - 자료가 소실되었거나 다른 이유로 사용할 수 없는 경우
 - 연구대상자에게 더 이상 중재결과가 발생할 수 없음(예: 사망)

이 영역은 중재결과 자료 결측 또는 이를 해결하기 위한 결측값 대체법 적용 절차에 의해 발생한 비뚤림의 위험을 다룬다.

일부 연구대상자는 중재결과 자료 결측 이외의 이유로 분석에서 제외될 수 있다. 특히 프로토콜대로(per protocol, PP) 분석은 의도한 중재를 받은 연구대상자로 제한된다(섹션 1.3.1 참조). 그러한 분석에 의해 발생하는 비뚤림위험이나 중재결과 자료를 이용할 수 있는 적격 연구대상자가 다른 이유로 배제되어 발생하는 잠재적 비뚤림은 ‘의도한 중재에서 이탈로 인해 발생하는 비뚤림’(섹션 5 참조) 영역에서 다루어지며, 최종 신도 질문은 분석 접근법이 적절한지 여부를 검토한다. 이는 불완전한 중재결과 자료로 인한 비뚤림위험 영역에 실제로 결측된 자료와 임상시험 연구자가 의도적으로 배제한 자료를 모두 다루는 이전 버전 Cochrane RoB 도구와 특징적으로 달라진 것이다.

6.1.1. 중재 배정 효과를 추정할 때, 중재결과 자료의 결측

중재 배정 효과는 ‘배정된 대로 분석’해야 한다(8). 섹션 1.3.1에서 언급한 바와 같이 배정된 대로 분석의 기본 원칙은 다음과 같다(9, 10).

- ① 실제 받은 중재에 관계없이 무작위 배정된 중재 그룹의 모든 연구대상자를 분석한다.
- ② 분석에는 모든 무작위 연구대상자를 포함하며, 이를 위해서는 모든 연구대상자의 중재 결과를 측정해야 한다.

중재결과 자료를 이용할 수 있는 연구대상자를 대상으로는 첫 번째와 두 번째 원칙을 항상 따를 수 있지만, 모든 연구대상자에 대한 중재결과 자료를 측정하는 것은 종종 어렵거나 실제로 달성하는 것이 불가능하다. 따라서 결과의 결측된 값에 대한 가정을 해야 하는 경우가 많다.

분석을 ITT라고 기술하는 경우에도 중재결과 자료가 결측된 연구대상자를 배제할 수 있으므로 비뚤림위험이 있다. 이러한 분석은 ‘수정된 ITT(modified ITT)’ 분석으로 설명되기도 하지만, 분석마다 이 용어를 서로 다른 방식으로 사용한다(84). 따라서 중재결과 자료 결측으로 인한 비뚤림위험 평가는 이 영역에 대한 신호 질문에서 다루어진 문제에 기초해야 하며, 임상시험 연구자의 분석 기술 방식에 근거해서는 안 된다.

6.1.2 중재준수 효과 추정에 중재결과 자료의 결측

위에서 언급한 바와 같이, 배정된 중재를 준수하지 않은 적절한 임상시험 연구대상자의 배제로 인한 비뚤림 가능성은 ‘의도한 중재에서 이탈로 인한 비뚤림’ 영역에서 다루어진다(섹션 5 참조). 이러한 비뚤림은 중재결과 자료가 결측되어 연구대상자가 제외된 경우 추가로 발생할 수 있다. 연구대상자가 중재결과 자료 결측으로 인해 배제되는 경우, 중재에 따른 효과를 추정하기 위한 적절한 방법(예를 들어, 도구를 이용해 중재의 효과 변수를 추정하는 도구변수분석(5))조차도 비뚤림이 있는 추정치를 산출할 수 있다.

중재결과 자료 결측 때문에 비뚤림이 생기는 상황은 관심의 효과에 관계없이 유사하므로 이 영역에 대한 신호 질문 집합은 하나이다.

6.1.3. 중재결과 자료의 결측이 비뚤림으로 이어지는 시점

중재결과 자료 결측이 있는 연구대상자를 제외한 통계적 분석은 ‘완료 증례(complete case)’ 분석(포함된 변수의 결측 값이 없는 개인에 국한된 분석)의 한 예이다. 그러한 분석이 비뚤림을 초래하는 경우를 이해하려면 다음 사항을 고려해야 한다.

- ① 중재결과 자료가 결측된 연구대상자 **중재결과의 참값**. 이는 측정해야 했지만 측정되지 않은 결과의 값이다.
- ② **결측 여부(missingness) 메커니즘**. 중재결과 자료가 결측되는 과정이다.

중재결과 자료 결측이 ‘완료 증례 분석’에서 비뚤림으로 이어지는지는 결측 여부 메커니즘이 중재결과의 참값과 관련되는지 여부에 따라 달라진다. 만일 참값과 관련되어 있다면 자료는 ‘무작위적으로 결측되지 않음(missing not at random, MNAR)’으로 설명된다(85). 마찬가지로 측정된(비결측) 결과가 결측된 결과(중재결과 자료가 결측된 연구대상자의 참값)와 체계적으로 다른지 고려할 수 있다. 예를 들어, 일반적인 우울증 치료와 비교하여 인지 행동 요법의 효과를 본 임상시험의 경우 더 우울해진 연구대상자들이 후속 조치를 위해 복귀할 가능성이 낮아진다면 우울증 결과가 결측되었는지 여부는 참값에 따라 달라지며, 이는 측정된 우울증 결과가 결측된 우울증 결과의 참값과 체계적으로 다를 것임을 의미한다.

아래에 ‘완료 증례 분석’에서 중재결과 자료 결측이 중재의 추정 중재 효과에 비뚤림을 가져오거나 가져오지 않는 상황을 요약하였다.

- ① 중재 결과의 결측 여부가 **중재결과의 참값**과 무관한 경우, 각 중재군 내에서 중재결과 자료가 결측되면 비뚤림이 생기지 않는다. 이 상황에서 완료 증례는 자료가 결측된 사례도 대표한다. 예를 들어, 우연히 자동 측정 장치가 고장나서 중재결과 자료가 결측된 경우 비뚤림은 일어나지 않을 수 있다.
- ② **중재결과의 결측 여부**가 중재군과 중재결과의 참값에 따라 다를 경우, 중재결과 자료가 결측되면 비뚤림이 초래된다. 예를 들어, 우울증 증상에 대한 위약 대조 임상시험에서 완료 증례 분석의 결과는 (1) 지속적인 우울증 증상을 가진 연구대상자들이 추적관찰 소실될 가능성이 더 높고, (2) 약물에 배정된 연구대상자들의 부작용은 그 약물 투여의 중단을 유발한다.
- ③ 중재결과와 결측 여부는 참값과 관련이 있고, 추가적으로, 임상시험군 중재의 효과가 대조군 중재의 결과와 다를 경우, 아래에 기술된 특별한 경우를 제외하고 중재결과 자료가 결측되면 비뚤림이 발생한다. 예를 들어 위에서 설명한 항우울제 임상시험에서

- (1) 지속적인 우울증 증상을 가진 연구대상자들이 추적관찰 소실될 가능성이 더 높고,
 (2) 약물이 위약과 비교하여 우울증 증상에 영향을 미칠 경우 우울증 증상에 대한 추정 효과는 비뚤림을 유발할 것이다.

- 특별한 경우로서 결측여부가 중재군 배정과 관련이 없다는 점을 전제한다면, 결과 변수가 이분법적이고 중재 효과 추정치가 오즈비일 때 중재결과 자료 결측은 결과의 참값에 따라 달라져도 비뚤림이 일어나지 않는다. 예를 들어, 임상시험과 대조군 중재가 전체 사망률에 미치는 영향을 비교한 임상시험의 오즈비는 비록 중재결과 자료가 사망한 연구대상자들에게서 더 많이 결측될 가능성이 있고, 임상시험군의 사망률이 대조군보다 낮았더라도 중재결과 자료가 결측되어 비뚤림이 일어나지 않을 것이다. 사망위험의 비례감소가 중재군과 대조군 모두에 적용되기 때문에 오즈비 계산에서 취소되기 때문이다. 이 예외는 결과가 수치변수(예: 측정된 혈압에 기초한 고혈압의 분류)에서 파생된 이분형 변수인 경우에는 적용되지 않는다. 많은 이분형 결과는 수치적 변수에서 도출되고, 중재결과 자료가 중재군 배정에 의존하는지 그렇지 않은지의 상황을 구별하기 어렵기 때문에, 이 영역에 대한 신호 질문은 오즈비를 다른 중재 효과 측정치와 구분하지 않는다. 또한 우리는 일반적으로 차이를 배제할 수 없기 때문에(근거의 부재이지 부재함의 근거가 아니다), 우리는 임상시험군 중재의 효과가 대조군 중재의 효과와 다르지 않은 상황을 구별하지 않는다.

- ④ 추가적인 예외도 있는데, 실제로는 식별하기 어렵고, 드물기 때문에 열거되지 않는다.

연구대상자의 하위군(subgroup)간에 중재 효과가 다른 효과 변경(effect modification)의 경우, 위의 고려사항은 하위군 내에서 별도로 적용된다. 결과의 결측 정도가 하위군마다 다를 경우, 분석된 하위군 비율은 원래 무작위 배정된 하위군 비율과 다르기 때문에 전체 중재 효과 추정치가 변경된다. 이 문제는 더 이상 고려되지 않는데, 왜냐하면 시험 사이의 효과변경 분포의 차이는 보통 비뚤림보다는 이질성의 원천으로 간주되기 때문이다.

6.1.3.1 기저상태에서 연구대상자의 특성을 보정하는 분석

일부 임상시험 분석의 경우 기저상태에서 연구대상자의 특성에 대해 보정하므로 기본 특성에 대한 자료가 결측된 연구대상자는 제외한다. 기저상태 특성에 대한 자료 결측으로 인한 제외는 (기저상태 특성에 의한 효과 수정이 없는 경우라면) 추가로 결측여부가 참값에 따라 달라지는 경우에만 비뚤림으로 이어진다. 따라서 그것들은 별도로 고려되지 않는다. 예외는 기저상태 특성이 후향적으로 수집되고 그 결과가 기저상태 특성에 결측여부가 발생하

는 경우로서, 이는 위에서 설명한 참값에 따라 결과의 결측 여부가 달라지는 경우와 동일한 비뚤림에 대한 의미를 갖는다. 예를 들어, 24시간 사망률을 결과로 삼은 새로운 급성 뇌졸중 치료법의 임상시험에서, 의식 회복 전에 사망한 사람들에게 증상의 지속시간과 같은 기저상태 특성을 수집할 수 없었다.

연구대상자 특성을 기저상태에서 고려하여 중재결과 자료 결측으로 인한 비뚤림을 줄이거나 제거할 수 있다. 예를 들어, 결과 변수가 혈압이고, 나이든 사람들과 임상시험 실험 중재군이 모두 임상시험에서 더 많이 탈락할 가능성이 있다고 가정해 보자. 만일 그렇다면 혈압에 대한 중재 추정 효과에 비뚤림이 초래될 것이다. 나이든 사람들은 혈압이 더 높은 경향이 있기 때문이다. 기준 연령을 보정하는 분석을 하게 되면 결과의 결측과 참값 사이의 관계를 완전히 설명하는 경우 비뚤림은 제거될 것이다.

6.1.3.2. 중재결과 자료 결측의 양이 비뚤림을 배제할 수 있을 정도로 작은 경우는?

안타깝게도, 중재결과 자료 결측의 비율과 관련하여 ‘충분히 작음’에 대한 합리적인 임계값은 없다.

중재결과 자료 결측이 비뚤림으로 이어지는 경우, 중재결과 자료 결측의 양이 증가함에 따라 비뚤림의 범위가 증가할 것이다. 5% 미만의 중재결과 자료 결측의 비율을 (비뚤림위험) “작음”으로, 20% 이상을 “큼”으로 평가하는 관례가 있다. 그러나 결측된 자료가 예상 중재 효과에 미치는 잠재적 영향은 결측된 자료가 있는 연구대상자의 비율, 결과의 유형 및 (이분형 결과에 대한) 사건의 위험에 따라 달라진다. 예를 들어, 1000명의 연구대상자가 중재군에 배정된 연구를 생각해 보자. 이중 중재결과 자료가 있는 900명의 연구대상자에 대해 관찰된 사망률이 2%(18명 사망)인 경우, 결측 자료의 비율은 10%에 불과하지만, 100명의 결측 연구대상자의 사망률이 20%(20명 사망)일 경우, 중재군의 전체 진짜 사망률은 관측된 자료에서 추정된 두 배(3.8% 대 2%)에 가까울 것이다.

6.1.4. 중재결과 자료가 결측되어 비뚤림이 발생한 근거를 어떻게 식별할 수 있는가?

결과가 결측될 확률은 참값에 따라 달라지는지 직접 조사할 수 없다. 비뚤림위험의 판단은 임상시험 상황에 따라 달라진다. 따라서 (1) 모든 연구대상자에서 결과를 측정하고 (2) 중재결과 자료가 결측된 연구대상자 수가 충분히 적기 때문에 중재의 추정 효과에 중요한 차이를 두지 않았거나 (3) (임상시험자나 문헌고찰 저자에 의한) 중재결과 자료 결측의 그럴듯한 값이 추정된 중재 효과에 중요한 차이를 만들 수 없음을 확인하는 민감도 분석이 중재

결과 자료 결측로 인한 비뚤림이 없음을 확인할 수 있을 뿐이다.

그러나 중재결과 자료가 결측될 가능성이 있다는 간접 근거는 (1) 실험 중재군 및 대조 중재군에서 중재결과 자료 결측의 비율과 (2) 중재결과 자료가 결측된 이유의 차이에서 얻을 수 있다.

6.1.4.1. 중재결과 자료 결측의 비율의 차이

중재결과에 대한 실험중재 및 대조중재의 영향이 다르고 중재결과와 결측 여부는 참 값에 따라 달라지는 경우, 자료가 결측된 연구대상자의 비율은 중재군 간에 다를 수 있다. 따라서 실험 중재군 및 대조 중재군에서 중재결과 자료 결측의 비율이 다르면 잠재적 비뚤림의 근거를 제공한다.

중재결과 자료 결측의 비율이 중재군 간에 달라도 중재결과 자료가 결측되는 것이 비뚤림을 초래하지 않을 수 있는데, 이는 결과가 결측될 가능성이 참값과 관련이 없는 경우에만 해당한다. 예를 들어, 임상시험에서 대조군보다 중재군에서 임상시험 참여율이 더 높았지만 결측된 자료와 결과와 관련이 없는 경우이다. 중재결과 자료 결측의 비율이 중재군간에 다를 경우, 임상시험 연구결과는 참값과 관련된 결과의 결측에 민감하기 때문에 비뚤림의 위험이 더 크다.

6.1.4.2. 중재결과 자료가 결측된 이유 검토

임상시험 보고서는 연구대상자가 자료를 결측한 이유를 제공할 수 있다. 예를 들어, 치매를 치료하기 위한 할로페리돌 임상시험에서 ‘효능 부족’, ‘부작용 경험’, ‘긍정적 반응’, ‘동의 철회’, ‘환자 사라짐’ 등 다양한 이유를 보고했다. 이들 중 일부(예: ‘효능 부족’과 ‘긍정적 반응’)는 중재결과 자료 결측의 참값과 관련이 있을 가능성이 있다. 따라서 실험 약물과 대조약의 영향이 다르거나, 중재군(예: ‘부작용 경험’)과 관련이 있는 경우 비뚤림의 위험을 증가시킨다.

비뚤림으로 이어질 가능성이 가장 높은 상황은 중재결과 자료가 결측된 이유가 중재군 간에 다를 때, 예를 들어 심각한 상태가 된 연구대상자들이 대조군에서 탈락하는 동안 중재군에서 호전되어 탈락하는 경우이다.

실제로, 비뚤림의 위험을 평가하는 우리의 능력은 시험 연구자들이 중재결과 자료가 결측된 이유를 수집하고 보고하는 정도에 의해 제한을 받을 것이다.

6.1.5. 사건발생 시간(time-to-event) 자료

많은 임상시험에서 ‘사건발생 시간’ 자료를 대상으로 분석한다. 여기서 중재결과는 각 연구대상자에서 중재결과 사건이 관찰되었는지 여부를 나타내는 이분형 변수다. 시간적인 추적관찰은 중재결과가 발생할 때 또는 다른 이유로 관찰이 중단될 때 종료된다. 중재결과가 관찰되기 전에 관찰이 중단된 연구대상자의 추적관찰 시간은 ‘중도절단(censored)’된 시간이라고 한다. 사건발생 시간 자료 분석의 중재효과는 일반적으로 발생비(rate ratio) 또는 위험비(hazard ratio)로 추정한다. 사건발생 시간 자료 분석 결과는 관찰 중도절단이 ‘비정보적(non-informative)’인 경우에만 비뚤림이 없다. 이는 결측된 연구대상자의 관찰 중도절단까지의 시간이 중도절단 이후 결과가 발생할 때까지의 시간과 무관하다는 의미이다. 예를 들어, 모든 연구대상자를 대상으로 특정 날짜까지만 추적관찰 하다가 이후 추적관찰을 중도절단하는 경우 해당 중도절단은 비정보적인 것으로 가정할 수 있다.

정보적 중도절단이란 결과가 관찰되지 않을 확률은 결과의 참값에 따라 결정된다는 것을 의미한다. 예를 들어, 계속 치료에 참여 중인 환자보다 중도탈락된 환자의 사망률이 더 높다면 이는 정보적 중도절단에 해당한다. 정보적 중도절단이 존재하는 상태에서는 다음과 같은 조건에서 사건발생 시간 분석에서 비뚤림이 발생할 수 있다.

- 추적관찰이 중단될 확률이 중재군에 따라 다를 때(예를 들어, 실험군이 심한 부작용으로 탈락가능성이 더 높음)
- 실험군 및 대조군 결과에 미치는 중재의 효과가 다를 때

중도절단율의 차이 또는 중도절단 이유의 차이는 중도절단이 정보적이란 근거를 제공할 수 있다.

예를 들어, 항암 임상시험에서 약물 독성 때문에 연구대상자가 2선 항암 치료로 전환했을 때와 같이 연구대상자가 자신들에게 배정된 중재를 중단하거나 바뀌서 추적관찰이 상실되는 경우 비뚤림위험은 발생한다. 추적관찰 중 연구참여 동의를 철회했을 경우 비록 수집된 자료가 있다 해도 이들의 자료는 중도절단자료로 간주해야 한다. 참고로 CONSORT 흐름도는 이 같은 연구대상자를 분석에 포함된 연구대상자로 보여줄 수 있다.

6.1.6. 중재결과 자료 결측을 처리하는 통계적 방법

임상시험 연구자는 중재결과 자료 결측으로 인한 비뚤림 가능성을 다루려는 통계 분석(완료 증례 분석에 추가로 또는 이를 대신하여)을 제시할 수 있다(83, 87-89). 가장 일반적인

접근방식은 다음과 같다.

- ① 단일 결측값 대치(즉, 하나의 중재결과 결측값을 입력하여 전체 자료셋을 생성)
- ② 다중 결측값 대치(중재결과 변수에 대한 예측 분포를 기반으로 여러 전체 자료셋을 생성)
- ③ 결측된 자료를 보정하기 위해 측정된 중재결과 자료를 가진 연구대상자에게 가중치를 부여하는 방법(90)
- ④ 생존 자료에 대한 우도 기반(likelihood-based) 방법, 모멘트 기반(moment-based) 방법, 반모수(semiparametric) 모델 등 전체 자료셋을 요구하지 않는 방법(83, 91).

결측값 대치 접근방식은 결측된 값을 하나 또는 그 이상의 새로운 값으로 대체한다. 단일 결측값 대치에서는 단 하나의 추정값만을 채워넣는다. 일반적으로 사용되는 접근방식은 '최종 관찰 이월(last observation carried forward, LOCF)'과 '기저상태 관찰 이월(baseline observation carried forward, BOCF)'이다. 이들 방법 모두 결과의 결측이 결과의 참값과 관련 있다면 마지막으로 측정된 후 결과에 변화가 없는 것이 아닌 이상 비뚤림을 제거하지 못할 가능성이 높다. 또한 이 방법들은 결측된 결과에 대한 불확실성을 반영하지 않기 때문에 일반적으로 정밀도를 인위적으로 높인다(따라서 중재 효과 추정의 신뢰 구간이 너무 좁다). 따라서 이러한 방법에 기초한 중재 효과 추정치는 명분이 분명한 경우에만 비뚤림의 위험이 낮은 것으로 간주해야 한다(86, 92).

다중 결측값 대치에서 결측된 결과의 다수의 값은 예측 분포에서 무작위로 도출되어 채워진 다중 자료 셋을 형성한다(93, 94). (단일 결측값 대치 방법과는 달리) 이러한 다중 자료셋을 분석하여 결측된 자료와 관련된 불확실성을 반영하는 단일 요약 추정값 및 신뢰 구간을 생성한다. 그러나 다중 결측값 대치법은 측정된 변수에 의해 그러한 결측이 설명될 수 없는 한, 결과의 결측이 참값에 따라 다를 때 발생하는 비뚤림을 제거하거나 감소시키지 않는다. 특히, 중재결과 자료 결측을 중재군에만 근거하여 결측값을 대치시키면 '완료 증례 분석'과 거의 동일한 결과를 얻을 수 있으며, 중재결과 자료가 MNAR일 때 비뚤림을 줄이거나 제거하지 않는다. 결측값 대치 모델(결과 결측값을 예측하는 데 사용되는 모델)이 올바르게 지정되지 않은 경우, 다중 결측값 대치법은 비뚤림을 제거하지 않으며 심지어 증가시킬 수도 있다. 다중 결측값 대치 추정값은 모델을 구성하는 변수들의 정당성이 만족되면서 결측값 대치 모델에 포함된 측정 변수들이 아닌 결과의 참값과 관계없는 결측의 발생이라는 가정에 정당성이 있는 경우에 비뚤림의 위험이 낮은 것으로 간주해야 한다.

중재결과 값이 시간에 따라 반복적으로 측정된 경우, 이를 중재결과 자료 결측을 예측하

는 데 사용할 경우 중재결과 자료 결측과 관련된 비뚤림을 줄일 수 있을 것이다(95). 그러한 접근법을 사용하는 경우에도 검토자는 추적관찰 소실이 마지막 측정 후의 결과 궤적과 신뢰성있게 관련되는지 신중하게 고려해야 한다.

추적관찰 소실로 인한 비뚤림은 시간 경과에 따른 확률을 모델링하고 탈락한 환자와 유사한 연구대상자에게 더 많은 가중치를 부여하는 중도탈락 확률을 사용하여 해결할 수 있다. 그 목적은 기저상태 이후 연구대상자의 특성이 후속 조치의 탈락과 무관한 가중치 분석을 수행하는 것이다. 이것은 결측된 자료를 처리하는 반모수 접근법의 예다(위의 접근법 (3) 참조). 결측값 대치 기반 접근법과 마찬가지로, 그러한 분석은 가중치를 도출하는 데 사용된 변수를 고려한 후에도 결과의 결측 여부가 참값에 따라 결정된다면 비뚤림을 제거하지 못할 것이다. 가중치는 추적관찰 소실의 확률을 정확하게 지정한 경우에만 비뚤림을 제거할 것이다.

6.1.7. 민감도 분석

민감도 분석을 수행하여 결과의 결측치와 참값 사이의 관계에 대한 가정을 기반으로 중재결과 자료 결측의 잠재적 영향을 평가할 수 있다. 이는 임상시험 연구자나 문헌고찰 저자에 의해 수행될 수 있다. 그러나, 그들은 결과의 결측여부와 참값 사이의 잠재적 관계를 다루는 경우에만 비뚤림의 위험을 판단하는 데 도움이 된다. 6.1.6절에 요약된 방법을 확장하여 그러한 민감도 분석을 수행할 수 있다(96, 97).

문헌고찰 저자는 결측된 자료가 있는 무작위 임상시험에서 얻은 결과의 견고성을 평가하기 위해 몇 가지 방법을 사용할 수 있다(98). 이분형 결과의 경우, 히긴스(Higgins) 등은 가정에 의해 도입된 불확실성을 고려하여 결측된 연구대상자 사이의 사건 위험이 관찰된 연구대상자 사이의 사건 위험과 어떻게 다른지에 대한 다른 가정을 포함하는 전략을 제안한다(86). 에이클(Akl) 등은 결측된 연구대상자와 관찰된 연구대상자의 사건의 상대적 위험에 기반한 히긴스 등의 접근법을 포함하여 간단한 결측값 대치 방법의 집합을 제안한다(99). 연속형 중재결과 자료에도 유사한 아이디어를 적용할 수 있다(100-102). 임상시험 보고서의 보고된 사건 수가 전체 무작위 표본에 적용되는지, 탈락하지 않은 사건에만 적용되는지 불분명할 수 있으므로 이중으로 카운트되는 것을 방지하기 위해 특별히 주의해야 한다(103).

이론적으로 가능한 것의 극한 경계를 명확히 하는 ‘최악의 경우’와 ‘최상의 경우’ 분석을 실시하는 전통이 있지만, 그러한 분석은 가장 그럴듯한 시나리오에 대해 유익하지 않을 수 있다(86).

6.2. 중재결과 자료 결측으로 인한 비뚫림의 경험적 근거

결측된 자료가 임상시험 보고서에서 다루어지는지에 대한 적절성을 조사한 경험적 연구가 시행되었다. 이 연구에서 4개의 의학 학술지의 71개 임상시험 보고서를 검토한 한 연구에서는 자료 결측은 흔히 있으며 통계 분석에서 불충분하게 처리된다고 결론지었다(104).

중재결과 자료 결측에 따른 비뚫림에 대한 우려는 주로 이론적 측면에서 제기되는데, 몇몇 경험적 연구를 통해 결측 자료의 다양한 측면이 효과 추정치의 크기와 연관되어 있는지 여부를 조사했다(18, 34, 74, 79, 105-107). 메타 역학 연구의 체계적 문헌고찰에서, 결측 자료는 일부 연구에서 효과크기의 과대추정과 관련이 있었지만, 나머지에서는 과소추정하거나 차이를 보이지 않았다(33). 대부분의 연구는 중재결과 자료 결측과 분석에서 제외되는 연구대상자를 구분하지 않는다(섹션 5.2 참조). 한 연구결과에서는 중도탈락율이 20%를 넘는 연구와 20% 이하인 연구 사이에 평균 효과추정치의 차이는 없었다(74).

6.3. 이 영역 도구의 사용

- ① 모든 또는 거의 모든 무작위 연구대상자의 중재결과 자료가 제공된다면 비뚫림의 위험은 낮을 것이다. 맥락상 '거의 모두'의 의미란 중재결과 자료가 결측된 연구대상자의 수가 너무 적기 때문에 결과가 어떻든 간에 중재의 추정 효과에 중요한 차이가 없다는 것이다. 만약 그렇다면, 더 이상의 신호 질문에 답할 필요가 없다. 중재결과 자료가 결측된 정도에 대한 정보(예를 들어, 시험 보고서에 CONSORT 흐름도가 제공되지 않은 경우)가 없으면 이 영역에 대한 비뚫림위험이 높다고 판단하게 된다.
- ② 임상시험 연구자 또는 문헌고찰 저자에 의해 수행된 민감도 분석(섹션 6.1.7 참조)에 의해 결과가 중재결과 자료 결측의 그럴듯한 값에 대해 견고하다는 것을 확인한다면 비뚫림의 위험은 낮을 것이다. 이러한 분석은 결측된 자료의 양이 중재의 추정 효과에 대한 잠재적 영향이 상당할 정도로 충분히 큰 경우에 특히 유용할 수 있다. 민감도 분석에 의해 결과가 견고하다는 것이 확인되면, 그 결과는 비뚫림의 위험이 낮은 것으로 간주될 수 있으며 더 이상의 신호 질문에 답할 필요가 없다.
- ③ 섹션 6.1.3에 설명한 바와 같이, 중재결과 자료가 결측된 경우, 결과가 결측될 가능성이 참 값에 따라 달라지는 경우에만 비뚫림으로 이어질 수 있다. 중재결과 자료가 결측된 것으로 보고된 이유에 근거하여 이를 배제할 수 있다(예를 들어, 측정기 고장으로 인한 중재결과 자료 또는 다기관연구에서 특정 센터의 폐쇄로 인한 중재결과 자료만

결측된 경우). 그러나 중재결과의 결측 여부는 참값에 따라 달라질 수 있다면, 검토 연구자들은 중재결과 자료의 비율과 결측 이유를 고려할 필요가 있을 것이다.

- ④ 결측된 예후 자료의 비율에 대한 실험군 및 대조군 간의 차이는 비뚤림의 위험을 나타낼 수 있다(섹션 6.1.4 참조). 사건발생 시간 자료의 경우 문헌고찰 저자는 중도절단(추적관찰 소실로 인한 손실) 비율이 비교군간에 다른지 여부를 고려해야 한다.
- ⑤ 임상시험 연구자가 보고한 결과 자료의 결측 이유 또는 임상시험의 상황 중 하나는, 중재결과의 결측 여부는 참값과 관련 있을 가능성이 높다고 문헌고찰 저자가 결론을 내리도록 유도할 수 있다.
- ⑥ 임상시험 및 대조군에서 중재결과 자료가 결측된 이유가 다를 경우에는 상당한 비뚤림을 초래할 수 있다. 예를 들어, 담배를 끊는 것을 목적으로 한 실험 중재 연구에서, 일부 대조군 연구대상자들이 새로운 중재에 포함되지 않은 것에 대해 실망해서(그리고 계속 담배를 피우면서) 연구를 그만둔 반면, 일부 실험군 연구대상자들은 흡연의 성공적인 중단으로 인해 연구를 그만둔다면 심각한 비뚤림이 있을 것이다.

6.4. 신호 질문 및 비뚤림위험 판단 기준

이 영역에 대한 신호 질문은 box 2-8에 제공된다. 비뚤림위험 판단의 기준은 표 2-11에 제시되어 있으며, 이를 구현하기 위한 알고리즘은 표 2-12와 그림 2-4에 제시되어 있다.

Box 2-8. 중재결과 자료의 결측으로 인한 비뚤림

신호질문	세부 설명	응답 옵션
3.1 현재 중재결과에 대해 모든 혹은 거의 모든 연구대상자의 자료가 이용가능한가?	중재배정 효과 분석(ITT effect analysis)을 위해서는 무작위 임상 시험 연구대상자 전부가 필요하다. “거의 모두”는 결측치가 있는 연구대상자의 수가 충분히 적기 때문에 결과가 어떨든 간에 중재의 추정 효과에 중요한 차이가 없는 것으로 해석해야 한다. 연속변수 중재결과의 경우, 연구대상자의 95%에서 자료를 이용할 수 있으면 충분하다고 할 수 있을 것이다. 이분변수 중재 결과의 경우, 요구되는 비율은 사건의 위험과 직접적으로 연관된다. 관찰된 사건 수가 결측치가 있는 연구대상자 수보다 훨씬 많은 경우, 비뚤림은 매우 적을 것이다.임상시험 보고서에 결측치의 범위에 대한 정보를 제공하지 않는 경우에만 ‘정보 없음’으로 응답한다.	Y/PY/PN/N/NI

	이러한 상황에서는 대개 '결측치로 인한 비뚤림위험'을 높다고 판단하게 될 것이다. 결측치 대체(imputed)된 데이터는 결측치로 간주되어야 하며, 이 질문의 맥락에서는 '중재결과 자료'로 간주되지 않아야 한다.	
3.2 3.1에 대해 N/PN/NI인 경우: 결측치로 인해 연구 결과의 비뚤림이 없다는 근거가 있는가?	결측치로 인해 결과가 편향되지 않았다는 근거는 (1) 비뚤림을 교정하는 분석 방법, 또는 (2) 결측여부와 그 참값 사이의 관계에 대한 그럴듯한 가정 하에 연구결과가 거의 변하지 않는다는 것을 보여주는 민감도 분석에서 비롯될 수 있다. 그러나 최종관찰 이월('last-observation-carried-forward')과 같은 방법 또는 중재군에만 기초한 복수의 결측치 대체는 결측치로 인한 비뚤림위험을 교정하는 것으로 가정해서는 안 된다.	NA/Y/PY/PN/N
3.3 3.2에 대해 N/PN인 경우: 중재 결과의 결측여부가 참값에 의존할 가능성이 있는가?	추적관찰 실패 또는 연구 참여 철회는 연구대상자의 건강 상태와 관련될 수 있다면, 중재 결과의 결측여부는 참값에 의해 영향을 받았을 수 있다. 그러나 모든 결측치가 중재 결과와 무관하다는 것이 문서화 되어 있으면, 결측치로 인한 비뚤림위험은 낮을 것이다(예: 측정 장치의 고장 또는 정규 자료 수집 중단). 사건발생시간 분석에서, 예를 들어 연구 참여 철회로 인해 추적관찰이 중단된 경우는 일부 추적 관찰 자료가 분석에 포함되더라도 결측치로 간주해야 한다. 이러한 연구대상자는 CONSORT 흐름도에는 분석에 포함된 것으로 표시된다.	NA/Y/PY/PN/N/NI
3.4 3.3에 대해 Y/PY/NI: 중재 결과의 결측여부가 참값에 의존할 가능성이 높은가?	이 질문은 (i) 결측여부가 참값에 좌우될 가능성이 있는 상황('일부 우려'로 평가됨)과 (ii) 자료 누락이 참값에 좌우될 가능성이 높은 상황('비뚤림위험'이 높음으로 평가됨)을 구별한다. '예'라고 대답하는 다섯 가지 이유는 다음과 같다. <ol style="list-style-type: none"> 1. 중재군간 결측치 비율의 차이. 실험 중재와 대조 중재가 중재결과에 미치는 영향이 차이가 있고, 결측여부가 참값에 영향을 받으면, 결측치의 비율이 중재군간에 다를 수 있다. 이러한 차이는 결측치로 인한 비뚤림위험을 시사한다. 왜냐하면 시험 결과는 참값에 따른 중재결과에 누락에 민감할 것이기 때문이다. 사건발생시간 분석의 경우, 중도절단 비율(추적관찰 실패)이 중재군마다 다르다는 것이 유사한 점이다. 2. 결측치에 대해 보고된 이유가 결측여부가 참값에 영향을 미친다는 근거를 제공한다. 3. 결측치 이유로 보고된 내용이 군간에 다르다. 4. 임상시험의 정황으로 보아 결측여부는 참값에 의존할 가능성이 높다. 예를 들어, 정신분열증을 치료하기 위한 중재 임상시험에서 증상이 지속되는 경우 탈락이 더 많다는 사실은 널리 알려져 있다. 5. 사건발생시간 분석에서, 연구대상자가 약물 독성 때문에 또는 암 임상시험에서 2차 화학 요법으로 전환할 때, 배정된 중재를 중지하거나 변경하는 경우 중단이 발생하는 경우 결측여부와 참값 사이의 관계를 설명하는 것이 연구대상자 특성인 경우 '아니오'에 답한다. 	NA/Y/PY/PN/N/NI

비뚤림위험 평가	알고리즘 참조	낮음/일부우려/높음
선택사항: 중재결과 자료의 결측으로 인한 비뚤림의 예측 방향은 어떠한가?	비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향함, 혹은 실험 중재 선호, 비교중재 선호로 기술한다.	적용불가 / 실험 중재선호 / 비교중재선호/무효점을 향함/ 무효점에서 멀어짐/예측불가

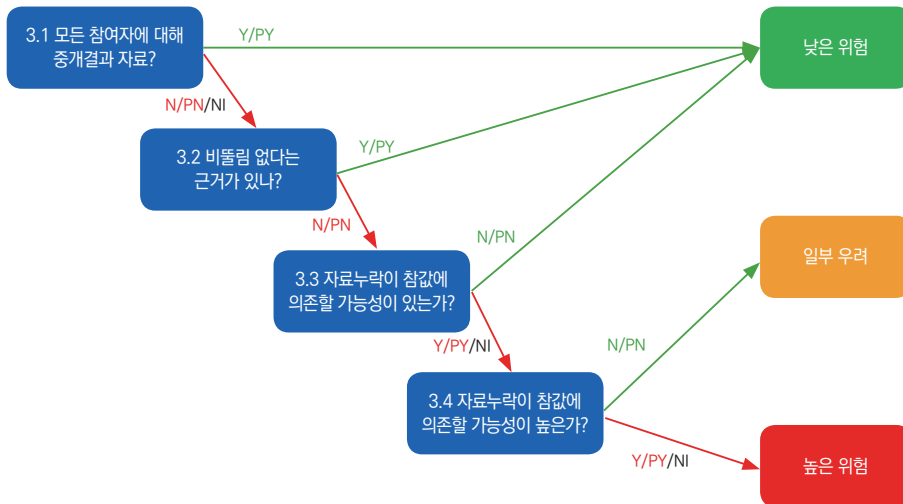
표 2-9. 중재결과 자료의 결측으로 인한 비뚤림 판단 기준

비뚤림위험 낮음	<ul style="list-style-type: none"> (i) 중재 결과 자료가 연구대상자 모두 혹은 거의 모두에서 이용 가능하다 혹은 (ii) 연구 결과가 중재결과 자료 결측에 의해 편향되지 않는다는 근거가 없다 혹은 (iii) 중재결과 자료 결측여부는 참값에 좌우되지 않는다
비뚤림위험 일부 우려	<ul style="list-style-type: none"> (i) 중재 결과 자료가 연구대상자 모두 혹은 거의 모두에서 이용 가능하지 않다 그리고 (ii) 연구 결과가 중재결과 자료 결측에 의해 편향되지 않는다는 근거가 없다 그리고 (iii) 중재결과 자료 결측여부는 참값에 좌우될 수 있다 그리고 (iv) 결과자료의 결측여부는 참값에 좌우될 것 같지는 않다.
비뚤림위험 높음	<ul style="list-style-type: none"> (i) 중재 결과 자료가 연구대상자 모두 혹은 거의 모두에서 이용 가능하지 않다 그리고 (ii) 연구 결과가 중재결과 자료 결측에 의해 편향되지 않는다는 근거가 없다 그리고 (iii) 중재결과 자료 결측여부는 참값에 좌우될 수 있다 그리고 (iv) 결과자료의 결측여부는 참값에 좌우될 것 같다.

표 2-10. 중재결과 자료의 결측으로 인한 비밀리위험 판단에 대한 신호질문 맵핑

신호질문				영역 수준 판단
1.1 완전한 자료?	1.2 비밀리없다는 근거?	1.3 참값에 의존?	1.4 참값에 의존할 가능성 높음?	비밀리위험
Y/PY	NA	NA	NA	낮음
N/PN/NI	Y/PY	NA	NA	낮음
N/PN/NI	N/PN	N/PN	NA	낮음
N/PN/NI	N/PN	Y/PY/NI	N/PN	일부우려
N/PN/NI	N/PN	Y/PY/NI	Y/PY/NI	높음

그림 2-4. 중재결과 자료의 결측으로 인한 비밀리위험 판단 알고리즘



7. 중재결과 측정의 비뚤림

7.1 배경

연구대상자의 중재결과 변수 측정 오류는 측정값이 참값 또는 기초값(underlying value)과 같지 않을 때 발생한다. 이 같은 오류는 무작위 임상시험의 중재 효과 추정에 비뚤림을 유발할 수 있다. 이러한 오류는 연속형 결과의 경우에는 측정 오류, 이분형 또는 범주형 결과의 경우에는 오분류(misclassification), 사건인 경우에는 과소확인/과다확인(under/over-ascertainment)으로 표현된다. 측정 오류는 중재 배정에 차등적일 수도 비차등적일 수도 있다.

- 차등적 측정 오류는 중재 배정과 관련이 있다. 그러한 오류는 실험군과 대조군 간의 체계적인 차이로 나타나며, 결과 평가자가 중재 배정에 눈가림되어 있다면 발생 가능성이 낮다.
- 비차등적 측정 오류는 중재 배정과 관련이 없는 오류이다.

이 영역은 차등적 오류와 관련이 있다. 비차등적 오류는 중요한 것으로 간주되지만 구체적으로 다루지 않는 경우, box 2-9에 설명된 이유로 해결할 수 있다.

Box 2- 9. 중재 효과 추정치에 대한 비차등적 측정 오류의 의미

중재 배정과 무관한 중재결과의 측정 또는 분류 오류를 '비차등적' 오류라고 한다. 그것은 일반적으로 중재군간 연속 변수의 평균 차이를 보는 중재 효과 추정치에 비뚤림을 일으키지는 않지만, 중재결과가 이분 변수이거나 범주 변수일 때 오즈비, 위험비, 상대위험비 등과 같은 중재 효과 추정치를 감소시킬 가능성(즉 무효점을 향하게 하는 비뚤림)이 높다. 비차등 측정 오류가 효과 추정치를 무효점에서 벗어나게 하는 방향일 상황이 있지만, 이러한 상황은 보통 무작위 시험에서는 발생할 가능성이 낮은 것으로 간주된다.

비차등 측정 오류가 중재결과에 미치는 영향은 측정 오차의 성격에 따라 달라진다. 예를 들어

- 효과 추정치가 (평균 차이와 같은) 차이 측정치이고, 측정 오류가 중재결과에 부가적으로 작용한다면 비차등 오류로 인한 비뚤림이 없을 것이다. 예를 들어, 양쪽 중재군에 사용된 혈압 측정

장치가 체계적으로 10 mmHg 높은 측정치를 생성한다면(부가 오류), 이 오류는 실험군 및 대조군 중재 그룹의 평균값과 평균 간의 차이를 계산할 때 동일하게 적용된다.

- 효과 측정치가 차이 측정치 (1) 측정 오차가 중재결과에 곱셈으로 작용하고 (2) 중재 효과가 0이 아닌 경우 비차등 오류로 인한 비뚤림이 있을 것이다. 예를 들어, 혈압 측정 장치가 참값의 120%에 해당하는 측정치를 체계적으로 생성하는 경우(승수 오차)
 - 중재 효과가 없을 경우 평균 차이에 편향은 발생하지 않는다(정확히 0이 나옴).
 - 중재 효과가 있을 경우 평균의 차이는 20% 초과 추정된다(예: 실제 평균 100 및 110 mmHg(차이 10 mmHg)는 120 및 132 mmHg(차이 12 mmHg)로 측정된다).
 - 연구대상자를 고혈압환자와 정상으로 분류하기 위해 혈압 측정을 하는 경우, 위험 비 또는 오즈비 추정치는 중재 효과가 있는 경우 비뚤림이 있을 수 있다.

연속 변수의 경우 측정 오류는 일반적으로 곱셈이 아닌 덧셈으로 가정한다는 점에 유의한다.

- 이분형 변수 중재결과의 경우, 비차등적 측정 오류의 영향은 다른 효과 측정치(즉, 오즈비, 상대위험비, 위험도 차이)에 대해 다르다. 연속 변수 중재결과에 대해, 중요한 고려사항은 오차가 효과 측정과 '동일' 방식으로 작용하는지 여부다. 자세한 논의는 Rothman 등의 문헌(108)을 참조한다.

이 영역의 첫 번째 신호 질문은 중재결과 측정 방법이 부적절했는지를 묻고, 이것을 비차등적 측정 오류가 특히 중요한 상황을 강조하기 위해 사용할 수 있다. 비차등적 측정 오류를 더 자세히 다루면 상당히 복잡해질 수 있다. 더욱이 의미 있는 평가는 임상시험 1차 보고서가 중재결과 및 측정 방법을 명확히 구별하고, 오류가 선택된 중재 효과치와 '합치'(congruent)하는지에 대한 판단을 할 수 있도록 참값과 측정값 사이의 관계를 고려해야 한다. 예를 들어, 고혈압은 보통 혈압 측정을 하나 이상 기준으로 진단되는데, 각각 측정 오류의 대상이 된다. 단, 연구대상자가 추적 관찰 종료 후 고혈압 진단을 받았는지를 기준으로 한 임상시험 중재결과는 일반적으로 비차등적 측정 오류로 인해 본질적으로 편향되지 않은 것으로 간주된다.

이 영역의 비뚤림위험 고려사항은 다음과 같다.

- (1) 적절한 중재결과 측정 방법
- (2) 중재군 간에 다르거나 다를 수 있는 중재결과의 측정 또는 확인
- (3) 중재결과 평가자
- (4) 중재결과 평가자의 중재 배정에 대한 눈가림
- (5) 제공받은 중재에 대한 지식이 중재결과 평가에 미칠 가능성

(1) 무작위 임상시험의 중재결과는 적절히 평가되어야 한다. 예를 들어, 인슐린 중재와 위약을 비교한 임상시험에서 연구대상자가 사용하는 휴대용 혈당계가 3.1mmol 미만의 수치를 신뢰할 수 있게 측정할 수 없는 것이라면 그 혈당계는 심각한 저혈당증 비율의 차이를 감지할 수 없을 것이며, 결과적으로 이 부작용의 실제 발병률을 과소추정하게 될 것이다. 따라서 이런 측정 방법은 이 중재결과에 적합하지 않을 것이다. 반대로 해당 혈당계는 낮은 타당도로 인해 관심 중재결과 변수를 적절하게 측정하지 못한다는 것이 입증되었을 수 있다.

(2) 중재결과는 중재군 간에 비교할 수 있는 방법을 사용하여 측정하거나 확인해야 한다. 이는 일반적으로 사전에 정의된 중재결과이다. 그러나 중재결과 자료를 수동적 수집하게 되면 문제가 생길 수 있는데 예상치 못한 부작용 등에서 그런 문제가 생길 수 있다.

- 위약-대조 임상시험에서 위약에 배정된 연구대상자들보다 신약에 배정된 연구대상자들에게 심한 두통이 더 자주 발생할 수 있다. 이러한 두통은 실험 중재군에서 더 많은 MRI 스캔을 하게 할 수 있고, 따라서 약이 뇌종양의 발생률을 증가시키지 않더라도 증상을 더 많이 진단하게 할 수 있다. 이는 중재결과가 뇌종양의 존재로 정의되는 경우(중재결과가 뇌종양의 진단으로 정의되는 경우는 아님) 비뒤림으로 이어질 수 있다.
- 클레멘스(Clemens) 등은 결핵에 대한 BCG 백신의 예방 효과 시험에서 이른바 “진단 검사 비뒤림(diagnostic test bias)”의 가능성을 확인했다(109). BCG 예방접종은 보통 쉽게 식별되는 흉터를 발생시키며, 수동적인 추적검사를 통해서만 결핵을 확인하는 경우 흉터가 있는 연구대상자가 의료기관을 찾을 가능성이 적거나 평가자가 방사선 영상 검사를 지시할 가능성이 적다. 체계적 문헌고찰 결과, 비뒤림의 위험이 더 높은 것으로 평가된 임상시험에서 추정된 결핵 예방효과가 높았다(110).

그러한 비뒤림에 대해 Sackett은 다음과 같이 말하였다. “무고한 노출은 질병을 일으키는 것이 아니라 질병 확인을 촉진하는 징후나 증상을 유발하는 경우 의심될 수 있다”(111).

사전에 정의된 중재결과 측정이라 할지라도 중재의 특성상 중재결과 측정에 중재군 간에 공평하지 않을 수 있다. 예를 들어, 의료인을 추가로 방문하는 것이 중재의 내용에 포함된다면 대조군에 비해 중재군에서 중재결과 확인이 더 빈번하게 일어날 기회가 많아질 것이다.

(3) 중재결과 평가자는 다음과 같은 사람이 해당될 수 있다.

- 중재결과가 고통, 삶의 질 또는 우울증, 불안 또는 기능을 평가하는 자기기입 설문지와 같이 연구대상자가 보고하는 중재결과일 때에는 연구대상자
- 중재결과가 임상 검사 결과, 임상적 사건의 발생, 수술 중재의 제공이나 환자의 퇴원과

같은 치료적 의사결정의 결과일 때에는 중재 제공자

- 심사 위원회, 자동화된 검사를 수행하는 생물학자, 의료기록이나 질병 레지스트리에 포함여부를 결정하는 의료인처럼 연구대상자에게 제공되는 중재에 직접적으로 관여하지 않는 관찰자인 중재결과 평가자

(4) 중재결과 평가자의 눈가림은 임상시험 연구대상자와 임상시험 인력에게 눈가림이 가능하지 않은 경우에도 종종 가능할 수 있으며 실제로 종종 이루어진다. 그러나 연구대상자가 보고하는 중재결과의 경우 특히 어렵다. 예를 들어, 3개월째 연구대상자가 느끼는 통증을 수술과 약물치료를 비교하는 임상시험에서 중재결과평가자(연구대상자)를 눈가림하는 것은 불가능하다. 중재결과 평가자의 눈가림이 불가능한 것이 중재결과 비뚤림의 가능성을 무시할 수 있다는 것을 의미하는 것은 아니다. 문헌고찰 저자는 항상 중재결과 측정의 오류로 인한 비뚤림위험을 평가해야 한다.

(5) 중재결과 평가자가 눈가림되지 않은 임상시험의 경우, 중재결과에 대한 평가가 수행된 중재의 지식에 의해 영향을 받을 가능성이 있는지의 여부는 관찰자의 선입견과 중재결과 평가에 관련된 판단 정도에 따라 달라진다. 후자는 중재결과의 유형에 따라 달라지는데, 일부 중재결과는 자의적 판단의 여지가 없거나 적지만(예: 모든 원인 사망률(all-cause mortality)) 다른 중재결과는 자의적 판단의 여지가 상당히 있다(예: 우울증 척도 평가). 우리는 다음과 같이 다섯 가지 다른 유형의 중재결과를 구별한다.

① 연구대상자 보고 중재결과(participant-reported outcomes)

연구대상자 보고 중재결과는 임상 의사 또는 연구자가 연구대상자의 반응을 해석하지 않고 건강 상태 및 특정 치료법과 관련하여 연구대상자가 어떻게 효과를 보았고 느끼는지에 대해 연구대상자가 직접 중재결과를 보고하는 것이다. 연구대상자 보고 중재결과에는 인터뷰, 자기기입 설문지, 환자일지 또는 휴대용 장치와 웹 기반의 자료 수집 도구를 통해 연구대상자로부터 직접 획득한 중재결과 평가가 포함된다(112). 그 예로는 고통, 메스꺼움, 그리고 건강 관련 삶의 질이 있다.

눈가림된 인터뷰 진행자가 연구대상자에게 질문하고 그들을 대신하여 설문지를 작성하는 경우에도 여기서의 중재결과 평가자는 연구대상자다. 인터뷰 진행자는 엄격한 의미에서 중재결과 평가자로 간주되지 않고 오히려 측정을 돕는 자로 간주된다.

연구대상자 보고 중재결과의 경우, 중재결과에 대한 평가는 받은 중재에 대한 지식에 의

해 잠재적으로 영향을 받아 최소한 ‘일부 우려’로 평가를 받는다. 문헌고찰 저자는 연구대상자 보고 중재결과가 수행된 중재에 대한 지식에 의해 영향을 받았을 가능성이 높은지 판단해야 하며, 이 경우 비뚤림의 위험이 높다고 간주된다. 예를 들어, 중재 시작 얼마 후 기록된 심각하거나 예기치 않은 부작용은 받은 중재에 대한 지식에 의해 영향을 받지 않을 것으로 간주될 수 있다. 반면 침술과 무치료를 비교한 연구에서 연구진행 과정 마지막에 보고된 고통의 정도는 받은 중재에 대한 지식에 의해 영향을 받을 가능성이 높다.

② 판단과 무관한 관찰자 보고 중재결과(observer-reported outcomes not involving judgement)

이러한 중재결과는 관찰자의 판단을 수반하지 않는 외부 관찰자(예: 영상의학과 전문의 같이 연구대상자에게 제공되는 치료에 관여하지 않는 독립적 연구자 또는 의사)가 보고한 결과물이다. 모든 원인 사망률 또는 자동화된 시험의 중재결과를 예로 들 수 있다.

여기서의 중재결과 평가자는 관찰자이다. 판단을 수반하지 않는 관찰자 보고 중재결과의 경우, 중재결과 평가는 일반적으로 수행된 중재 지식에 의해 영향을 받지 않을 가능성이 높다.

③ 일부 판단이 수반되는 관찰자 보고 중재결과(observer-reported outcomes involving some judgement)

이것들은 임상 검사에 관여하는 것과 같이 일부 판단을 수반하는 외부 관찰자(예: 중재 제공자)에 의해 보고된 중재결과들이다. 예를 들면, 방사선 촬영, 임상적 관찰, 사망을 제외한 임상적 사건(예: 심근경색) 등 의무기록에 기초한 판단을 포함하는 임상시험이 있다.

여기서의 중재결과 평가자는 관찰자이다. 관찰자가 수행된 중재를 인지하는 경우, 중재결과 평가는 잠재적으로 이 지식에 의해 영향을 받을 수 있어 최소한 ‘일부 우려’로 판단하게 된다. 문헌고찰 저자는 중재결과 평가가 수행된 중재의 지식에 의해 영향을 받았을 가능성이 있는지 판단해야 하며, 이 경우 비뚤림의 위험이 높다고 간주된다.

④ 중재 제공자가 내린 결정을 반영하는 중재결과(outcomes that reflect decisions made by the intervention provider)

이러한 중재결과는 중재 제공자의 결정을 반영한다. 이 결정의 기록에는 어떠한 판단도 수반되지 않는다. 그러나 그 결정 자체는 받은 중재에 대한 지식에 의해 영향을 받을 수 있다. 예를 들어, 복강경 대 최소절개 담낭절제술이 재원일에 미치는 영향을 비교하는 임상시험에서, 연구대상자의 퇴원 결정이 연구대상자의 임상 평가에만 영향을 미치도록 하기 위해 의료진이 연구대상자가 받은 중재에 눈가림을 하는 것이 필수적이었다. 중재 제공자 의사결

정 중재결과의 일반적인 예로는 입원, 치료 중단, 타 병동 이동, 제왕절개 수행, 기계환기 중단, 퇴원 등이 있다. 여기서의 중재결과 평가자는 의사 결정권을 가진 의료 제공자다. 의료 제공자가 이러한 사실을 알고 있다면, 중재결과 평가는 일반적으로 수행된 중재에 대한 지식에 의해 영향을 받을 가능성이 높다. 이는 특히 임상시험 중재의 효과에 관한 선호도, 기대치 또는 예감이 강할 때 중요하다.

⑤ 복합 중재결과(composite outcomes)

복합 중재결과는 여러 개의 중재결과를 하나의 중재결과로 결합한다. 일반적으로 특정 중재결과의 집합 중 하나를 경험한 연구대상자는 복합 중재결과를 경험한 것으로 간주한다. 대표적인 예로는 주요 심장 및 뇌혈관 유해사건(major adverse cardiac and cerebrovascular events, MACCE)이 있다. 복합 중재결과는 연속형 변수 측정으로도 구성할 수 있다.

복합 중재결과에 대한 비뚤림위험 평가는 복합 중재결과의 각 구성요소의 빈도나 기여도를 고려해야 하며 가장 영향력 있는 구성요소로 인한 비뚤림위험을 고려해야 한다.

7.2. 중재결과 측정에서 비뚤림의 경험적 근거

중재결과 측정에서 비뚤림의 경험적 근거는 주로 “이중 눈가림(double blinding)”이 중재 효과와 관련이 있는지를 탐구하는 연구에서 나온다. 이러한 연구는 섹션 5.2에 요약되어 있다. 무작위 임상시험에서 중재결과 평가자의 눈가림이 없는 경우 평균적으로 중재의 오즈비가 34% 더 과다추정 된다.

7.3. 이 영역 도구의 사용

이 영역의 첫 번째 질문은 드물긴 하지만 중재결과 측정 방법이 부적절하지는 않았는가를 식별하기 위한 선별 질문이다. 이것은 무작위 임상시험에서 사전에 지정된 중재결과의 경우에서 이 같은 일이 일어날 가능성은 낮지만, 특히 부작용을 정의하거나 이의 발생을 식별하기 위한 시스템이 마련되어 있지 않은 경우에는 해당될 수 있다. 이 질문은 평가 대상 중재결과의 선택이 합리적이었는지를 평가하기 위한 것이 아니다(예: 주요 관심중재결과에 대한 적절한 대리 중재결과).

두 번째 질문은 중재결과의 측정 또는 확인이 중재군간에 다를 수 있었는지를 다룬다. 대부분의 무작위 임상시험에서 이를 비교할 수 있다. 이 질문은 예를 들어, 중재군이 의료인과 더 많이 만나서 대조군보다 중재결과를 확인할 수 있는 기회가 더 많았기 때문에 군간에 체계적인 차이가 있었던 상황 또는 진단 결과 확인 비뚤림(diagnostic detection bias)을 식별하는 데 목적이 있다(섹션 7.1 참조).

다음 질문은 중재결과 평가에 대한 눈가림과 잠재적 영향을 다룬다. 중재결과 평가가 중재 배정에 눈가림되었는지 여부를 결정하는 것이 중요하다. 눈가림이 성공적으로 구현되었다면, 차등적 측정 오류로 인한 비뚤림의 위험이 낮다. 연구대상자와 자료 수집자 사이의 상호작용을 사용하여 평가되는 중재결과(예: 자료 수집자가 질문을 하거나 연구대상자가 설문지를 작성하도록 도울 때) 연구대상자가 눈가림되고 자료 수집자가 눈가림하지 않거나 반대로 연구대상자가 눈가림되지 않고 자료 수집자가 눈가림된 경우라면 설득력 있는 근거가 제공되지 않는 한 평가자는 수행된 중재를 인지하는 것으로 간주해야 한다.

중재결과 평가자의 눈가림 부재의 중요성은 배정된 중재에 대한 지식에 의해 평가가 영향을 받을 수 있는 정도에 따라 달라진다. 모든 원인 사망률과 같은 중재결과의 경우, 중재결과 평가는 영향을 받았을 가능성이 없지만, '호전(improvement)의 임상적 판단'과 같은 주관적 중재결과의 경우, 받은 중재에 대한 지식이 큰 영향을 미칠 수 있다.

중재결과 평가자가 눈가림하지 않고 중재결과가 수행된 중재에 대한 지식에 의해 영향을 받을 수 있었을 때, 고찰 저자는 그러한 영향력이 발생했을 가능성이 있는지 평가해야 한다. 그렇게 함으로써, 그들은 중재결과 평가자에 미치는 의식적 및 잠재의식적 영향을 모두 설명해야 한다('관심'과 '예상' 모두). 이는 다음에 따라 달라질 수 있다.

- 대조군(대조군에 행해지는 중재가 아무 것도 없거나 통상적인 중재일 경우 또 다른 적극적 중재를 대조군에 시행한 것에 비해 비뚤림위험이 높아진다. 대조군에도 적극적 중재가 시행되는 경우 실험군과 대조군 모두에 행해지는 적극적 중재 중 하나가 다른 중재보다 더 유익하다고 생각하는 사전 믿음을 갖지 못할 수 있기 때문이다).
- 중재결과 평가자가 연구대상자 치료에 참여(중재결과 평가자가 치료효과 입증에 이해관계가 없는 독립 연구자인 경우 비뚤림위험이 낮음).
- 다른 대상자의 영향(예를 들어, 면접을 통해 기록된 연구대상자 보고 중재중재결과의 경우, 면접자가 독립적인 연구자인 경우 면접자가 치료 제공자일 때와 비교해서 비뚤림의 위험이 더 낮을 수 있다)

중재결과 평가 시점 또한 수행된 중재에 대한 지식이 중재결과 평가에 영향을 미치는 것과 마찬가지로 영향을 미칠 수 있다. 예를 들어, 감독 하에 이루어진 12주간의 운동과 충격과 치료를 비교한 임상시험에서 평가시점이 무작위 배정 후 1년이라면, 이루어진 중재에 대한 지식은 그다지 중요하지 않을 수 있다(114). 이것은 특히 그 사이에 연구대상자들에게 많은 다른 치료법이 사용되었기 때문에, 연구대상자들의 회상이 처음에 받은 치료법에 영향을 받지 않을 가능성이 높은 경우가 될 수 있다.

7.4. 신호 질문 및 비뿔림의 위험성 판단 기준

이 영역에 대한 신호 질문은 box 2-10에 제공된다. 비뿔림위험 판단을 위한 기준은 표 2-11에 제시되어 있으며, 이를 구현하기 위한 알고리즘은 표 2-12와 그림 2-5에 제시되어 있다.

Box 2-10. 중재결과 측정의 비뿔림위험

신호질문	세부 설명	응답 옵션
4.1 중재결과 측정 방법이 부적절한가?	이 질문은 평가하고자 하는 중재 결과에 적합하지 않은 결과 측정 방법(자료 수집)을 확인하는 것이 목적이다. 이 질문은 평가 되는 중재 결과의 선택이 합리적이었는지를 평가하기 위한 것이 아니다(예: 관심 주요 결과에 대한 대리결과 또는 중간결과). 대부분의 경우 사전 지정된 중재 결과에 대해 이 질문에 대한 응답은 'N' 또는 'PN'이 될 것이다. 예를 들어 다음과 같은 이유로 결과를 측정하는 방법이 적절하지 않은 경우 'Y' 또는 'PY'로 응답한다. (1) 중재효과를 확인하기 어려운 측정방법 사용(예: 중재 결과 값의 중요한 범위가 측정 방법을 사용하여 확인 가능한 수준을 벗어나는 경우) (2) 측정기기의 타당도가 불량하다는 것이 입증된 경우	Y/PY/PN/N/NI
4.2 중재기간 중재결과 측정 혹은 확인 방법이 다른가?	비교 가능한 중재결과 측정 방법(자료 수집)이란 유사한 시점에 동일한 측정 방법과 임계치로 측정하는 것이다. 중재군 간의 차이가 발생하는 경우는 중재결과 자료의 수동적 수집이라는 맥락에서 '진단 결과 확인 비뿔림' 때문이거나, 중재에 의료 제공자를 추가로 방문하는 것이 포함되어서 중재결과 사건을 확인할 수 있는 기회가 추가되는 경우이다.	Y/PY/PN/N/NI

<p>4.3 4.1과 4.2에 대해 <u>NA/PN/NI</u> 인 경우: 중재결과 평가자는 연구대상자가 받은 중재를 알고 있는가?</p>	<p>중재 결과 평가자가 중재 상태에 눈가림되어 있으면 'N'로 응답한다. 연구대상자 보고 결과의 경우, 중재 결과 평가자는 연구대상자이다.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p>4.4 4.3에 대해 <u>Y/PY/NI</u> 인 경우: 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 있는가?</p>	<p>배정된 중재에 대한 지식은 연구대상자 보고 결과(통증 수준 등), 일부 판단이 포함된 관찰자 보고 결과 및 중재 제공자의 의사결정 결과에 영향을 미칠 수 있다. 판단을 수반하지 않는 관찰자가 보고한 중재 결과(예를 들어, 모든 원인 사망률과 같은)에 영향을 미칠 가능성은 낮다.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p>4.5 4.4에 대해 <u>Y/PY/NI</u> 인 경우: 중재결과에 대한 평가가 중재에 대한 지식에 영향을 받았을 가능성이 높은가?</p>	<p>이 질문은 (i) 중재 상태에 대한 지식이 결과 평가에 영향을 미칠 수 있었지만 그렇게 했다고 믿을 이유는 없는 상황('일부 우려'로 평가)과 (ii) 중재 상태에 대한 지식이 결과 평가에 영향을 미칠 가능성이 있는 상황('높음'으로 평가됨)을 구별한다. 중재가 유익할거라는 혹은 유해할 거라는 믿음이 강할 때, 중재 결과가 받은 중재에 대한 지식에 영향을 받았을 가능성이 더 높다. 동종요법 임상시험에서 환자 보고증상이나, 중재를 제공한 물리치료사가 기능회복에 대한 평가를 시행하는 경우 등이 예가 될 수 있다.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p>비뚤림위험 평가</p>	<p>알고리즘 참조</p>	<p>낮음/높음/일부 우려</p>
<p>선택사항: 중재결과 측정의 비뚤림의 예측 방향은 어떠한가?</p>	<p>비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향함, 혹은 실험 중재선호, 비교중재 선호로 기술한다.</p>	<p>적용불가/실험 중재선호/비교중재 선호/무효점을 향함/무효점에서 멀어짐/예측불가</p>

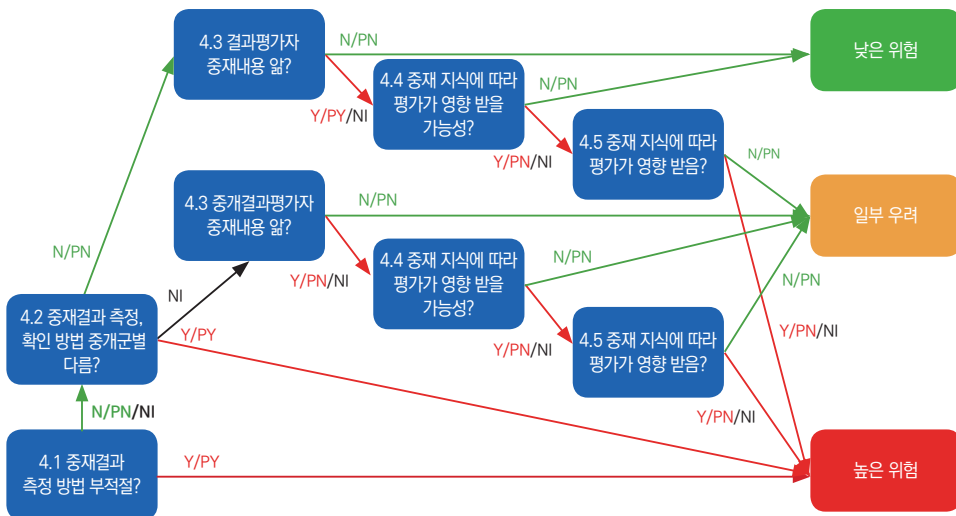
표 2-11. 결과자료 측정의 비돌림위험 판단 기준

<p>비돌림위험 낮음</p>	<p>(i) 중재 결과 측정 방법이 적절하다. 그리고 (ii) 중재군간 중재결과 측정 혹은 확인 방법이 다르지 않다. 그리고 (iii) 1) 중재결과 평가자는 연구대상자가 받은 중재를 알지 못한다. 혹은 (iii) 2) 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 없다.</p>
<p>비돌림위험 일부 우려</p>	<p>(i.1) 중재 결과 측정 방법이 적절하다. 그리고 (i.2) 중재군간 중재결과 측정 혹은 확인 방법이 다르지 않다. 그리고 (i.3) 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 있다. 그리고 (i.4) 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 높지 않다. 혹은 (ii.1) 중재 결과 측정 방법이 적절하다. 그리고 (ii.2) 중재군간 중재결과 측정 혹은 확인 방법이 다른지에 대한 정보가 없다. 그리고 (ii 3-1) 중재결과 평가자는 연구대상자가 받은 중재를 알지 못한다. 혹은 (ii 3-2) 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 없다.</p>
<p>비돌림위험 높음</p>	<p>(i) 중재 결과 측정 방법이 부적절하다. 그리고 (ii) 중재군간 중재결과 측정 혹은 확인 방법이 다를 수 있다. (iii) 중재결과에 대한 평가가 중재에 대한 지식에 의해 영향을 받을 가능성이 높다.</p>

표 2-12. 결과자료 측정의 비틀림위험 판단 신호질문 맵핑

신호질문					영역 수준 판단
4.1 부적절?	4.2 군간 차이?	4.3 알고 있음?	4.4 영향받을 가능성?	4.5 가능성 높음?	비틀림 위험
N/PN/NI	N/PN	N/PN	NA	NA	낮음
N/PN/NI	N/PN	Y/PY/NI	N/PN	NA	낮음
N/PN/NI	N/PN	Y/PY/NI	Y/PY/NI	N/PN	일부우려
N/PN/NI	N/PN	Y/PY/NI	Y/PY/NI	Y/PY/NI	높음
N/PN/NI	NI	N/PN	NA	NA	일부우려
N/PN/NI	NI	Y/PY/NI	N/PN	NA	일부우려
N/PN/NI	NI	Y/PY/NI	Y/PY/NI	N/PN	일부우려
N/PN/NI	NI	Y/PY/NI	Y/PY/NI	Y/PY/NI	높음
Y/PY	Any response	Any response	Any response	Any response	높음
Any response	Y/PY	Any response	Any response	Any response	높음

그림 2-5. 결과자료 측정의 비틀림위험 판단 알고리즘.



8. 보고된 연구결과 선택의 비뚤림

8.1 배경

이 영역은 임상시험 연구자에 의해 계산된 여러 증재 효과 추정치 중에서 선택된(방향, 크기 또는 통계적 유의성에 기초하여) 연구결과 보고 때문에 발생하는 비뚤림을 다룬다. 우리는 이를 ‘보고된 연구결과 선택의 비뚤림’이라 부른다. 비뚤림위험의 고려사항으로 다음과 같은 구분이 필요하다.

- **증재결과 영역.** 이는 관심 상태 또는 종결점이다. 어떻게 측정되는지와는 관계없다(예 : 우울증의 증증도).
- **증재결과 측정.** 이는 증재결과 영역을 측정하는 구체적인 방법이다(예 : 증재 6 주 후 Hamilton 등급 척도를 사용한 우울증 측정)
- **증재결과 분석.** 이는 하나 이상의 증재결과 측정을 분석하여 얻어진 특정 연구결과이다(예 : 임상시험군과 비교군간의 기저치에서 6 주후의 Hamilton 등급 척도 점수의 평균 변화량의 차이).

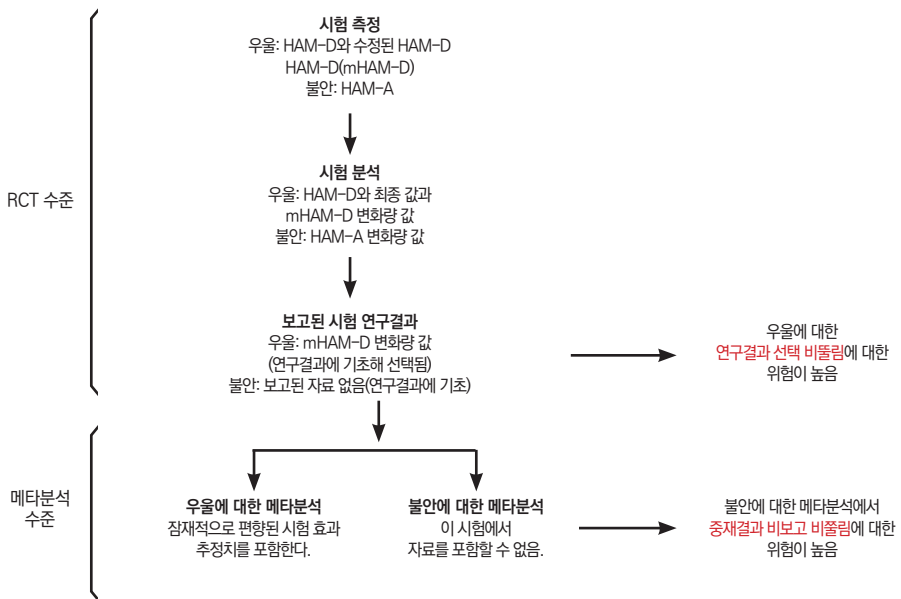
이 영역은 임상시험 연구자가 측정 및 분석된 증재결과 영역이 선택적 비보고(또는 불완전한 보고)되어서 생기는 비뚤림을 다루지 않는다(115). 예를 들어, 임상시험 연구대상자의 사망은 임상시험연구자에 의해 기록될 수 있지만, 임상시험보고서에는 사망률 자료가 포함되어 있지 않거나 사망률에 대한 증재 효과 추정치가 통계적으로 유의하지 않다고 진술할 수 있다. 이러한 비뚤림은 합성의 결과를 방향, 크기 또는 통계적 유의성에 따라 생략하기 때문에, 위험에 처하게 된다. 따라서 이러한 비뚤림은 보고 비뚤림의 위험에 대한 통합 평가의 일부로 문헌고찰 수준에서 다루어져야 한다(116).

증재결과 영역, 측정, 분석의 차이에 대한 예가 그림 6에 나와 있으며, 보고된 결과를 선택할 때 비뚤림위험이 결과 누락으로 인한 비뚤림위험과 어떻게 다른지 보여준다. 이 가상의 예에서, 우울증과 불안의 두 가지 결과 영역이 관심 대상이다. 우울증은 HAM-D (Hamilton Depression Rating Scale)와 수정된 버전(mHAM-D)의 두 가지 척도를 사용하여 측정된다. 이러한 결과는 최종 값과 변화량 값(최종값 - 기저치)을 모두 사용하여 분석된다. 우울증에 대해 보고된 효과 추정치는 임상시험군과 비교군 간의 평균 mHAM-D 변화량의 차이가 실험증재에 가장 유리했기 때문에, 임상시험 연구자에 의해 선택되었다. 따라

서, 우울증에 대한 임상시험으로부터 이용 가능한 결과는 이 RoB 2 영역에서 보고된 연구 결과 선택의 비뚤림위험이 있다. 대조적으로, 임상시험연구자들에 의해 불안이 측정되고 분석되었지만, 효과 추정이 실험중재에 유리하지 않았기 때문에 자료가 보고되지 않았다. 이러한 비보고는 불안에 대한 연구결과의 메타 분석에서 비뚤림을 유발할 것이지만, 그 비보고 자체만으로는 보고된 우울증 효과 추정치나 동일한 임상시험의 다른 중재결과를 비뚤림 위험에 처하게 하지는 않는다.

‘보고된 연구결과 선택’과 ‘선택적 비보고 영역’에 대한 분리는, Cochrane RoB 1 버전에 비해 특징적으로 달라진 것이다.

그림 2-6. ‘보고된 연구결과 선택의 비뚤림’과 ‘중재결과 비보고 비뚤림’의 예시



8.1.1. 핵심 중재결과 세트의 역할(The role of core outcome sets)

임상시험 연구자에 의한 중재결과의 선택적 비보고가 있을 때 생기는 심각한 영향에 대한 인식과, 같은 임상 영역 내에서 시행되는 임상시험에서 측정된 중재결과의 변이가 크다는 인식에 따라 핵심 결과 세트를 개발하려는 이니셔티브가 나타났다(117). 그것은 특정 구체

적 임상 환경에서 모든 임상시험에서 측정해야 하는 적은 수의 필수 중재결과 영역 목록이다. 이것은 임상가와 환자들이 어떤 중재결과를 중요하게 생각하는지를 도출하는 공식적인 합의 과정을 거쳐 개발되었다(118).

핵심 결과 세트를 광범위하게 채택하여 중재결과에 대한 선택적 비보고의 발생을 줄여야 한다. 그러나 보고된 연구결과 선택에 미치는 영향은 핵심 결과를 정의하고 측정하는 방법에 대한 권고 사항을 일관성 있게 포함하지 않았기 때문에 더욱 제한적일 수 있다. 일부 핵심 결과 세트는 임상시험 연구자가 중요한 중재결과를 측정하는 데 사용해야 하는 도구를 권고한다.(119, 120) 핵심 중재결과 측정 세트가 출판된 후 설계된 임상시험에서 이러한 권고 사항을 따르지 않는 경우, 문헌고찰 저자는 선택적 보고가 있을 가능성을 고려할 수 있다.

8.1.2. 합성에 기여하는 연구결과의 선택적 보고

이 영역은 (i) 중재결과 영역 내에서 평가된 다수의 측정치로부터 **특정 중재결과 측정의 선택적 보고** 및 (ii) 특정 중재결과 측정의 다중 분석으로부터 **특정 분석의 선택적 보고**를 둘 다 고려한다. 선택이 효과 추정치의 방향, 크기 또는 통계적 유의성에 기초한 경우, 선택적 보고 유형 중 하나가 비둘림으로 이어질 수 있다.

특정 중재결과 측정에 대한 선택적 보고는 여러 결과에 대한 효과 추정치 중에서 문헌고찰 저자가 관심을 가질 수 있는 중재결과 영역 내의 보고된 효과 추정치를 (연구결과에 기초하여) 선택할 때 발생한다. 예를 들면 다음과 같다.

- 중재결과가 측정된 시점 중 하나 또는 일부만을 보고하는 것(예를 들어, 6 주 및 8 주에서도 중재결과를 측정 했음에도 불구하고 3주 효과만 보고)
- 측정 도구 중 하나 또는 일부만 보고(예 : 통증 척도)
- 중재결과 평가자가 보고한 것 중 하나 또는 일부만을 보고하는 것(예 : 환자 평가, 임상 의 평가 척도는 제외)
- 여러 측정도구 결과 중 하나 또는 하위척도 일부만 보고

문헌고찰 저자가 관심이 있는 다중 분석 방법으로 추정된 중재 효과에서 보고된 결과를 (연구결과에 따라) 선택하면, **분석에 대한 선택적 보고**가 발생된다. 예를 들면 다음과 같다.

- (i) '보정되지 않은 중재 효과 추정치' 및 (ii) '중재결과 측정의 기저 값에 대해 보정된 효과' 중 하나만 보고
- 상이한 세트의 예후 인자에 대해 보정된 다중 분석 중 하나 또는 하위 세트만 보고하는 것
- 다중 복합 중재결과(예 : 사망, 관상 동맥 심장 질환, 심근 경색 및 뇌졸중의 다른 조합으로 정의된 심혈관 질환 사건)에 대한 중재 효과 추정치 중 하나 또는 하위 세트만 보고

보고된 연구결과 선택의 비뚤림은 일반적으로 출판을 위한 주목할 만한 결과에 대한 열망에서 비롯된다. 이는 이전의 근거(또는 이전의 가설)가 뒷받침되거나 모순되는 경우에 해당될 수 있다. 이러한 종류의 비뚤림은 위해와 이득으로 대해 발생할 수 있지만, 위해와 이득에 대한 효과 추정치의 선택에 대한 동기 부여(및 비뚤림 방향)는 다를 수 있다.

예를 들어, 위약과 비교하는 임상시험에서, 임상시험 중재가 유의하고 안전하다는 것을 보여주는데 이해관계가 있는 임상시험 연구자는 임상시험 중재가 통계적으로 유의하고 유리한 효능 추정치를 보고하는데 선택적일 수 있다. 반면, 위해는 구간 유의하지 않은 추정치에 대해 선택적으로 선호하게 된다. 대조적으로, 다른 경우, 위해의 존재를 공개하는 것이 영향력이 큰 저널에 출판될 가능성을 증가 킨다고 연구자가 믿는 경우, 통계적으로 유의미하고 실험 중재에 불리한 유해 추정치를 선택적으로 보고할 수 있다. 그러한 동기가 항상 해독하기 쉬운 것은 아니다. 예를 들어, 행동 중재를 비교하는 임상시험에서, 임상시험 연구자들 중에는 여러 중재 중 한쪽 혹은 다른 쪽의 효능 및 안전성에 대해 서로 다른 선입견을 가질 수 있다.

문헌고찰 저자가 여러 자료를 찾은 후 보고된 연구결과 중 무엇을 선택할 것인가가 중요할 수 있다(예 : 메타 분석에 포함을 위해서). 예를 들어, 변화량 값이 최종 값보다 우선하고 보정된 HAM-D 값이 보정되지 않은 HAM-D 값보다 우선적으로 적용되도록 체계적인 문헌고찰 프로토콜에서 사전 지정되었다고 가정해보자. 그림 1에 제시된 시나리오를 다시 고려하면, 이 경우 보고된 자료는 임상시험에서 보고된 결과를 선택하는 과정에서 잠재적 비뚤림의 근거가 있음에도 불구하고 문헌고찰 저자가 찾은 자료이다. 신호 질문에 “적격”이라는 단어를 포함 시키면 문헌고찰 저자는 모든 결과가 완전히 보고된 경우 연구에서 선택한 자료와 관련하여 질문에 대답할 수 있다. 이 예에서 문헌고찰 저자는 선택적 보고의 근거에도 불구하고 계획된 메타 분석이 해당 문제의 영향을 받지 않는다고 생각할 수 있다.

8.2. 보고된 연구결과를 선택할 때 비뚤림에 대한 경험적 근거

임상시험 보고서 내의 여러 섹션을 비교하고 임상시험 보고서를 프로토콜과 비교하는 체계적인 문헌고찰이 시행되었고, 복합(composite) 중재결과에 대한 정의, 결측된 자료 처리, 보정되지 않은 분석과 보정된 분석, 하위군 분석의 정의 등에서 종종 불일치가 발견되었다(121). 이러한 불일치는 보고된 연구결과 선택의 비뚤림이 있음을 시사하지만, 학회지 편집자 또는 동료 심사자의 요청과 같은 다른 이유 때문일 수도 있다.

8.3. 도구의 이 영역 사용

8.3.1 임상시험의 분석 의도를 찾는 것의 중요성

문헌고찰 저자는 사전에 지정한 분석 의도를 검색해 보는 것을 강력하게 추천한다. 그렇게 하면 임상시험 보고서에서 생략되거나 추가된 중재결과를 식별할 수 있다. 임상시험 레지스트리 항목(예 : ClinicalTrials.gov 기록), 임상시험 프로토콜 또는 설계 문서(저널에 게시되거나 임상시험 펀드 웹 사이트를 통해 제공될 수 있음) 등을 포함하는 다양한 출처로부터 분석 의도가 확인 될 수 있다. 통계 분석 계획(statistical analysis plan, SAP)은 종종 눈가림 되지 않은 결과 자료가 이용 가능하기 전에 사전에 지정된 계획이 완료되었는지 여부에 대한 가장 자세한 정보를 제공하지만, 공개되지 않을 수 있다. 연구자의 사전 지정된 의도가 충분히 상세하게 이용가능하게 되면 계획된 중재결과 측정 및 분석을 발표된 보고서에 제시된 결과와 비교할 수 있다.

분석 의도와 연구 결과를 보고한 출처와 비교할 때, 해당 문서의 날짜를 신중하게 고려해야 한다. 임상시험 연구자(모니터링위원회에 기밀 자료를 제공한 직원 제외)에게 눈가림 되지 않은 결과 자료가 이용가능해지기 전에 분석 의도가 확정되었음을 확인하는 ‘날짜 스탬프’가 있어야 한다. 중재결과 자료는 때때로 눈가림된 형태로 제공되어 분석가가 어떤 중재군에 배정되었는지를 모르는 채 분석 전략을 수정할 수 있게 해준다. 이런 경우는 자료에 중재를 명확하게 식별하는 변수를 제공하지 않는 한 일반적으로 허용 가능한 것으로 간주해야 한다.

분석 의도 자체의 수정 또는 업데이트도 검색하여 원래 의도와 비교해야 한다. 일반적으로 임상시험 등록 기관 또는 저널 간행물에 해당 날짜가 기록되어 있으므로 문헌고찰 작성자는 변경시기를 결정할 수 있어야 한다.

문헌고찰 저자는 연구 프로토콜과 통계 분석 계획을 공개적으로 사용할 수 없는 경우, 연구자에게 연구 프로토콜을 요청해야 한다. 또한, 논문 또는 프로토콜에 언급되어 있는 중재 결과 측정 및 분석 내용이 보고되지 않은 경우, 연구 저자에게 이러한 중재결과가 실제로 분석되었는지 물어보고 그렇다면 자료를 제공하도록 요청할 수 있다.

임상시험 프로토콜에는 예상치 못한 위해 결과(예상치 못한 피해를 반영하는)가 어떻게 수집되고 분석되는지 설명하는 내용이 있어야 한다. 그러나 자발적으로 보고된 유해한 결과의 경우 주목할 만한지에 따라 선택될 수 있는 점을 고려하여야 한다.

8.3.2. 분석 의도를 사용할 수 없을 때 선택적 보고 추론

일부 임상시험의 경우 분석 의도를 쉽게 알 수 없는 경우도 있다. 그럼에도 불구하고 보고된 연구 선택의 비뮴림위험을 평가하는 것은 여전히 가능하다. 예를 들어, 논문의 방법 섹션에 나열된 중재결과 측정 및 분석을 보고된 것과 비교할 수 있다. 또한, 중재결과 측정 및 분석을 임상시험 및 그 연구결과를 설명하는 여러 출처에서 비교할 수 있다. 또한 다음과 같은 질문을 하면 문헌고찰 저자가 선택적 보고를 추론하는 데 도움이 될 수 있다.

- ① 하위척도(subscales)가 통상적이지 않은 방식으로 종합되었나?
- ② 일차 및 이차 결과 또는 특정 결과의 지정에 있어 보고서 간에 불일치가 있나?
- ③ 다수의 보정된 분석이 수행되었지만 단 하나(또는 하위세트만) 보고되었나? 하나 이상의 보정된 분석이 수행되었으나 보고된 바가 없었나?
- ④ 연구자들은 연속형 결과 측정을 통상적이지 않은 방식으로 범주화했나? 같은 연구와 관련하여 여러 출판물에서 범주를 만들기 위한 다양한 절단점이 보고되었나?
- ⑤ 분석된 연구대상자 표본에서 다른 보고서 간에 불일치가 있었나?
- ⑥ 통상적이지 않은 복합 중재결과가 보고되었나? 예를 들어 “주요 부작용 사례” 또는 “경미한 부작용 사례” 범주에서 예상치 못한 부작용 사례의 다른 조합을 그룹화하는 것처럼 복합 중재결과에 대한 서로 다른 정의를 고려할 수 있었나?

분석 의도와 출판 사이의 일부 차이점은 프로토콜의 합법적인 변경으로 인한 것일 수 있음을 인식하는 것이 중요하다. 예를 들어, 계획된 하위군 분석 또는 연속형 결과 측정을 위한 계획된 절단점은 자료 분포가 예상한 것과 다르게 나와서 혹은 하위군이 자료가 없거나 매우 고르지 않은 분포를 초래하여 수정해야 할 수 있다. 그러한 변화는 출판물에 보고되어야 하지만, 그렇게 하는 연구자는 거의 없다(121). 또한, 임상 임상시험 연구자는 분석을 수

행하기 전에 분석 의도를 수정할 수 있지만 공개적으로 이용 가능한 임상시험 등록 기록이나 프로토콜은 업데이트 하지 않을 수 있다. 따라서 식별된 불일치에 대한 설명을 얻으려면 저자와 연락해야 한다.

일부 문서에 세부 정보가 충분하지 않으면 문헌고찰 저자가 평가 중인 현재 연구결과를 생성한 자료가 사전에 지정된 계획에 따라 분석되었는지 여부를 결정하는 데 제한이 있을 수 있다. 예를 들어, 통계 분석 계획에 날짜가 찍히지 않았거나 임상시험 연구자가 레지스트리 자료에 통증을 측정한다고만 기록하고, 시점, 측정 척도를 명시하지 않을 수 있다.

문헌고찰 저자는 불완전한 정보를 기반으로 추론 할 때 RoB 2 신호 질문에 대한 답변을 프리텍스트 박스에 표시하고 이에 대한 근거를 제공해야 한다.

8.4. 비뚫림의 위험을 판단하기 위한 신호 질문 및 기준

이 영역에 대한 신호 질문은 Box 2-11에 제공되어있다. 비뚫림위험 판단에 도달하기 위한 기준은 표 2-13에 나와 있으며 이를 구현하기 위한 알고리즘은 표 2-14와 그림 2-7에 나와 있다.

Box 2-11. RoB 2 도구: 보고된 연구결과 선택의 비뚫림위험

신호질문	세부설명	응답 옵션
5.1 이 연구 결과를 도출한 자료는 분석할 수 있도록 미리 지정된 분석 계획에 따라 중재 결과 자료 눈가림이 해제되기 전에 분석되었는가?	사전에 지정된 연구자의 의도를 충분히 정도로 상세히 이용할 수 있는 경우, 계획된 중재 결과 측정과 보고서에 제시된 것을 비교할 수 있다. 보고된 결과의 선택 가능성을 피하기 위해, 분석 의도가 최종화되는 시점이 자료의 눈가림이 해제되는 시점보다 선행되어야 한다. 자료의 눈가림이 풀리기 전에 분석계획을 변경한 경우 혹은 그러한 변화가 결과와 무관한 경우(예를 들어, 고장난 기계로 인해 자료수집이 불가능함) 보고 결과의 선택에 대한 비뚫림 우려가 제기되지 않는다.	Y/PY/PN/N/NI
평가되고 있는 수치형 연구 결과는 다음과 같은 연구 결과에 기초하여 선택되었을 가능성이 있는가?		

<p>5.2 중재 결과 영역 내에서 여러 개의 적합한 중재 연구 결과 측정(예: 척도, 정의, 시점)?</p>	<p>특정 중재 결과 영역(즉, 실제 상태 또는 관심 종결점)은 여러 가지 방법으로 측정될 수 있다. 예를 들어, 통증 영역은 여러 척도(예: 시각적 아날로그 척도 및 맥길 통증 설문도구)를 사용하여 각각 여러 시점(예: 치료 후 3, 6, 12주)에서 측정할 수 있다. 복수의 측정이 이루어졌지만 결과(예: 통계적 유의성)에 근거하여 1개 또는 하나의 하위세트만 보고되는 경우, 비뮌림위험은 높을 것이다. RoB 2 도구 사용자가 고려하기에 적격인 중재결과에만 주의를 기울이면 된다.</p> <p>예를 들어, 특정 측정 척도를 사용한 결과만 메타 분석(예: 해밀턴 우울증 평가 척도)에 포함할 수 있고, 임상시험에 이 중재결과가 보고된 경우, 다른 측정 척도(예: Beck depression Inventory)에서 나온 결과보다 우선적으로(통계적 의미성 등에 근거하여) 이 결과가 보고되었다 하더라도 선택에 대한 문제는 없을 것이다.</p> <p>다음과 같은 경우 'Y' 또는 'PY'로 응답한다.</p> <p>한 영역을 복수의 적격 방법으로 측정했다는 명확한 근거(보통 임상시험 프로토콜이나 통계 분석 계획서의 검토를 통해)가 있지만, 여러 측정 결과 중 (사유 없이) 하나 또는 일부 subset만 완전히 보고되고, 최종 보고 결과는 그 결과(통계적 유의성 등에 근거하여 선택되었을 가능성이 높다. 결과에 기초하여 선택하는 것은 연구 결과가 뉴스 가치가 있고, 출판할 가치가 있거나, 사전 가설을 확인하기에 충분히 주목할 만한 것이어야 한다는 열망에서 비롯될 수 있다. 예를 들어, 임상시험 중재가 유익하다는 선입견을 가지고 있거나 이해상충이 있는 임상시험 연구자들은 임상시험 중재에 유리한 결과 측정을 선택적으로 보고하는 경향이 있을 수 있다.</p> <p>다음과 같은 경우 'N' 또는 'PN'으로 응답한다.</p> <p>(일반적으로 임상시험 프로토콜 또는 통계 분석 계획의 조사를 통해) 결과 영역에 대해 모든 적격 보고 결과가 의도된 모든 중재 결과 측정과 일치한다는 명확한 근거가 있다.</p> <p>혹은</p> <p>중재 결과 영역을 측정할 수 있는 방법은 단 한 가지뿐이다(여러 번 측정해서 하나를 선택할 기회가 없는 경우).</p> <p>혹은</p> <p>동일한 임상시험에 대한 서로 다른 보고서 간에 중재 결과 측정들이 일관되지 않지만, 임상시험 연구자들이 불일치의 이유를 제공했고, 연구결과와 본질과 관련이 없다.</p> <p>다음과 같은 경우 '정보 없음'으로 응답한다</p> <p>분석 의도를 이용할 수 없거나, 분석 의도를 충분히 상세히 보고하지 않아 평가를 할 수 없고, 중재 결과 영역을 측정할 수 있었던 방법이 두 개 이상 있다.</p>	<p>Y/PY/PN/N/NI</p>
---	--	---------------------

<p>5.3 데이터의 다중 적격 분석(multiple eligible analyses)?</p>	<p>특정 중재 결과를 여러 가지 방법으로 분석할 수 있다. 예를 들어, 비교 정 vs 보정 모형; 최종값 vs 기저 상태에서부터 변화량 vs 공분산 분석; 변수 변환; 복합 중재결과에 대한 다른 정의(예: '주요 유해사건'); 다양한 역할을 가진 연속형 변수를 범주형 변수로 변환; 다양한 공변량 집합에 의한 보정; 결측치를 처리하기 위한 다양한 전략 등이 그것이다. 복수의 방법을 적용하면 특정 결과 측정에 대한 복수의 효과 추정치가 생성된다. 복수의 추정치가 생성 되지만 결과(예: 통계적 유의성)에 근거하여 1개 결과 또는 하나의 하위 세트만 보고되는 경우, 최종 보고된 결과는 비뚫림위험이 높다. RoB 2 도구 사용자가 적격으로 고려할 수 있는 분석으로 제한하도록 주의를 기울여야 한다. 예를 들어, 중재 직후(post-intervention) 측정치만 메타 분석(예: 무작위배정 후 12주)에 포함하는 것이 적격이고, 이 결과가 보고된 경우, 기저치로부터 변화량에 대한 분석 결과보다 우선하여 (결과에 기초해서) 해당결과를 보고했을지라도 선택에 대한 문제는 없을 것이다.</p> <p>다음과 같은 경우 'Y' 또는 'PY'로 응답한다.</p> <p>(보통 임상시험 프로토콜이나 통계 분석 계획의 조사를 통해) 측정이 복수의 적격 방법으로 분석되었다는 명확한 근거가 있지만, 분석의 한 부분 또는 일부에 대한 데이터는 충분히 보고되고, 최종 보고 결과는 하나만 혹은 하나의 하부세트만 보고한 경우(이에 대한 정당화 없이), 최종 보고 결과는 결과에 기초해서 선택되었을 가능성이 높은 경우이다. 결과에 기초하여 선택하는 것은 연구 결과가 뉴스 가치가 있고, 출판할 가치가 있거나, 사전 가설을 증명하기에 충분히 주목할 만한 것이어야 한다는 열망에서 비롯될 수 있다. 예를 들어, 임상시험 중재가 유익하다는 선입견을 가지고 있거나 이해상충이 있는 임상시험 연구자들은 임상시험 중재에 유리한 결과 측정을 선택적으로 보고하는 경향이 있을 수 있다.</p> <p>다음과 같은 경우 'N' 또는 'PN'으로 응답한다.</p> <p>중재결과 측정에 대한 모든 적격하게 보고된 연구결과가 의도된 분석과 동일하다는 명확한 근거가 있다(일반적으로 또는 결과 측정을 분석할 수 있는 방법은 단 한 가지뿐이다(다중 분석에서 선택할 기회가 없는 경우).</p> <p>또는</p> <p>동일한 임상시험에 대한 서로 다른 보고서 간에 중재 결과 측정들이 일관되지 않지만, 임상시험 시행자들이 불일치의 이유를 제공했고, 연구결과와 본질과 관련이 없다.</p> <p>다음과 같은 경우 '정보 없음'으로 응답한다</p> <p>분석 의도를 이용할 수 없거나, 분석 의도를 충분히 상세히 보고하지 않아 평가를 할 수 없고 중재 결과 영역을 측정할 수 있었던 방법이 두 개 이상 있다.</p>	<p>Y/PY/PN/N/ NI</p>
---	--	--------------------------

비뚤림위험 평가	알고리즘 참조	낮음/높음/일부 우려
<p>선택사항: 보고된 연구결과 선택의 비뚤림의 예측 방향은 어떠한가?</p>	<p>비뚤림의 방향을 예측할 수 있으면, 이를 기술하는 것이 도움이 된다. 방향은 무효점(혹은 무효점의 반대쪽)을 향한, 혹은 실험 중재선호, 비교중재 선호로 기술한다.</p>	<p>적용불가 / 실험 중재선호 / 비교중재선호/ 무효점을 향한/ 무효점에서 멀어짐/예측불가</p>

표 2-13. 보고된 연구결과 선택의 비뚤림위험 판단

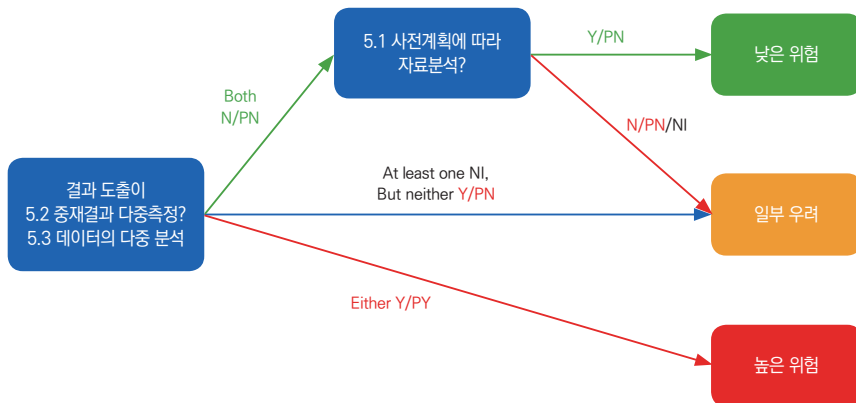
<p>낮은 비뚤림위험</p>	<p>(i) 자료는 눈가림되지 않은 중재결과 자료가 분석에 사용되기 전에 확정된 사전의 구체적인 계획에 따라 분석되었다. 그리고 (ii) 평가된 연구결과는 중재결과 영역 내에서 다중 적격 중재결과 측정(예 : 척도, 정의, 시점)의 연구결과에 기초하여 선택되지 않은 것 같다. 그리고 (iii) 보고된 중재결과 자료가 다중 적격 자료 분석 결과에 따라 서 선택되지 않은 것 같다</p>
<p>일부 우려</p>	<p>(i.1) 자료는 눈가림되지 않은 결과 자료를 분석하기 전에 확정된 사전 지정된 계획에 따라 분석되지 않았다. 그리고 (i.2) 평가된 연구결과는 중재결과 영역 내에서 다중 적격 중재결과 측정(예 : 척도, 정의, 시점)의 연구결과에 기초하여 선택되지 않은 것 같다. 그리고 (i.3) 연구결과가 다중 적격 자료 분석의 연구결과에 기초하여 선택되지 않은 것 같다. 혹은 (ii) 평가된 연구결과는 중재영역 내의 다중 적격 중재결과 측정(예 : 척도, 정의, 시점)과 다중 적격 자료 분석으로부터 연구결과에 기초하여 선택되었는지에 대한 정보가 없다.</p>

높은 비돌림위험	i) 평가된 연구결과는 중재결과 영역 내에서 다중 적격 중재결과 측정(예 : 척도, 정의, 시점) 결과를 기반으로 선택되었을 가능성이 있다. 혹은 (ii) 평가된 연구결과는 연구결과에 기초하여 다중 적격 자료 분석에서 선택되었을 가능성이 있다.
----------	--

표 2-14. 보고된 연구결과 선택의 비돌림위험 판단 신호질문 맵핑

신호질문			영역 수준 판단
5.1 계획에 일치?	5.2 다중 중재결과에서 선택?	5.3 다중분석에서 선택?	비돌림위험
Y/PY	N/PN	N/PN	낮음
N/PN/NI	N/PN	N/PN	일부우려
Any answer	N/PN	NI	일부우려
Any answer	NI	N/PN	일부우려
Any answer	NI	NI	일부우려
Any answer	Either 5.2 or 5.3 Y/PY		높음

그림 2-7. 보고된 연구결과 선택의 비돌림위험 판단 알고리즘



References

1. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2016 (Published online 29 September).
2. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med*. 2002;21(11):1513-24.
3. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-12.
4. Savović J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med*. 2012;157(6):426-38.
5. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377(14):1391-8.
6. Insight Start Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med*. 2015;373(9):795-807.
7. Lodi S, Sharma S, Lundgren JD, Phillips AN, Cole SR, Logan R, et al. The per-protocol effect of immediate versus deferred antiretroviral therapy initiation. *AIDS*. 2016;30(17):2659-63.
8. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325(7365):652-4.
9. Menerit CL. *Clinical Trials – Design, Conduct, and Analysis*. Second Edition: Oxford University Press; 2012.
10. Piantadosi S. *Clinical Trials: A Methodologic perspective*. Second Edition: Wiley; 2005.
11. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319(7211):670-4.
12. Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials*. 2007;4(4):350-6.
13. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs: a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118.
14. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.
15. Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: what is the right question? *Clin Trials*. 2014;11(1):28-37.
16. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA*. 2014;312(6):606-15.

17. Savovic J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane Reviews: the ROBES meta-epidemiologic study. *Am J Epidemiol*. 2017.
18. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-12.
19. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-6.
20. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med*. 2002;136(3):254-9.
21. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359(9305):515-9.
22. Schulz KF, Grimes DA. *The Lancet Handbook of Essential Concepts in Clinical Research*. Edinburgh, UK: Elsevier; 2006.
23. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet*. 2002;359(9310):966-70.
24. Berger VW, Ivanova A, Knoll MD. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat Med*. 2003;22(19):3017-28.
25. Berger VW. Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biometrical journal Biometrische Zeitschrift*. 2005;47(2):119-27; discussion 28-39.
26. Brown S, Thorpe H, Hawkins K, Brown J. Minimization--reducing predictability for multi-centre trials whilst retaining balance within centre. *Stat Med*. 2005;24(24):3715-27.
27. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? *BMJ*. 2016;355:i5663.
28. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274(18):1456-8.
29. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-6.
30. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *The Cochrane database of systematic reviews*. 2012;11:MR000030.
31. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol*. 2007;36(4):847-57.
32. Corbett MS, Higgins JPT, Woolacott NF. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. *Research synthesis methods*. 2014;5(1):79-85.
33. Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hrobjartsson A, Savovic J. Empirical Evidence

- of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PLoS One*. 2016;11(7):e0159267.
34. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002;287(22):2973-82.
35. Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. *PLoS One*. 2014;9(2):e88008.
36. Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. *J Clin Epidemiol*. 2013;66(11):1271-80.
37. Chapin RB, Williams DC, Adair RF. Diabetes control improved when inner-city patients received graphic feedback about glycosylated hemoglobin levels. *J Gen Intern Med*. 2003;18(2):120-4.
38. Saugstad OD, Rootwelt T, Aalen O. Resuscitation of asphyxiated newborn infants with room air or oxygen: an international controlled trial: the Resair 2 study. *Pediatrics*. 1998;102(1):e1.
39. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614-8.
40. Bellomo R, Chapman M, Finfer S, Hickling K, Myburgh J. Low-dose dopamine in patients with early renal dysfunction: a placebo-controlled randomised trial. Australian and New Zealand Intensive Care Society (ANZICS) Clinical Trials Group. *Lancet*. 2000;356(9248):2139-43.
41. Smilde TJ, van Wissen S, Wollersheim H, Trip MD, Kastelein JJ, Stalenhoef AF. Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *Lancet*. 2001;357(9256):577-81.
42. de Gaetano G. Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice. Collaborative Group of the Primary Prevention Project. *Lancet*. 2001;357(9250):89-95.
43. Brightling CE, Monteiro W, Ward R, Parker D, Morgan MD, Wardlaw AJ, et al. Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet*. 2000;356(9240):1480-5.
44. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274(18):1456-8.
45. Altman DG, Bland JM. How to randomize. *BMJ*. 1999;319:703-4.
46. Shrier I, Platt RW, Steele RJ. Re: "Variable selection for propensity score models". *Am J Epidemiol*. 2007;166(2):238-9.
47. Senn S. Seven myths of randomisation in clinical trials. *Stat Med*. 2013;32(9):1439-50.
48. Fu R, Vandermeer BW, Shamliyan TA, O'Neil ME, Yazdi F, Fox SH, et al. AHRQ Methods for

- Effective Health Care: Handling Continuous Outcomes in Quantitative Synthesis. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
49. Wright IS, Marple CD, Beck DF. Report of the Committee for the Evaluation of Anticoagulants in the Treatment of Coronary Thrombosis with Myocardial Infarction: a progress report on the statistical analysis of the first 800 cases studied by this committee. *Am Heart J*. 1948;36(6):801-15.
 50. Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *J Clin Epidemiol*. 2007;60(12):1229-33.
 51. Carlisle JB. Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia*. 2017;72(8):944-52.
 52. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. 2016;355:i4919.
 53. Hernan MA, Scharfstein D. Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat. *Ann Intern Med*. 2018;168(7):515-6.
 54. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology*. (in press).
 55. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005;365(9453):82-93.
 56. Worsley SD, Oude Rengerink K, Irving E, Lejeune S, Mol K, Collier S, et al. Series: Pragmatic trials and real world evidence: Paper 2. Setting, sites, and investigator selection. *J Clin Epidemiol*. 2017;88:14-20.
 57. Madsen MV, Gotzsche PC, Hrobjartsson A. Acupuncture treatment for pain: systematic review of randomised clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups. *BMJ*. 2009;338:a3115.
 58. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med*. 2006;3(10):e425.
 59. Jensen JS, Bielefeldt AO, Hrobjartsson A. Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. *J Clin Epidemiol*. 2017;87:35-46.
 60. Bello S, Moustgaard H, Hrobjartsson A. Unreported formal assessment of unblinding occurred in 4 of 10 randomized clinical trials, unreported loss of blinding in 1 of 10 trials. *J Clin Epidemiol*. 2017;81:42-50.
 61. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ*. 2004;328(7437):432.

62. Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Contemp Clin Trials*. 2005;26(1):25-37.
63. Hrobjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol*. 2007;36(3):654-63.
64. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol*. 2007;36(3):664-5.
65. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285(15):2000-3.
66. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol*. 2005;58(12):1220-6.
67. Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials*. 2006;3(4):360-5.
68. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol*. 2002;55(8):787-90.
69. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Bmj*. 2010;340:c332.
70. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA*. 2006;296(20):2441-50.
71. De Bruyne B, Fearon WF, Pijls NH, Barbato E, Tonino P, Piroth Z, et al. Fractional flow reserve-guided PCI for stable coronary artery disease. *N Engl J Med*. 2014;371(13):1208-17.
72. Biere SS, van Berge Henegouwen MI, Maas KW, Bonavina L, Rosman C, Garcia JR, et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multicentre, open-label, randomised controlled trial. *Lancet*. 2012;379(9829):1887-92.
73. St Peter SD, Adibe OO, Juang D, Sharp SW, Garey CL, Laituri CA, et al. Single incision versus standard 3-port laparoscopic appendectomy: a prospective randomized trial. *Ann Surg*. 2011;254(4):586-90.
74. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med*. 2012;157(6):429-38.
75. Hrobjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *Int J Epidemiol*. 2014;43(4):1272-83.
76. Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum*. 2009;61(12):1633-41.

77. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine-- selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326(7400):1171-3.
78. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol*. 2007;60(7):663-9.
79. Abraha I, Cherubini A, Cozzolino F, De Florio R, Luchetta ML, Rimland JM, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. *BMJ*. 2015;350:h2445.
80. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol*. 2005;34(1):79-87.
81. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*. 2010;340:c2697.
82. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual participant data (IPD) meta- analyses of randomised controlled trials: guidance on their use. *PLoS Med*. 2015;12(7):e1001855.
83. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010.
84. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319(7211):670-4.
85. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019.
86. Higgins JPT, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials*. 2008;5(3):225-39.
87. Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med*. 2011;154(2):113-7.
88. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Second Edition. Hoboken, NJ: Wiley; 2002.
89. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367(14):1355-60.
90. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-95.
91. Robins J, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. *Journal of the American Statistical Association*. 1995;90:106-21.
92. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation*. 1981;64(4):669-73.
93. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987.

94. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
95. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2016;962280216683570.
96. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials*. 2012;9(4):396-407.
97. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *Pharmacoeconomics*. 2018.
98. Akl EA, Kahale LA, Agoritsas T, Brignardello-Petersen R, Busse JW, Carrasco-Labra A, et al. Handling trial participants with missing outcome data when conducting a meta-analysis: a systematic survey of proposed approaches. *Syst Rev*. 2015;4:98.
99. Akl EA, Johnston BC, Alonso-Coello P, Neumann I, Ebrahim S, Briel M, et al. Addressing dichotomous data for participants excluded from trial analysis: a guide for systematic reviewers. *PLoS One*. 2013;8(2):e57132.
100. Ebrahim S, Akl EA, Mustafa RA, Sun X, Walter SD, Heels-Ansdell D, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol*. 2013;66(9):1014-21 e1.
101. Ebrahim S, Johnston BC, Akl EA, Mustafa RA, Sun X, Walter SD, et al. Addressing continuous data measured with different instruments for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol*. 2014;67(5):560-70.
102. Mavridis D, White IR, Higgins JP, Cipriani A, Salanti G. Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. *Stat Med*. 2015;34(5):721-41.
103. Akl EA, Kahale LA, Ebrahim S, Alonso-Coello P, Schunemann HJ, Guyatt GH. Three challenges described for identifying participants with missing data in trials reports, and potential solutions suggested to systematic reviewers. *J Clin Epidemiol*. 2016;76:147-54.
104. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 2004;1(4):368-76.
105. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med*. 2001;135(11):982-9.
106. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med*. 2007;26(14):2745-58.
107. Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological

- study. *BMJ*. 2009;339:b3244.
108. Rothman KJ, Greenland S, Lash TL. Validity in epidemiological studies. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*. 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2008. p. 128- 47.
 109. Clemens JD, Chuong JJ, Feinstein AR. The BCG controversy. A methodological and statistical reappraisal. *JAMA*. 1983;249(17):2362-9.
 110. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PE, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis*. 2014;58(4):470-80.
 111. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1-2):51-63.
 112. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and quality of life outcomes. 2006;4:79.
 113. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*. 2012;344:e1119.
 114. Engebretsen K, Grotle M, Bautz-Holter E, Ekeberg OM, Juel NG, Brox JI. Supervised exercises compared with radial extracorporeal shock-wave therapy for subacromial shoulder pain: 1-year results of a single-blind randomized controlled trial. *Phys Ther*. 2011;91(1):37-47.
 115. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365.
 116. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Systematic reviews*. 2016;5(1):108.
 117. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy*. 2012;17(1):1-2.
 118. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med*. 2017;14(11):e1002447.
 119. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev*. 2016;5:11.
 120. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument selection using the OMERACT Filter 2.1: the OMERACT methodology. *J Rheumatol*. 2019.
 121. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JPT, Sterne JAC, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med*. 2014;11(6):e1001666.

PART 3

ROBINS-I : 원본 매뉴얼 번역

- 1. 배경 148
- 2. 도구를 사용하기 위한 지침: 일반적인 고려사항 155
- 3. 도구 사용 지침: 각 비틀림 영역에 대한 자세한 지침 168



1. 배경

중재 효과 체계적 문헌고찰의 목표는 하나 이상의 중재결과에 대한 효과의 인과성을 결정하는 것이다. 포함된 연구가 무작위 임상시험인 경우 임상시험이 방법론적으로 타당하면 인과관계를 추론할 수 있다. 충분히 많은 사람에 대해 무작위 배정 임상시험을 하면 관찰되거나 관찰되지 않은 예측 요인이 중재군과 비교군 간에 비슷한 분포를 보이기 때문이다. 그러나 무작위 임상시험의 근거가 환자나 보건의료 제공자가 관심있는 질문에 대답하기에 충분하지 않을 수 있으므로, 체계적 문헌고찰 저자는 중재 효과에 대한 비무작위 임상연구(NRSIs)를 문헌고찰에 포함시키길 원할 수 있다.

ROBINS-I 도구는 두 개 이상의 중재의 건강 효과를 비교하는 NRSI 연구 결과에서의 비뚤림위험을 평가하는 것과 관련이 있다. 이 도구를 사용하여 평가할 수 있는 NRSI는 중재의 효과(위해 또는 이득)를 측정하고, 비교군(개인이나 집단) 배정에 무작위 방법을 사용하지 않는 정량적 연구이다. 이러한 연구를 종종 ‘관찰 연구’라고 한다. NRSI에는 여러 유형이 있는데 코호트 연구(cohort study), 환자-대조군 연구(case-control study), 통제 전후 연구(controlled before-and-after study), 단속 시계열 연구(interrupted-time-series), 그리고 완전한 무작위 배정에 미치지 못하는 방법을 사용하여 배정되는 통제된 임상시험(‘준 무작위(quasi-randomized)’ 연구라고 함) 같은 것들이 있다. 이 문서는 특히 서로 다른 중재를 받는 사람들이 시간에 따라 추적관찰이 되는 코호트 유형의 설계를 가진 연구에 대해 ROBINS-I를 사용하는 지침을 제공한다.

ROBINS-I 도구는 2008년에 만들어져 2011년에 수정된, 무작위 임상시험을 위한 Cochrane RoB 도구를 기반으로 한다(Higgins et al, 2011). 무작위 임상시험을 위한 도구에서와 같이, 비뚤림위험은 정해진 비뚤림 영역 내에서 평가되며, 문헌고찰 저자는 판단의 근거가 되는 정보를 문서화해야 한다. 또한 ROBINS-I는 진단 정확도 연구 평가를 위한 QUADAS 2 도구(Whiting et al, 2011)와 같은 관련 도구를 기반으로 하여 비뚤림의 가능성을 나타내는 신호 질문을 제공하여 작성자가 비뚤림위험을 판단하는데 도움이 되도록 한다. ROBINS-I 도구는 NRSI의 비뚤림위험과 관련된 사용 가능한 근거를 정리하고 보여주는 체계적인 방법을 제공한다.

1.1. 도구의 맥락

NRSI의 체계적 문헌고찰에서 비뚤림위험을 평가하려면 방법론적인 지식과 내용에 대한 전문적인 지식이 필요하다. 이 과정은 무작위 임상시험에서 비뚤림위험을 평가하는 과정보다 더 복잡하며, 보통 3단계로 구성된다.

첫째, 계획 단계에서 문헌고찰 질문(review question)을 명확하게 설명하고 NRSI의 중요한 잠재적 문제를 확인해야 한다. 여기에는 주로 교란(표 3-1 및 3.1 절 참조)과 공동 중재(3.4절 참조)에 대한 사전 설명을 포함한다.

둘째, 비뚤림위험에 처할 수 있는 모든 방법을 고려하여 각 연구를 면밀히 검토해야 한다. 평가는 예상하지 못했을 수도 있는 중요한 문제를 확인하기 위해 사전 고려사항을 도출해야 한다. 예를 들어, 추가 핵심 교란, 또는 중재 정의의 문제, 또는 중요한 공동 중재 같은 것들이 확인될 수 있다.

셋째, 관찰된 중재 효과가 원인이 될 수 있는 정도의 결론을 도출하기 위해, 연구를 비교하고 대조하여 이들의 장단점을 같이 고려할 수 있도록 해야 한다. 서로 다른 설계의 연구는 다른 유형의 비뚤림을 나타낼 수 있으며, 이러한 연구들 사이에 발견되는 “삼각 측량(triangulation)”은 비뚤림들이 최소한이라거나 실재한다는 확신을 제공할 수 있다.

이 문서는 주로 NRSI에서 비뚤림위험을 평가하는 도구를 제안함으로써 이러한 단계 중 두 번째 단계를 다룬다. 각 연구의 평가에 정보를 제공하는 것이 필요해, 일부 첫 번째 단계의 고려사항도 다룬다.

1.2. 목표 임상시험(target trial)과 관련해서 비뚤림위험을 평가하기

ROBINS-I와 무작위 임상시험에 대한 Cochrane RoB 도구는 모두 연구의 내적 타당도에 초점을 맞춘다. 여기에서는 ‘비뚤림’을 ‘무작위 임상시험에서 예상되는 연구결과와 체계적으로 다른 연구결과의 경향성’으로 정의한다. 비뚤림위험이 없는 임상시험은 전형적으로 무작위 배정을 은폐하고, 추적 조사 내내 받은 중재에 대해 환자, 의료 전문가 및 결과 평가자의 눈가림 상태를 유지하며, 무작위화된 모든 연구대상자의 결과를 확인하고, 모든 측정 결과에서 중재 효과를 보고하는 대규모 임상시험일 것이다. 이러한 방식으로 정의된다면, 비뚤림은 연구에 포함되지 않은 사람들에게 적용할 수 있는 **일반화 가능성(적용가능성 또는 전이성(transportability))**의 이슈와는 구별된다. 예를 들어 동반질환이 없는 연구대상자로

제한하는 것은 동반질환이 흔한 일반적인 임상 현장에 일반화하기 어렵기 때문에 연구 결과의 유용성을 제한할 수 있다.

NRSI의 연구결과에서 비뿔림위험을 평가하는 것은 각 NRSI가 가상의 임상시험을 모방(흉내)한다고 가정해야 가능하다. 이는 동일한 연구대상자 그룹에서 수행된 동일한 중재의 건강 효과를 보기 위해 시행되는 비뿔림위험이 없는 가설적인 실용적 무작위 임상시험이다(Hernán 2011; IOM 2012). 우리는 이러한 가상의 무작위 임상시험을 ‘목표’ 무작위 임상시험이라고 부를 것이다(자세한 내용은 2.1.1절 참조). 중요한 것은 목표 무작위 임상시험이 실현 가능하거나 윤리적인 필요는 없다는 것이다.

ROBINS-I는 NRSI가 모방하려는 목표 임상시험에서 비교될 중재를 명시적으로 확인하도록 요구한다. 관찰 연구에서 제공되는 정보가 목표 임상시험을 정의하기에 불충분하기 때문에 종종 중재에 대한 설명을 위해 연구주제에 관한 지식이 필요하다. 예를 들어, 저자는 ‘치료의 사용[A]’을 언급할 수 있는데, 이는 목표 임상시험의 ITT 분석에서 시험하는 ‘치료의 시작[A]’과 직접적으로 일치하지는 않는다. 잘 정의된 중재가 없는 경우 비뿔림위험에 대한 의미 있는 평가가 어려울 수 있다. 예를 들어 비만환자에 대한 체중 감소 중재(예: 칼로리 제한)가 사망률에 미치는 효과보다는 비만이 사망률에 미치는 효과에 대한 교란 요인을 평가하는 것이 더 어려울 것이다.

목표 임상시험과 유사성을 유지하기 위해, 연구자에 의해 실제 중재가 수행되지 않았음에도 불구하고, 이 문서에서는 관찰연구에서의 ‘치료’ 또는 ‘노출’ 그룹에 대해 ‘중재’ 그룹이라는 용어를 사용한다.

1.3. 비뿔림의 영역

ROBINS-I 도구는 비뿔림이 NRSI에 개입될 수 있는 7개의 영역을 다룬다. 이러한 영역은 모든 유형의 NRSI를 고려하는 틀을 제공하며, 표 3-1에 요약되어 있다. 첫 두 영역은 비교할 중재가 시작되기 전의 문제를 다루고(“기저 상태(baseline)”) 세 번째 영역은 중재 자체의 분류를 다룬다. 나머지 네 개의 영역은 중재가 시작된 후의 문제를 다룬다. 처음 세 가지 영역의 경우, NRSI에서 비뿔림위험을 평가하는 것은, (무작위 임상시험의) 무작위화가 중재의 시작 전에 나타나는 비뿔림을 방지하기 때문에, 무작위 임상시험을 평가하는 것과는 구별된다. 그러나 무작위화는 중재 시작 후 발생하는 비뿔림을 방지하지는 못하기 때문에 NRSI와 무작위 임상시험에서 비뿔림위험 평가의 마지막 네 가지 영역이 상당히 중복된다.

연구자 간, 그리고 연구 분야 간 용어의 다양성은 ROBINS-I를 개발하고 지침을 만들 때 도전 과제였다. 같은 용어가 다른 유형의 비뚤림을 나타내는 경우도 있고, 다른 유형의 비뚤림이 많은 다른 용어에 의해 설명되기도 한다.

‘선택 비뚤림(selection bias)’이라는 용어는 특히 혼란의 원천이다. 그것은 종종 하나 이상의 예후 인자가 어떤 개인이 받는 하나 이상의 관심 중재를 받는지 여부를 예측하는 (무작위 임상연구에서 RoB를 평가하기 위해 현재의 Cochrane 도구에 포함된) ‘교란(confounding)’과 동의어로 되기도 한다. 하지만 우리는 일부 적격 연구대상자, 또는 일부 연구대상자의 처음 추적조사 시간, 또는 일부 중재결과 사건들이, 목표 임상시험의 완전한 추적조사에서 관찰될 수 있는 중재와 중재결과 사이의 연관성과는 다른 방향으로 나타나게 하는 방식으로 제외될 때, 이러한 종류의 비뚤림에 선택 비뚤림이라는 용어를 사용하는 것으로 제한한다. 우리는, 교란을 지칭하기 위해 선택 비뚤림이라는 용어를 사용하지 않는다. 무작위 임상시험의 RoB를 평가하는 현재의 Cochrane 도구와 ROBINS-I 도구 간의 이러한 용어 차이를 해결하기 위한 작업이 진행 중이다.

무작위 임상시험과 달리 NRSI에서는 연구대상자의 특성이 중재 그룹마다 다를 것이다. 그러므로 통제되지 않은 교란으로 인한 비뚤림위험의 평가는 ROBINS-I 평가의 주요 구성 요소이다. 중재 효과의 교란은 하나 이상의 예후 인자(관심 결과를 예측하는 인자)가 어떤 개인이 하나 이상의 관심 중재를 받는지 여부를 예측할 때 발생한다. 예를 들어 항 레트로 바이러스 요법 A와 항 레트로 바이러스 요법 B의 시작으로 인한 사망 위험을 비교하는 HIV 감염 환자에 대한 코호트 연구를 생각해보자. 교란이 성공적으로 통제되면, 표본의 변이를 제외한다면, 이 관찰 연구의 효과 추정치가 동일한 연구 모집단의 개인을 중재 A 또는 B에 무작위로 배정하는 임상시험의 효과 추정치와 동일할 것이다. 그러나 핵심 교란요인의 통제를 실패하는 경우 치료 A와 B를 받는 사람들 사이의 비교성이 낮아져 비뚤림을 초래할 것이다. 교란의 평가에 대한 자세한 토의내용은 3.1절에 나와 있다.

분석 과정에 모든 연구대상자를 포함하지 않거나 중재의 시작 이후 추적 조사에 포함되지 않은 경우 선택 비뚤림이 발생할 수 있다. ROBINS-I 도구는 두 가지 유형의 선택 비뚤림을 다룬다:

- ① 일부 사람들에서 중재 시작 후 추적 조사의 모든 기간 또는 추적 기간이 누락된 경우 발생하는 비뚤림(예를 들어 어떤 중재를 처음 사용하는 사람보다 이미 사용하고 있는 사람들이 포함되면서 나타나는 비뚤림),
- ② 처음에 포함되고 추적하고 있는 사람들인데 나중 추적조사가 누락되어 나타나는 비뚤림

(예를 들어 예측 요인에 의해 영향을 받는 추적 조사의 서로 다른 손실로 인한 비뚤림).

우리는 첫 번째 유형의 선택 비뚤림을 “연구대상자 선택에서의 비뚤림”(3.2절)이라 하고, 추적 조사 손실과 관련된 것을 “결측치로 인한 비뚤림”(3.5절)이라 할 것이다. 이러한 비뚤림의 유형들의 예시는 관련 절(section)에서 나온다.

표 3-1. ROBINS-I 도구에 포함된 비뚤림 영역

영역	관련 용어	설명
증재 전		
교란으로 인한 비뚤림	종종 임상시험과 관련해서 사용되는 용어로서 선택 비뚤림(현재 Cochrane에서 널리 사용됨); 배경 비뚤림; 증례-혼합 바이어스; Channelling 비뚤림	기저 상태 교란 요인은 하나 이상의 예측 변수(관심 결과를 예측하는 요인)가 기저상태에서 받게되는 증재를 예측할 때 발생한다. ROBINS-I는 시간 변동 교란 요인도 다루는데, 이것은 개인들이 비교 증재로 바뀔 때와 기저 상태 후 예측 변수가 기저 상태 후 받은 증재에 영향을 미칠 때 발생한다.
연구대상자 선택의 비뚤림	종종 관찰연구와 임상시험과 관련해서 일반적으로 사용되는 용어로서 선택 비뚤림; inception bias; lead time bias; immortal time bias. 이 비뚤림은 외적 타당도가 부족한지에 대해서는 다루지 않는데, 외적 타당도가 부족한 것은 비뚤리지 않은(내적 타당도를 확보한) 효과 추정치를 다른 인구집단에 대해 일반화하거나 이행하기에 실패한 것으로 보이는 것과 관련이 있다.	증재와 결과에 모두 연결되어 있는 일부 적격 연구 대상자 혹은 일부 참가자의 초기 추적시간 혹은 일부 증재결과 발생을 배제하는 경우, 증재의 영향이 동일하더라도 증재와 증재결과 사이에 연관성이 발생한다. 이러한 유형의 비뚤림은 교란 요인과는 구별된다. 예로는 신규 사용자가 아닌 기존 사용자가 증재에 포함됨으로 인한 비뚤림이다.
증재 중		
증재 분류의 비뚤림	오분류 비뚤림; 정보 비뚤림; 회상 비뚤림; 측정 비뚤림; 관찰자 비뚤림	증재 상태의 차등적 또는 비차등적 오분류에 의해 도입된 비뚤림. 비차등적 오분류는 증재결과와 무관하며 일반적으로 증재의 추정효과를 무효점(null)으로 향하게 한다. 차등적 오분류는 증재 상태의 오분류가 증재결과 또는 증재결과의 위험과 관련이 있고 이로 인해 비뚤림을 야기할 때 발생한다.
증재 후		

<p>의도한 증재에서 이탈되어 나타나는 비뚫림</p>	<p>실행 비뚫림; 시간-변이 교란</p>	<p>실험 증재군과 대조군 사이에 제공된 관리가 체계적으로 다를 때 발생하는 비뚫림이다. 이는 의도한 증재로부터의 벗어난 것을 의미한다. 이 영역의 비뚫림 평가는 관심 효과의 유형(증재에 배정 효과 또는 증재 시작 및 준수의 효과)에 따라 달라진다.</p>
<p>결측치로 인한 비뚫림</p>	<p>탈락 비뚫림; 종종 관찰연구와 관련하여서 사용되는 용어로서 선택 비뚫림</p>	<p>포함 되었던 연구 대상자가 추적 관찰에서 누락되어 발생하는 비뚫림(예: 예측 요인에 의해 영향을 받는 추적 관찰의 차등적 차이). 증재 상태나 교란 변수와 같은 변수의 정보가 누락되어 발생한 비뚫림</p>
<p>증재결과 측정 비뚫림</p>	<p>결과 확인(detection) 비뚫림; 회상 비뚫림; 정보 비뚫림; 오분류 비뚫림; 관찰자 비뚫림; 측정 비뚫림</p>	<p>증재결과 자료의 측정에서 차등적 또는 비차등적 오분류에 의해 발생하는 비뚫림. 증재결과 평가자가 증재 상태를 인지하고 있거나, 다른 증재군의 증재결과를 평가하기 위해 다른 방법을 사용하거나, 측정 오류가 증재 상태 또는 효과와 관련이 있는 경우 발생한다.</p>
<p>보고된 연구결과 선택의 비뚫림</p>	<p>증재결과 보고 비뚫림; 분석 보고 비뚫림</p>	<p>연구결과에 따라 선택적으로 연구결과를 보고하여 발생한다.</p>

1.4. 연구설계

이 문서는 코호트 연구, 준 무작위 임상시험 및 기타 동시 통제 연구(other concurrently controlled study)와 같은 코호트와 유사한 디자인을 가진 NRSI와 가장 밀접하게 관련되어 있다. 이 자료의 대부분은 환자-대조군 연구, 단면 연구, 중단된 시계열 연구, 통제된 전후 연구와 같은 설계와도 관련이 있지만, 현재 우리는 이러한 다른 유형의 연구에서 신호 질문 에 대한 수정이 필요한지 여부를 고려 중이다.

1.5. 비뚤림위험 평가는 특정 중재 효과와 관련이 되어야 함

이 섹션은 연구가 정량화하려는 중재의 효과와 관련이 있다. 목표 임상시험에서 관심 효과는 아래의 두 가지에 해당한다.

- 추적 조사 동안 받은 중재의 정도에 관계없이, 기저상태(추적조사 시작)에서의 중재를 배정하는 효과(중중 무작위 임상시험의 상황에서 “중재배정(intention-to-treat)” 효과라고 부름). 또는
- 임상시험 계획서에 명시된대로 중재를 시작하고 준수하는 효과(중중 무작위 임상시험의 상황에서 “프로토콜 준수(per-protocol)” 효과라고 부름)

예를 들어 특정 보건 시스템에 대한 중재를 권고할 것인지에 대한 보건 정책에 정보를 주기 위해 우리는 아마도 중재 배정의 효과를 측정할 것이다. 반면에 개별 환자의 진료 결정에 정보를 주기 위해 우리는 특정 비교중재와 비교하여 특정 프로토콜에 따라 치료를 시작하고 준수하는 효과를 추정하려고 할 것이다. 문헌고찰 저자는 각 NRSI에서 관심 있는 중재 효과를 정의하고 이 효과에 비뚤림위험 도구를 적절하게 적용해야 한다. 중재 효과의 선택과 관련된 문제는 아래 2.2.2절에서 자세히 설명한다.

ROBINS-I의 맥락에서 중재 효과의 구체화는 상대적 혹은 절대적 측정치의 선택이나 문헌고찰 질문의 PICO(patients, intervention, comparator, outcome)의 요소와는 관련이 없다.

1.6. 이 문서의 구조

이 문서의 2,3절은 ROBINS-I 사용에 대한 자세한 지침을 제공한다. 여기에는 문헌고찰 프로토콜 작성(2.1절), 관심 효과를 정하는 문제(2.2.2절), 비뚤림위험 평가에서 신호 질문의 사용(2.3절), 영역 수준에서 비뚤림 평가할 때 요구사항(2.4절), 이러한 내용들을 비뚤림 위험에 대한 전반적인 판단을 위해 어떻게 사용해야하는지(2.5절), 그리고 결과-수준의 평가에서 사용(2.6절)할 때 고려할 사항들이 포함된다. 각 영역별 비뚤림 평가에 대한 자세한 지침은 3절에서 볼 수 있다.

2. 도구를 사용하기 위한 지침: 일반적인 고려사항

2.1. 프로토콜 단계

2.1.1 연구질문을 구체화하기

연구질문은 문헌고찰의 목표와 직접적으로 연결된다. 연구질문에서는 인구, 실험 중재군, 비교군 및 관심 중재결과를 다룬다. 비교군은 무중재, 일상적인 진료, 혹은 대안적인 중재가 모두 가능하다.

NRSI의 문헌고찰을 시작할 때는 연구질문의 맥락에서 NRSI에 기초한 관심 중재의 효과에 대한 인과성 평가를 할 때 어떤 문제가 발생할 수 있는지를 고려하여야 한다. 무엇에 대해 연구하는지, 왜 연구하는지, 어떤 형태의 연구가 발견될 가능성이 있는지, 이러한 연구에서 어떤 문제가 발생할 수 있는지 생각해보는 것이 도움이 된다. 발생할 수 있는 문제를 파악하는 것은 부분적으로 연구 주제 내용 전문가의 지식에 기초할 것이다. 또한 연구팀은 이해 상충이 전문가의 판단에 영향을 미칠 수 있는지 여부를 다루어야 한다.

연구질문의 특징은 NRSI에서 평가되는 중재를 정의하는데 어려움이거나, 중재결과 영역 또는 측정 시기를 측정하는데 사용되는 도구와 관련하여 발생할 수 있는 복잡성을 강조할 수 있다. 이상적으로는 프로토콜 단계에서 비돌림위험 평가를 준비할 때나 문헌고찰을 수행할 때 이러한 복잡성을 담기 위해 문헌고찰 저자가 계획하는 방법을 명시할 것이다.

2.1.2. 문헌고찰 대상인 모든 또는 대부분의 연구와 관련된 교란 영역 작성하기

관련 있는 교란 영역은 어떤 사람이 하나 또는 다른 관심 중재를 받는지를 예측하는 예후 인자이다. 이들은 문헌고찰 그룹의 구성원인 연구 주제 전문가의 지식을 통해서, 그리고 문헌고찰 초창기 범주화를 통해서 정해질 수 있다. 목표 환자 또는 인구 집단에 대한 중재 결정을 내리는 의료 전문가와의 토론 또한 도움이 될 수 있다. 이런 이슈는 3.1절에서 자세히 설명한다.

2.1.3. 중재 그룹 간에 다를 수 있고 연구 중재결과에 영향을 줄 수 있는 가능한 공동-중재 작성하기

관련 있는 공동-중재는 어떤 사람이 관심 중재의 시작과 함께 혹은 그 이후에 받을 수 있는 중재 혹은 노출이며, 이는 그들이 받는 중재와 관련이 있고 관심 결과에 대해 예후적이다. 이것들은 또한 문헌고찰 초창기 범주화를 통해서 문헌고찰 그룹의 구성원인 연구 주제 전문가의 지식을 통해서 정해질 수 있다. 이런 이슈는 3.4절에서 자세히 설명한다.

Box 3-1. ROBINS-I 도구(1단계): 프로토콜 단계

문헌고찰 질문을 명확하게 기술하십시오.

연구대상자	
실험 중재	
비교 중재	
중재결과	

모든 혹은 대부분의 연구에서 중요한 교란 영역의 목록을 작성하십시오.

중재군에 따라 다를 수 있고 중재결과에 영향을 줄 수 있는 공동 중재의 목록을 작성하십시오.

2.2. 개별 연구에서 사전에 고려할 사항

2.2.1. 연구에 특정한 목표 임상시험을 구체화하기

NRSI를 실용적 무작위 임상시험을 모방하는 시도로 간주함으로써 비풀림위험의 평가를 용이하게 할 수 있는데, 이를 목표 임상시험이라고 한다. 특정 연구에 대한 ROBINS-I 평가의 첫 번째 과정은 목표 임상시험을 구체화하는 것이다(Box 3-2). 목표 임상시험은 고려중

인 NRSI에 비뚤림이 없는 경우 연구결과가 동일할 것으로 예상되는 가설적인 무작위 임상 시험이다. 그것의 주요 특징은 (배제/포함기준을 비롯한) 연구대상자의 유형과 실험 및 비교 중재에 대한 설명이다. 이러한 이슈는 Hernán(2001)에 의해 더 자세히 고려된다. 개별 NRSI에 대한 목표 임상시험과 문헌고찰에 대한 일반적인 연구질문의 차이는 비뚤림위험보다는 이질성 그리고/혹은 일반화 가능성과 관련이 있다.

목표 임상시험을 구체화할 때, 이는 가설일 뿐이기 때문에 윤리적인 문제와 실제 가능한 지에 대해서는 고려할 필요가 없다. 예를 들어 담배를 피우는 사람과 피우지 않는 사람을 비교하는 목표 임상시험에 대해 반대하지 않는다. 비록 그러한 임상시험이 실제로는 윤리적이 아니거나 실행 가능하지는 않더라도 말이다.

목표 임상시험에 적합한 환자 그룹을 선택하려면 자세한 고려사항이 필요할 수 있으며 많은 환자가 제외될 수 있다. 예를 들어 Magid 등(2010)은 고혈압의 2차 치료법으로 베타 차단제와 비교하여 ACE 억제제의 비교효과를 연구하였다. 1.6백만명의 환자의 초기 코호트에서 연구자들은 분석 대상을 (1) 고혈압이 발생한 환자, (2) 티아지드(thiazide)제로 처음 치료받은 사람, (3) 관심 있는 두 약제 중 하나를 조절되지 않은 고혈압에 대한 두 번째 제제로 추가한 사람, (4) 어떤 약물에도 금기사항이 없는 사람으로 선정하였다. 그들의 “비교 효과” 코호트에서는 15,540명의 사람들이 포함되었는데 원래 코호트의 1% 미만에 해당한다.

2.2.2. 관심 효과를 구체화하기

목표 임상시험에서 어떤 특정 연구문제에 대한 관심 효과는 추적 관찰기간 동안 받는 중재의 정도에 관계없이 기저 상태에서 중재에 배정하는 효과이거나 프로토콜에서 명시한 중재의 시작 및 준수의 효과이다(Box 3-2). 이러한 효과들 사이에서 선택은 문헌고찰 저자의 결정 사항이다.

무작위 임상시험의 맥락에서 중재의 배정 효과는 연구대상자가 무작위 배정된 중재 그룹에서 분석되는, 배정된 대로(ITT) 분석을 통해 추정될 수 있다. 무작위된 중재에 대한 비준수가 있는 경우, 위약-대조 임상시험의 ITT 분석에서 모든 연구대상자가 무작위 배정을 준수하였을 경우에도 중재 효과는 과소 추정된다. ITT 효과는 위약-대조 임상시험에서 추정된 중재의 바람직한 효과에 비해 보수적인 것으로 간주될 수 있지만 둘 이상의 활성 중재를 비교하는 임상시험에서는 보수적이지 않을 수 있으며, 비열등성 연구나 동등성 연구, 또는 위해를 추정할 때는 문제가 있다.

환자와 다른 이해 당사자들은 종종 임상시험 프로토콜 준수 효과에 관심이 있다. 프로토

콜 준수 효과는 중재의 유해효과(또는 의도하지 않은 효과)를 고려할 때 관심을 가질 만한 효과이기도 하다. 무작위 임상시험의 자료를 사용하여 중재 시작 및 준수의 효과를 추정할 수 있다. 그러나 무작위 임상시험에 대해 보고한 자료만을 가지고 그렇게 하는 접근 방식은 때로 문제가 된다. 특히, 실제로 받은 치료에 기초한 보정되지 않은 분석 또는 임상시험 프로토콜을 준수하는(또는 이들이 받은 후속 조치에서) 각 중재 그룹의 개인에게만 제한되는 naïve ‘per protocol’ 분석은 예후요인이 치료에 영향을 미치는 경우 편향될 수 있다. 고급 통계방법은 이러한 비틀림에 대해 적절한 보정을 허용하지만, 이러한 방법을 적용하는 경우는 비교적 드물다. 무작위 배정 상태를 도구 변수로 사용하는 대안적인 방법은 예후 요인에 대한 보정의 필요성을 건너뛰지만, 항상 적용할 수 있는 것은 아니다.

이러한 내용을 NRSI에도 정의할 수 있다. 예를 들어 ‘중재 배정’ 효과는 실험 중재의 시작 대 비교 중재의 시작의 효과에 의해 거의 정확하게 계산할 수 있으며, 이는 중재에 배정된 연구대상자가 항상 그 중재를 시작하는 임상시험의 중재배정 효과에 해당한다. 이는 무작위 임상시험의 ITT 효과와는 약간 다른데, 왜냐하면 어떤 중재에 무작위 배정된 일부 사람들은 그것을 시작하지 않을 수 있기 때문이다. 임상시험 프로토콜에서 중재준수효과와의 유사점은 의료적 이유(예: 독성)로 중단되지 않는 한 실험 중재로 시작 및 준수하는 것 대 비교 중재의 시작 및 준수하는 것을 의미한다.

예를 들어 암 검진 연구에서 관심 효과는 검진에 대한 초청 여부(검진의 무작위 임상시험의 ITT 분석에서 측정된 효과), 혹은 검진 초청의 수락 여부와 관련될 수 있다.

무작위 임상시험과 NRSI의 경우, 중재 시작 및 준수의 효과에 대한 비틀림 없는 추정은 의도한 중재로부터의 이탈을 예측할 수 있는 예후인자에 대해 적절히 보정하는 것이 필요하다(“시간-변이 교란요인”, 3.1.9절과 3.4절의 자세한 논의 참조). 문헌고찰 저자는 시간-변이 교란에 대해 보정하는 방법을 사용하여 추정된 중재 효과를 평가할 때 전문가의 조언을 구해야 한다.

무작위 임상시험과 NRSI에서, 비틀림위험 평가는 관심있는 특정 효과와 관련이 있어야 한다. **관심 효과가 기저상태에서 배정되는 효과(무작위 임상시험)이거나 기저상태에서 시작된 중재(NRSI)일 때**, 두 연구 유형의 비틀림위험 평가에서 잠재적으로 편향된 연구자들의 행동 외에 어떤 사건의 자연스러운 과정을 반영하는, **기저상태 이후 의도한 중재로부터의 이탈(예를 들어 환자의 상태가 갑자기 악화되어 임상적 필요성에 의해 무작위 중재에서 이탈함)에 관심을 가질 필요는 없다.** 관심의 효과가 중재 개시 및 준수인 경우, 무작위 임상시험과 NRSI 모두 비틀림위험 평가는 중재 그룹간의 추가적인 중재(공동-중재)의 준수와 차이를 고려해야 할 수 있다. 이러한 이슈의 보다 자세한 논의는 3.1.8, 3.1.9, 3.4절에서 설명될 것이다.

2.2.3. 교란요인과 공동중재에서 사전에 고려할 사항

우리는 도구를 올바르게 완성하기 전에 두 가지 주요 영역에서 연구를 자세히 검토할 것을 권고한다(Box 3-2). 이 두 영역은 교란요인과 공동 중재이다. 이 과정은 임상시험 프로토콜에 명시된 중요한 교란요인과 공동 중재가 현재 연구에서 측정 또는 관리되는지, 그리고 추가적인 교란요인과 공동중재가 이 연구에서 확인되었는지 여부를 결정해야 한다. 평가에 대한 추가적인 지침과 구조는 3.1절과 3.4절에서 설명될 것이다.

Box 3-2. ROBINS-I 도구(2단계, 파트 1): 개별 연구에서: 평가 셋업하기

연구에 특정한 목표 임상시험을 지정하십시오.

연구설계	개인 무작위 / 집단 무작위 / 매치 (예: 교차)
연구대상자	
실험 중재	
비교 중재	

이 연구의 목적은?

중재에 대한 배정의 효과를 평가함

중재의 시작 및 준수의 효과를 평가함

중재결과를 지정하십시오.

비돌림위험평가를 시행하는 중재결과를 구체적으로 기술하십시오(일반적으로 결과 요약표에 표시되어 있는 것 중). 이것이 중재의 이득인지 위해인지 명시할 것

평가할 수치형 연구결과를 구체적으로 기술하십시오.

분석이 여러 번 시행된 경우 어떤 결과인지(예: RR=1.52 (95% CI 0.83-2.77)), 어디에서 자료를 얻었는지 명시할 것(예: 표, 그림, 단락)

Box 3-3. ROBINS-I 도구(2단계, 파트 2) : 개별 연구에서: 교란 영역과 공동 중재의 평가

교란인에 대한 사전 고려사항

중요한 교란 영역에 대한 행을 작성하십시오. i) 프로토콜에 작성된 것, ii) 이 연구의 세팅과 관련이 있거나 저자가 잠재적으로 중요하다고 간주하는 것. "중요한" 교란 영역은 본 연구의 맥락에서 보정이 중재의 추정 효과에 임상적으로 중요한 변화를 초래할 것으로 예상되는 영역이다. "타당도"는 교란 변수 또는 변수가 영역을 완전히 측정하는지 여부를 나타내며, "신뢰도"는 측정의 정밀도를 나타낸다(측정 오류가 클수록 신뢰도가 떨어짐).

(i) 문헌고찰 프로토콜에 작성된 교란 영역		
교란 영역	측정된 교란 변수	이 변수에 대한 통제가 불필요하다 는 근거가 있는가?*
		이 변수는 교란 영역을 타당하고 신뢰성 있게 측정하였는가?
		그렇다 / 아니다 / 정보 없음
		선택 사항: 이 변수에 대한 보정 실패는 실험 중재나 비교 중재 중 어느 쪽에 유리할 것으로 예상되는가?
		실험 중재 선호 / 비교 중재 선호 / 정보 없음

(ii) 이 연구의 세팅과 관련되거나 저자가 중요하다고 판단한 추가 교란 영역		
교란 영역	측정 변수	이 변수에 대한 통제가 불필요하 다는 근거가 있는가?*
		이 변수에 의해 교란 영역이 타당 하고 신뢰성 있게 측정되는가?
		그렇다 / 아니다 / 정보 없음
		선택 사항: 이 변수에 대한 보정 실패는 실험 중재나 비교 중재 중 어느 쪽에 유리할 것으로 예상되는가?
		실험 중재 선호 / 비교 중재 선호 / 정보 없음

* 특정 연구의 맥락에서, 변수는 교란요인이 아닌 것은 분석에 포함되지 않는 것으로 나타날 수 있다. (a) 중재결과를 예측하지 않는 경우, (b) 중재를 예측하지 않는 경우, (c) 보정이 1차 매개 변수(따라미 타)의 추정된 효과와 거의 또는 전혀 차이가 없는 경우. "통계적으로 유의미한 연관성이 없음"은 "예측하지 않음"과는 같지 않음에 주의하십시오.

공동 증제에 대한 사전 고려사항

중요한 공동 증제에 대한 행을 작성하십시오. (i) 프로토콜에 작성된 것. (ii) 이 연구의 세팅과 관련이 있거나 저자가 중요하다고 여기는 것. "중요한" 공동 증제는 본 연구의 맥락에서 보정이 증제의 추정 효과에 임상적으로 중요한 변화를 초래할 것으로 예상되는 것이다.

(i) 프로토콜에 작성된 공동 증제

공동 증제	이 증제에 대한 통제가 불필요하다는 근거가 있는가?(예: 관리하지 않았기 때문에)	이 공동 증제는 실험 증제나 비교 증제 중 어느 쪽에 유리할 것으로 예상되는가?
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음

(ii) 이 연구의 세팅과 관련되거나 저자가 중요하다고 판단한 추가 공동 증제

공동 증제	이 증제에 대한 통제가 불필요하다는 근거가 있는가?(예: 관리하지 않았기 때문에)	이 공동 증제는 실험 증제나 비교 증제 중 어느 쪽에 유리할 것으로 예상되는가?
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음
		실험 증제 선호 / 비교 증제 선호 / 정보 없음

2.3. 신호 질문

이 도구의 주요 특징은 각 비플림 영역 내에 신호 질문을 포함시키는 것이다. 이것들은 본질적으로 비플림위험에 대한 판단을 용이하게 하는 것을 목적으로 한다.

신호 질문에서 응답 옵션은 다음과 같다.

- ① 그렇다
- ② 아마도 그렇다
- ③ 아마도 아니다
- ④ 아니다
- ⑤ 정보 없음

첫 신호 질문(1.1 교란으로 의한 비플림 평가)의 경우 ‘정보 없음’의 옵션이 없다.

몇 가지 신호 질문은 특정 상황에서만 응답한다. 예를 들어 이전 질문에 대한 대답이 ‘그렇다’ 또는 ‘아마도 그렇다’(또는 ‘아니다’ 또는 ‘아마도 아니다’)만 응답을 하고 그렇지 않은 경우 ‘해당 사항 없음’이라는 옵션을 선택할 수 있다. 도구에서 **녹색으로 밑줄 친** 대답은 비플림위험이 낮을 가능성이 있는 표시이고 **빨간색**으로 표시된 것은 비플림위험이 발생할 가능성이 있는 표시이다. 신호 질문 답변의 일치도를 구할 경우 이러한 질문의 대답 쌍을 그룹화하는 것을 권장한다.

2.3.1. 신호 질문과 함께 있는 프리 텍스트 박스(free-text box)

각 신호 질문과 함께 프리 텍스트를 위한 공간이 있다. 각 대답에 대한 판단근거를 제공하기 위해 사용해야 한다. 가능한 경우 연구 논문의 본문에서 간단히 이용하도록 한다.

2.4. 비플림위험에 대한 영역-수준의 판단

ROBINS-I는 계층적으로 구성된다: 신호 질문에 대한 대답(상대적인 사실, “무슨 일이 있었는지” 또는 “연구자들이 한 일”)은 RoB에 대한 영역-수준 판단의 기초를 제공하며, 이는 특정 중재결과에 대해 전반적인 RoB 판단의 기초를 제공한다. 두 번째와 세 번째 단계를 설명하기 위해 ‘판단’이라는 단어를 사용하는 것은 매우 중요한데, 왜냐하면 문헌고찰 저자는 특정 영역에서의 비플림의 심각성과 다른 영역에서의 비플림의 상대적인 결과를 모두 고려

해야 하기 때문이다. 이 도구를 적용하는 핵심은 연구결과의 신뢰성에 대해 비뚤림의 영향에 대한 우려와 관련하여 영역 전체에서 동일한 비뚤림위험이 있는지 영역 수준에서 판단을 하는 것이다. 영역-수준의 판단이 일관되게 이루어지면 특정 중재결과에 대한 전체 RoB를 판단하는 것이 비교적 간단하다(2.5절 참조).

7가지 영역에서 비뚤림위험을 판단하기 위한 기준을 제공한다. 어떤 영역에서 신호 질문의 대답이 잠재적인 문제를 시사하지 않는다면, 영역에 대한 비뚤림의 위험이 낮다고 판단할 수 있다. 그렇지 않다면 비뚤림 위험이 있는 것으로 판단한다. 그런 다음 문헌고찰 저자는 연구결과가 편향될 정도를 판단해야 한다. “비뚤림위험”은 “실제적 비뚤림위험(risk of material bias)”으로 해석된다. 즉 연구에서 타당한 결론을 도출할 수 있는 능력에 영향을 줄 수 있는 문제에 대해서만 우려를 표명해야 한다: 사소한 비뚤림위험이 높다고 “높은 비뚤림위험”으로 간주되어서는 안된다.

“정보 없음”의 범주는 판단을 허용하기에 충분하지 않은 자료가 보고된 경우에만 사용해야 한다.

각 영역-수준의 RoB 판단에서 응답 옵션은 다음과 같다.

- ① 낮은 비뚤림위험(이 영역에 관련하여 잘 수행된 무작위 임상시험과 유사하다)
- ② 중등도 비뚤림위험(이 영역과 관련하여 비무작위연구로서는 적합하지만 잘 수행된 무작위 임상시험과 유사하다고 간주할 수 없다)
- ③ 높은 비뚤림위험(이 영역에서 몇 가지 중요한 문제가 있다)
- ④ 매우 높은 비뚤림위험(이 영역에서 중재의 효과에 대해 유용한 근거를 제공하기에는 너무 문제가 많다)
- ⑤ 이 영역에 대한 비뚤림위험을 판단하기 위한 근거가 되는 정보가 없다.

무작위 임상시험과 비무작위 연구 사이의 구별을 강조하기 위해 “낮은 비뚤림위험”의 범주가 존재한다. 이러한 차이점은 “중재 전”과 “중재 시” 영역에 적용된다(표 3-1 참조). 특히 우리는 비무작위 연구의 연구설계 특성상 매우 드문 연구만이 교란으로 인한 낮은 비뚤림위험의 범주로 평가될 것으로 예상된다. 무작위화로는 중재 후 비뚤림을 막을 수 없으므로, 중재 후 영역에서 무작위 임상시험의 평가와 NRSI의 평가 사이에 더 많은 중복이 예상된다. 그러나 연구대상자, 보건의료 전문가 또는 중재결과 평가자의 눈가림과 같은 무작위 임상시험의 다른 특징은 중재 후 비뚤림을 막을 수 있다.

2.4.1. 비뚤림위험 판단과 함께 있는 프리 텍스트 박스

각 RoB 판단과 함께 판단을 뒷받침하는 이유를 설명할 수 있는 프리 텍스트의 공간이 있다. 비뚤림위험이 “높은” 혹은 “매우 높은” 경우 판단의 이유를 제공하는 것이 필수적이다.

2.4.2. 비뚤림의 방향

확인된 잠재적 비뚤림의 규모와 방향을 아는 것이 매우 바람직하지만, 이는 비뚤림위험을 판단하는 것보다 훨씬 더 어렵다. 이 도구에는 각 영역 및 전체에 대한 비뚤림 방향을 판단하는 선택적 구성요소가 포함되어 있다. 일부 영역의 경우, 비뚤림은 무효점(null)을 향하거나 무효점에서 멀어지는 것으로 매우 쉽게 판단할 수 있다. 예를 들어 통계적으로 유의하지 않은 연구 결과의 선택적 비보고에 대한 의혹이 있다면 무효점에서 멀어지는 비뚤림을 암시한다. 그러나 다른 영역(특히 교란효과, 선택 비뚤림, 차등적 오분류와 같은 측정 비뚤림 형태)은 무효점과 연관하여 판단하지 말고 효과 추정치의 증가 또는 감소로 생각해야 한다(즉 실험 중재나 비교 중재를 선호하게 됨). 예를 들어 교란 비뚤림은 실제 위험비(risk ratio)가 1보다 크면 무효점으로 향하게 하여 효과 추정치를 감소시키며, 위험비가 1보다 작으면 무효점에서 멀어지게 한다. 문헌고찰 저자가 판단할 명확한 논리적 근거가 없는 경우 비뚤림의 방향을 추측하려고 시도해서는 안된다.

2.5. 비뚤림위험에 대해 전체적으로 판단하기

전체 RoB 판단에 대한 응답 옵션은 다음과 같다.

- ① 낮은 비뚤림위험(이 영역에 관련하여 잘 수행된 무작위 임상시험과 유사하다)
- ② 중등도 비뚤림위험(이 영역과 관련하여 비무작위 연구에는 적합하지만 잘 수행된 무작위 임상시험과 유사하다고 간주할 수 없다)
- ③ 높은 비뚤림위험(이 영역에서 몇 가지 중요한 문제가 있다)
- ④ 매우 높은 비뚤림위험(이 영역에서 중재의 효과에 대해 유용한 근거를 제공하기에는 너무 문제가 많다)
- ⑤ 이 영역에 대한 비뚤림위험을 판단하기 위한 근거가 되는 정보가 없다.

표 3-2는 영역 내 RoB 판단을 중재결과에 대한 영역 전체의 단일 RoB 판단에 매핑하는데 사용되는 기본 접근 방식을 보여준다.

표 3-2. 특정 중재결과에서 전체 RoB 판단하기

응답 옵션	기준
① 낮은 비플립위험(연구는 이 영역에 관련하여 잘 수행된 무작위 임상시험과 유사하다)	모든 영역에서 낮은 비플립위험에 있다고 판단된다.
② 중등도 비플립위험(연구는 이 영역에 관련하여 비무작위연구로는 적절한 근거를 제공하지만 잘 수행된 무작위 임상시험과 유사하다고 간주할 수 없다)	모든 영역에서 낮거나 중등도의 비플립위험에 있다고 판단된다.
③ 높은 비플립위험(연구는 이 영역에 관련하여 몇 가지 중요한 문제가 있다)	최소한 하나의 영역에서 높은 비플립위험에 있다고 판단되나, 어떤 영역에서도 매우 높은 비플립위험에 있지는 않다.
④ 매우 높은 비플립위험(연구는 이 영역에 관련하여 중재의 효과에 대해 유용한 근거를 제공하기에는 너무 문제가 많다)	최소한 하나의 영역에서 매우 높은 비플립위험에 있다고 판단된다.
⑤ 연구는 이 영역에 관련하여 비플립위험을 판단하기 위한 근거가 되는 정보가 없다.	높거나 매우 높은 비플립위험에 있다는 명확한 내용이 없고, 하나 이상의 핵심 비플립 영역에서 정보가 부족하다.

연구를 개별 영역에 특정 수준의 비플립위험에 있다고 선언하는 것은 연구 전체가 적어도 이 정도로 심각하게 (평가되는 중재결과에 대해) 비플립위험이 있다는 것을 의미할 것이다. 따라서 어떤 영역에서든 “높은 비플립위험”으로 판단한 경우 연구 전체에서 어떤 영역이 평가되고 있는지에 관계없이 유사한 의미를 가져야 한다.

NRSI가 교란으로 인한 비플립위험이 낮다고 판단되는 경우는 드물기 때문에, 대부분의 NRSI는 전체적으로 최소한 중등도의 비플립위험이 있는 것으로 판단된다.

영역-수준 판단을 설명된 전체 판단에 매핑하는 것은 프로그래밍 가능한 알고리즘이다. 그러나 실제로 일부 ‘높은’ 비플립위험(또는 ‘중등도’ 비플립위험)은 부가적인 것으로 간주될 수 있으므로 여러 영역에서 ‘높은’ 비플립위험은 전체적인 판단에서 ‘매우 높은’ 비플립

위험으로 이끌 수 있다(그리고 유사하게 여러 영역에서 ‘중등도’ 비뚤림위험은 전체적인 판단에서 ‘높은’ 비뚤림위험으로 이끌 수 있다).

2.6 문헌고찰에서 여러 중재결과에 대해 비뚤림위험을 평가하기

ROBINS-I는 NRSI의 특정 연구 결과에서의 비뚤림위험을 다룬다. 중재결과에 따른 비뚤림위험은 동일한 중재결과에 대한 다른 분석(예: 다른 분석이 다른 교란에 따라 보정되는 경우) 뿐만 아니라 다른 중재결과에 따라서도 매우 다를 수 있다. 체계적 문헌고찰에 포함된 NRSI는 (일반적이진 않지만) 빈번하게 여러 중재결과에 대한 연구 결과를 제공하므로 각 연구에 대해 몇 가지 비뚤림위험 평가가 필요할 수 있다. 표 3-3은 O1(예: 사망률), O2(예: 바이러스 부하), 그리고 O3(예: 삶의 질)의 세 가지 결과를 다루는 가상의 NRSI에 대한 가능한 평가의 예를 보여준다.

표 3-3. 특정 중재결과에 대해 전체 RoB 판단하기

영역	중재결과 평가	코멘트
교란으로 인한 비뚤림	O1: 높은 위험	예: 계수 변수만 이용 가능함(교란요인에 대해 보정하지 않음)
	O2: 중등도 위험	예: 적절히 보정됨
	O3: 높은 위험	예: 불충분한 보정
연구대상자 선택의 비뚤림	O1, O2, O3: 낮은 위험	예: 모든 중재결과에 적용되는 것으로 생각되는 같은 이슈들
중재 분류의 비뚤림	O1, O2, O3: 낮은 위험	예: 모든 중재결과에 적용되는 것으로 생각되는 같은 이슈들
의도한 중재에서 이탈되어 나타나는 비뚤림	O1, O2, O3: 중등도 위험	예: 모든 중재결과에 적용되는 것으로 생각되는 같은 이슈들
결측치로 인한 비뚤림	O1: 낮은 위험	예: 모든 사람이 의무기록에서 추적조사됨.
	O2, O3: 정보 없음	예: 탈락으로 인해; 같은 연구대상자

중재결과 측정 비플림	O1, O2: 낮은 위험	예: 둘 다 객관적 측정
	O3: 높은 위험	예: 중재결과 평가자 눈가림을 잘못해서 나타나는 비플림의 가능성
보고된 연구결과 선택의 비플림	O1: 중등도 위험	예: 조작된 것 같지 않음
	O2: 중등도 위험	예: 조작된 것 같지 않음
	O3: 높은 위험	예: 타당한 이유 없이 사용된 것 포인트

이를 통해 표 3-4에 나와있는 RoB 프로파일이 나올 수 있다(이 프로파일은 메타분석 그 리고/혹은 GRADE 평가와 함께 제시할 수 있음).

표 3-4. 서로 다른 중재결과에 대한 다른 RoB 판단 예시

영역	O1	O2	O3
교란으로 인한 비플림	높은 위험	중등도 위험	높은 위험
연구대상자 선택의 비플림	낮은 위험	낮은 위험	낮은 위험
중재 분류의 비플림	낮은 위험	낮은 위험	낮은 위험
의도한 중재에서 이탈되어 나타나는 비플림	중등도 위험	중등도 위험	중등도 위험
결측치로 인한 비플림	낮은 위험	정보 없음	정보 없음
중재결과 측정 비플림	낮은 위험	낮은 위험	높은 위험
보고된 연구결과 선택의 비플림	중등도 위험	중등도 위험	높은 위험
전체	높은 위험	중등도 위험	높은 위험

3. 도구 사용 지침: 각 비뚤림 영역에 대한 자세한 지침

3.1. 세부 지침: 교란으로 인한 비뚤림

3.1.1. 배경

교란 영역은 개인이 하나 또는 다른 관심 중재를 받는지 여부를 예측하는 중재 전 예후 인자이다. 몇 가지 일반적인 예는 기존 질병의 심각성, 동반 질환의 존재, 보건의료서비스 이용 정도, 비판 및 사회 경제적 상태이다. **교란 영역**은 하나 이상의 특정 변수들의 범위를 측정함으로써 특징화할 수 있다. 관련 교란변수 영역은 연구 세팅에 따라 다르다. 예를 들어 관심있는 중재에 대한 접근이 보편적인 국가에서 수행된 연구에서는 사회 경제적 상태가 교란을 일으키지 않을 수 있으며 이 때 사회 경제적 상태는 제공받은 중재에 영향을 미치지 않는다.

이 도구는 기저상태 교란(baseline confounding)과 시간 변동 교란(time-varying confounding) 두 가지 유형의 교란을 다룬다.

3.1.2. 기저상태 교란

기저상태 교란은 하나 이상의 중재 전 예후 인자가 추적 시작시 받은 중재를 예측할 때 발생한다. 중재 전 변수는 관심있는 중재가 시작되기 전에 측정되는 변수이다. 예를 들어, 2개의 항레트로바이러스 약물 요법을 비교하는 비무작위연구는 항레트로바이러스 요법 시작 전에 측정된 CD4 세포 수를 통제해야 한다. 이는 AIDS 및 사망에 대해 강하게 예후적이고 요법의 선택에 영향을 미칠 수 있기 때문이다. 기저상태 교란변수는 대부분 또는 모든 NRSI에서 문제가 될 수 있다.

3.1.3. 시간 변동 교란

시간 변동 교란은 제공받은 중재가 시간이 지남에 따라 바뀔 수 있는 경우(예를 들어, 개인이 비교 중인 중재를 교차하는 경우) 및 기저상태 후 예후인자(post-baseline prognostic factors)가 기저상태 이후 제공받은 중재에 영향을 미칠 때 발생한다. 기저상태 후 변수는 기저상태 이후에 측정되는 변수로, 예를 들어 치료 개시 6개월 후 측정된 CD4 세

포 수이다. 시간 변동 교란은 개별 연구대상자의 추적조사 기간을 다른 중재 그룹에서 보낸 시간에 따라 분할하는 연구에서 고려될 필요가 있다.

예를 들어, HIV 치료 환자에 대해 추적조사 기간을 다른 항레트로바이러스요법을 받았던 기간으로 분할하고, 분석에서 이 기간 동안의 의료결과를 비교하는 연구를 가정하자. (기저상태 후 예후 변수로서) CD4 세포 수는 관심 요법들 간의 변화에 영향을 미칠 수 있다. 기저상태 후 예후 변수가 중재 그 자체에 의해 영향을 받는 경우(예를 들어 항레트로바이러스요법이 기저상태 후 CD4 수에 영향을 미칠 수 있음), 통계 분석에서 이들에 대한 통상적인 보정은 교란변수를 통제하는 수단으로서 적합하지 않다. 개인들이 비교되는 중재를 교차할 때, 관심 효과는 중재에 대한 배정 효과가 아니라 중재 시작과 준수 효과임을 명심해야 한다.

다른 예로, 기존 요법을 받는 대조군과 당뇨병을 위해 새로운 약을 복용하는 환자들의 심혈관계질환 사건을 비교하는 대규모 비교 개방(open) NRSI가 있다고 해보자. 연구의 추적조사기간 동안 출판된 연구근거는 새로운 당뇨병 약이 심혈관계 사건 위험을 증가시켰음을 시사하였다. 연구 참여 후 혈압 또는 지질 수준이 악화된 환자는 심혈관계 위험을 우려하여 새로운 약물에서 (기존 약으로) 변경되었다. 혈압과 지질 수준은 심혈관 사건에 대해 예후적이었고, 그리고 기저상태 이후에 받은 중재를 예측했기 때문에, 연구는 시간 변동 교란으로 인해 비뚤림의 위험이 있었다. 이러한 문제는 3.1.8 및 3.1.9 절에서 설명한다.

3.1.4. 교란 영역 확인하기

중요한 교란 영역은 NRSI 문헌고찰 리뷰 프로토콜에서 미리 지정해야 한다. 잠재적 교란 영역을 조사하려면 주제에 대한 지식이 필요하다. 예를 들어 최소 침습적 수술 전략과 개방 수술 전략을 비교한 관찰 연구에서, ASA(American Society of Anesthesiologists) 등급 또는 Charlson 지수에 의해 측정된 수술에 대한 중재 전 적합성(동반질환)에 대해 보정하지 않으면 이 요인이 수술 전략의 선택을 예측한다면 이는 교란을 야기할 것이다.

수술 전문가가 수술 전략 선택과 관련이 있을 것으로 예상되는 예후 요인을 확인하는 데 가장 적합하다. 따라서 아래에 설명된 절차는 연구중인 주제에 대해 잘 알고 있는 평가자가 사용하도록 설계되었다. 문헌고찰 프로토콜을 작성하는 팀에 주제 전문가를 포함시키고 고찰 저자와 초기 토론과 기존 문헌 지식을 바탕으로 문헌고찰 프로토콜에 교란 영역을 나열할 것을 권장한다.

문헌고찰 프로토콜 작성 후, 심지어 체계적 문헌고찰에 포함시키기 위해 선택된 연구에서

자료수집을 예비로 수행한 후에도 교란 및 기타 잠재적인 비뿔림과 관련된 새로운 아이디어가 발견될 수 있다. 예를 들어, 이러한 이슈는 하나 이상의 논문의 서론 및/ 또는 논의에서 언급되기 때문에 확인될 수 있다. 이것은 잠재적인 교란이나 기타 비뿔림의 원인이 논문에 언급되어 있는지를 자료수집 부분에 명시적으로 기록함으로써 해결할 수 있다.

드물거나 비정상적인 부작용으로 인해, 근본적인 위험 요소를 알 수 없으며 사전에 교란의 출처를 확인하기가 어려울 수 있다. 예를 들어, 신성 전신 섬유증(nephrogenic system fibrosis)은 병인학적 요인과 자연사가 아직 밝혀지지 않은, 최근에 발견된 드문 부작용이다. 이러한 특정 상황에서, 문헌고찰 저자는 관련 교란의 출처를 미리 구체화하지 못하거나, 이러한 부작용을 평가하는 연구가 교란변수를 적절히 다루었는지를 판단할 수 없다.

반면에, 문헌고찰 저자는 중재를 배정한 사람들이 부작용의 가능성을 인식하지 못하여 그 부작용에 대한 위험 요소에 근거한 치료 결정을 내릴 수 없다고 생각하는 경우 교란을 인정하기 어렵다고 판단할 수 있다. 부작용이 알려진 부작용 사건(예를 들어, 위의 신성 전신 섬유증 예에서 신장 기능 저하)의 결과이거나 이와 관련이 있는 경우에도 교란은 여전히 존재할 수 있음을 유의하라.

3.1.5. 잔류 및 측정되지 않은 교란(Residual and unmeasured confounding)

교란 영역은 직접 측정되지 않을 수 있으므로, 연구자들은 이러한 교란변수 영역을 완전히 또는 부분적으로 보정하려는 시도에서 특정 변수(종종 교란변수라고 언급되는)를 측정한다. 예를 들어, 기저상태 CD4 세포 수 및 최근 체중 감소가 질병 중증도를 보정하기 위해 사용될 수 있으며; 기저상태 6개월 전의 입원 및 외래진료 수는 의료서비스 이용률을 보정하기 위해 사용하고; 의사 처방 실무를 보정하기 위해 지리적 위치; 비만을 보정하기 위해 체지방 지수와 허리 대 엉덩이 비율; 사회 경제적 상태를 보정하기 위해 소득과 교육수준이 사용될 수 있다.

교란이 완전히 통제되지 않는 두 가지 광범위한 이유가 있다. 잔류 교란은 교란변수 영역이 잘못 측정되거나 교란변수 영역과 중재결과 또는 노출(사용되는 분석 방법에 따라 다름) 사이의 관계가 불완전하게 모델링 될 때 발생한다. 예를 들어 두 가지 항고혈압제를 비교한 NRSI에서 중재 시작 3개월 전에 중재 전 혈압을 측정할 경우, 우리는 잔류 교란변수를 예측한 것이지만, 중재시점에 약물들을 결정하기 위해 임상수가 사용한 혈압은 우리의 데이터 세트에서는 접근가능하지 않았다.

측정되지 않은 교란변수는 교란변수 영역이 측정되지 않았거나 분석에서 통제되지 않은

경우에 발생한다. 이것은 중재 전 혈압 측정값이 없거나, 측정되었음에도 불구하고 분석을 통해 중재 전 혈압을 통제하는데 실패한 경우이다.

보건의로 전문가가 중재 결정을 내릴 때, 그들이 이용 가능한 정보에서 측정 오류가 있다고 반드시 잔류 교란을 유발하지는 않는다. 예를 들어, 중재 전 혈압은 기본 혈압을 완벽하게 반영하지 않는다. 그러나 두 번의 사전 중재 측정을 기반으로 중재 결정을 내리고 이러한 측정을 데이터 세트에서 사용할 수 있다면 교란을 충분히 보정할 수 있다.

일부 문헌고찰 질문에서, 치료 결정에 영향을 미치는 모든 교란 영역을 측정할 수 없기 때문에 교란이 다루기 어려울 수 있다. 예를 들어, 경구용 당뇨병 치료제가 실패할 때 인슐린으로 2형 당뇨병을 치료하는 효과에 대한 연구를 고려한다. 환자는 일반적으로 나이가 많으며 의사는 그들의 결정을 기록하지 않을 수 있으며, 대부분 인지 장애가 없고 충분한 손재주가 있는 환자들에게 인슐린 치료를 처방할 수 있다. 이것은 측정할 수 없는 잠재적으로 강력한 교란을 야기한다.

3.1.6. 교란의 통제(Control of confounding)

모든 교란변수가 오류없이 측정될 때, 교란은 연구설계(예: 기저상태 교란변수의 동일한 값을 가진 대상으로 자격을 제한함으로써)나 교란요인들을 보정(통제) 하는 통계분석을 통해 통제될 수 있다. 특정 연구 맥락에서, 교란요인이 중재와 관련이 없거나 결과와 관련이 없는 경우, 분석에서 이를 통제할 필요는 없다. 그러나 이 문맥에서 “비 관련(unrelated)”은 “연관성 없음(not associated)” (예: 위험 비율이 1에 가까움)을 의미하고, “통계적으로 유의미한 관련성이 없음(no statistically significant association)”을 의미하지 않는다는 점에 유의해야 한다.

교란을 적절히 통제하려면, 사용된 변수가 교란 영역을 측정하기에 타당하고 신뢰성이 있어야 한다. 이와 관련하여 “타당도”는 변수 또는 변수가 그 영역을 완전히 측정하는지 여부를 나타내며, “신뢰도”는 측정의 정밀도(측정 오류가 많을수록 신뢰성이 낮음)를 나타낸다 (Streiner and Norman, 2003). 일부 주제의 경우, 교란 영역에 대한 타당하고 신뢰할 수 있는 측정 목록이 미리 이용가능하며, 문헌고찰 프로토콜에서 구체화되어야 한다. 다른 주제의 경우 이러한 목록을 사용하지 못할 수 있다. 연구 저자는 특정 측정도구의 사용을 뒷받침하는 참고 문헌을 인용할 수 있다.

그런 다음 연구자는 이러한 인용을 기반으로 하여 측정도구의 타당도와 신뢰도에 대한 판단을 내릴 수 있다(Cook and Beckman, 2006). 일부 저자는 타당도 또는 신뢰도에 대한

입증없이 교란변수를 통제할 수 있다. 그러한 경우, 문헌고찰 저자는 그 측정도구의 주관성에 주의를 기울여야한다. 자가 보고(self-report)에 기반한 주관적 측정은 임상 보고서 및 실험실 소견과 같은 객관적인 측정에 비해 타당도와 신뢰도가 낮은 경향이 있다(Cook et al, 1990).

부적절한 보정이 이루어졌는지 고려하는 것이 중요하다. 특히 **중재 후 변수를 조정하는 것은 일반적으로 적합하지 않다. 매개 변수(중재에서 결과까지의 인과 경로에 있는 변수)에 대한 보정은 매개변수를 거치지 않는 중재 효과(“직접 효과”)에 대한 관심을 제한시켜, 무작위 임상시험에서도 교란을 일으킬 수 있다. 중재 및 결과의 공통 효과(common effect)에 대한 보정은 비뚤림을 유발한다. 예를 들어, 다른 항레트로바이러스 약물 조합을 비교하는 연구에서 일반적으로 중재 전 CD4 세포 수를 보정하는 것이 필수적이지만, 치료 시작 6개월 후 CD4 세포 수를 보정하는 것은 부적절하다.**

3.1.7. 음성 대조군(Negative controls)

음성대조군은 연관성이 없음을 입증해야 하는 대안적 분석에 흔히 사용된다. 이러한 “**음성 대조군**”을 때때로 측정되지 않은 교란가능성을 해결하기 위해 사용할 수 있다. Lipsitch et al(2010)은 이 문제에 대해 논의했으며 노출 대조군(exposure controls)과 중재결과 대조군(outcome controls)과 같은 두 가지 유형의 음성 대조군을 구분하였다.

이 저자들에 의해 논의된 한 예로 인플루엔자에 대한 예방 접종이 폐렴/인플루엔자 입원 위험의 감소와 모든 원인 사망률(all-cause mortality)과 관련이 있다고 제안한 노인들을 대상으로 한 관찰 연구와 관련이 있다. 이 가설을 검증하기 위해 Jackson et al(2006)은 인플루엔자 백신 접종자의 보호 효과에 대한 초기 추정치를 재현한 다음 두 세트의 음성 대조군 결과에 대해 분석했다.

먼저, 그들은 독감 시즌 이전, 도중 및 이후에 백신 접종을 한 사람과 예방 접종하지 않은 사람(“노출 대조군”)의 폐렴/인플루엔자 입원 위험과 모든 원인 사망률을 비교하였다. 그들은 이전 연구에서 측정된 효과가 인과 관계가 있는 것이라면, 그 효과는 인플루엔자 시즌에 가장 두드러져야 한다고 추론하였다. 교란을 통제하려는 노력에도 불구하고, 예방 효과(protective effects)는 실제로 인플루엔자 시즌 이전, 도중 및 이후에 모두 나타나는 것으로 나타났다. 그들은 이것이 나타난 사망감소 효과가 인플루엔자 예방효과라기 보다는 교란에 의한 것이 상당 부분을 차지한다는 근거라고 결론지었다.

둘째, 그들은 실제로 인플루엔자 접종이 예방 효과가 있다면, 인플루엔자와 관련이 있는

중재결과로 제한되어야 한다고 가정하였다. 그들은 분석을 반복했고, 최종 결과(“중재결과 대조군”)로 부상이나 외상으로 인한 입원을 보았다. 그들은 인플루엔자 예방 접종이 부상이나 외상 입원도 ‘예방’한다는 것을 발견하였다. 이것은 또한 폐렴/인플루엔자 입원 또는 사망에 대해 관찰된 일부 예방효과가 부적절하게 통제된 교란에 의한 것이라는 근거로 해석되었다. “중재결과 대조군”의 두 번째 예는 흡연과 자살에 대한 연구가 흡연과 살인 사이의 연관성도 발견했다는 것이다(Davey Smith et al, 1992).

3.1.8. 중재간 전환(Switches between interventions)

일부 (아마도 많은) NSRI, 특히 일상적으로 수집된 자료를 기반으로 하는 경우, 연구대상자들에게 제공된 중재는 추적조사 기간 동안 그들이 기저상태에서 제공받은 중재에서 문헌 고찰에서 비교되는 다른 중재로 변경될 수 있다.

이로 인해 “**관심있는 중재간에 전환**”이 발생할 수 있는데, 이러한 현상을 교란 영역에서 고려한다(아래의 “시간 변동 교란” 참조). 비교되는 중재 그룹 중 하나가 무중재인 경우, 이러한 전환은 적극적 치료의 중단 또는 대조군에 배정된 개인의 적극적인 치료 시작을 포함한다. 반면에, 기저상태 중재에서의 변경은 문헌고찰 질문에 관심이 없는 중재로 전환될 수 있다. 우리는 이러한 종류의 전환을 “의도한 중재로부터의 이탈”로 고려한다. 연구대상자가 중재를 전환하는 연구의 경우, 비돌림위험 평가는 관심 효과에 달려 있다. 두 가지 광범위한 접근 방식이 있다.

① 관심 효과는 실험 중재에 대한 배정(또는 시작)과 비교 중재에 대한 배정(또는 시작)에 대한 효과이며, 연구대상자는 초기 중재로 정의된 그룹으로 분석된다. 이 상황에서, 추적조사 기간 동안 중재간의 전환은 비돌림을 유발하지 않는다. 예를 들어, 김진에서 발견된 국소 전립선 암 환자는 즉각적인 수술이나 암의 적극적인 모니터링에 배정된 연구를 생각해 보자. 일부 남성은 수술을 받았지만 초기 중재에 따라 분석될 것이다. 또 다른 예로, 여성의 경구 피임약 선택이 유방암 위험에 미치는 영향을 조사한 연구에는 여성이 임신을 원했기 때문에 피임법 사용을 중단했는지 여부에 관계없이 모든 추적 관찰 시간이 포함된다.

② 관심 효과는 중재 시작 및 준수의 효과이며, 추적 기간은 다른 중재를 받는 시간으로 나뉜다. 예를 들어, 2개의 선택적 세로토닌 재흡수 억제제(SSRI) A와 B를 중재없음과 비교한 12개월의 연구에서, 환자는 A에서 6개월, 중재없음에서 2개월, B에서 4개월을 보냈을

수 있고, 이러한 추적기간은 분석에서 다른 중재에 배정된다. 이러한 연구 및 분석은 중재 변경 직후에 관심 중재결과의 위험이 변한다는 가정에 의존한다: 예를 들어, 연구 저자는 특정 경구 피임약의 사용과 관련된 정맥 혈전증 위험의 변화가 해당 중재 사용 직후 중단된다고 믿을 수 있다.

대조적으로, 연구 저자들은 경구 피임약이 중단된 후 유방암 위험의 변화가 상당한 기간 동안 지속될 것이라고 생각할 수 있다. “프로토콜 준수(per protocol)” 효과에 대한 추정은 특정 추적 기간 동안 유방암 위험에 대한 이전과 현재 중재의 기여도에 대한 강력한 가정이 필요하기 때문에 매우 어려울 것이다.

3.1.9. 시간 변동 교란

제공받은 중재에 따라 추적조사 시간이 분리될 때, 우리는 시간 변동 교란으로 인한 비뮴립의 위험을 평가해야 한다. 관심 중재결과에 대해 예후적이고 중재의 전환을 예측하는 요인들의 값이 시간에 따라 변하는 경우, 기저상태 교란에 대해서만 보정하는 것만으로는 충분하지 않다. 예를 들어, 비스테로이드성 항염증제(NSAIDs)가 사망률에 미치는 영향을 비교한 연구에서, 연구대상자들이 추적기간 동안 비교되는 NSAIDs들을 바꾼 경우, 추적조사 기간동안 위장 출혈 에피소드가 사망률에 대해 예후적이고 NSAIDs 간의 전환도 예측하였다면 시간 변동 교란이 발생할 것이다.

3.1.10. 기술 노트: 시간 변동 교란 보정

시간 변동 교란은 결과를 예측하는 시간 변동 인자가 중재의 변화에도 영향을 미칠 때 발생한다. 또한 중재의 과거 변경이 동일한 요인들의 후속 값에 영향을 미치는 경우, 관련 요인이 완벽하게 측정되고 그 효과가 완벽하게 모델링 되었을지라도, 표준 통계 방법들(시간 변동 요인을 포함하는 Cox 회귀 모델과 같은)은 교란에 대해 적절히 보정할 수 없다.

예를 들어 HIV 감염에 대한 항레트로바이러스 요법(ART)에 대한 연구가 있다: CD4 세포 수는 ART 준수를 예측할 수 있는 AIDS에 대한 예후 인자이며, ART 준수는 후속 CD4 수에 영향을 미친다. 이러한 상황에서, 지속적인 중재의 효과를 추정하는 것은 원칙적으로 가능하지만 시간 변동 교란을 다룰 수 있는 방법을 사용해야 한다. 일반적으로 사용되는 방법은 한계 구조 모델(marginal structural models)의 역확률 가중치(inverse probability weighting)이지만, 기술적으로 사용하기 어렵다. 이러한 방법을 사용하여 연구의 비뮴립위

험을 평가하기 위해서는 전문가의 조언을 구해야 한다.

3.1.11. 교란으로 인한 비뚫림에 대한 비뚫림위험 평가

신호 질문과 비뚫림위험 평가는 Box 3-4와 표 3-5에 제시되어 있다. 교란이 발생할 가능성이 있는 경우, 비뚫림위험의 판단은 연구대상자들이 추적조사 기간 동안 그들의 초기 중재 그룹에 남아있거나 시간 변동 교란이 예상되지 않는 연구에 대한 질문 1.4 ~ 1.6에 대한 응답과, 연구대상자들이 관심 중재들 간에 전환이 있었고 시간 변동 교란이 예상되는 연구에 대한 질문 1.7 및 1.8에 대한 응답에 근거해야 한다.

Box 3-4. ROBINS-I tool(2단계, 파트 3): 교란으로 인한 비틀림위험

신호 질문	세부 설명(elaboration)	응답 선택
<p>1.1 이 연구에서 중재 효과의 교란이 일어날 가능성이 있는가?</p> <p>1.101 N/PN 인 경우: 연구는 교란에 의한 비틀림위험이 낮은 것으로 간주될 수 있으며, 더 이상의 신호 질문을 고려할 필요가 없다.</p>	<p>드물게 치료 결정에 영향을 미치는 요인과 관련이 거의 없는 경우 교란이 예상되지 않아, 이 연구는 교란으로 인한 비틀림위험이 낮은 것으로 간주될 수 있다. 이 신호 질문에 대한 응답 옵션에 NI(정보 없음)는 없다.</p>	<p>그렇다(Y) / 아마도 그렇다(PY) / 아마도 아닐 것이다(PN) / 아니다(N)</p>
<p>1.101 Y/PY 인 경우: 시간 변동 교란요인을 평가할 필요가 있는지 결정하십시오.</p>		
<p>1.2 제공받은 중재에 따라, 연구대상자의 추적 관찰기간을 구분하여 분석하였는가?</p> <p>N/PN 인 경우, 기저 상태의 교란과 관련지어 대답하십시오.(1.4 ~ 1.6)</p> <p>Y/PY 인 경우 1.3으로 가시오.</p>	<p>연구대상자가 중재 기간 전환이 가능한 경우 시간 변동 교란에 의해 중재와 중재결과간의 연관성이 편향될 수 있다. 이는 예후 요인에 의해 의도한 중재기간 전환이 영향을 받을 경우 발생한다.</p>	<p>적용할 수 없음(NA) / Y / PY / PN / N / 정보 없음(NI)</p>
<p>1.3 중재의 중단 또는 교란이 결과에 대한 예후 인자와 관련이 있을 가능성이 있는가?</p> <p>N/PN 인 경우, 기저선의 교란과 관련지어 대답하라.(1.4~1.6)</p> <p>Y/PY 인 경우, 기저선과 시간 변동의 교란과 관련지어 대답하라.(1.7~1.8)</p>	<p>중재 전환이 중재결과와 관련이 없는 경우(예: 중재결과가 예상치 못한 위험인 경우) 시간-변동 교란은 존재하지 않으며 기저 상태 교란에 대한 통제만 필요하다.</p>	<p>NA / Y / PY / PN / N / NI</p>
<p>기저 상태 교란에만 관련이 있는 질문</p>		
<p>1.4 저자는 모든 중요한 교란 영역을 통제하는 적절한 분석 방법을 사용하였는가?</p>	<p>측정된 교란요인을 통제하는 적절한 방법으로는 중화, 회귀분석, 매칭, 표준화 및 역확률 가중(inverse probability weighting) 같은 방법들이 있다. 개별 변수 또는 추정된 상호작용으로 통제 할 수 있다. 역확률 가중치는 상호작용 함수를 기반으로 한다. 각각의 방법은 측정되지 않은 교란이나 잔류 교란이 없다는 가정에 기반한다.</p>	<p>NA / Y / PY / PN / N / NI</p>

<p>1.5. 1.40 Y/PY 인 경우: 이 연구에서 사용 가능한 변수에 의해 측정된 교란 영역이 타당하고 신뢰성 있게 통제되었는가?</p>	<p>교란을 적절히 통제하려면 보정된 변수가 교란 영역을 타당하고 신뢰성 있게 측정해야 한다. 일부 주제에서는 교란 영역에 대한 타당하고 신뢰할 수 있는 측정도구의 목록이 프로토콜에서 정해지지만, 어떤 경우는 해당 목록을 사용하지 못할 수 있다. 저자는 특정 측정 방법의 사용을 정당화하기 위해 참고문헌을 인용할 수 있다. 저자가 타당도 혹은 신뢰도 제시 없이 교란 변수를 통제하는 경우, 측정이 주관적이 될 수 있음에 주의를 기울여야 한다. 주관적 측정(예: 자기보고 기반)은 실험실 연구와 같은 객관적인 측정보다 낮은 타당도와 신뢰도를 가질 수 있다.</p>	<p>NA / Y / PY / PN / N / NI</p>
<p>1.6. 저자는 중재의 영향을 받을 수 있는 중재 후 변수를 통제하였는가?</p>	<p>중재에 의해 영향을 받게 되는 중재 후 변수를 통제하는 것은 적절하지 않다. 매개 변수에 대한 통제는 중재의 직접적인 효과를 추정하고 비동일성을 유발할 수 있다. 중재와 결과의 일반적인 효과에 대해 통제하면 비동일성이 생기게 된다.</p>	<p>NA / Y / PY / PN / N / NI</p>
<p>기저 상태와 시간 변동 교란에 대한 질문</p>		
<p>1.7. 저자는 모든 중요한 교란 영역과 시간-변동 교란을 보정하는 적절한 분석방법을 사용하였는가?</p>	<p>RCT와 NRS에서 중재의 시작 및 준수 효과를 추정하기 위해서는 시간-변동 교란에 대한 보정이 필요하다. 적절한 통계분석 방법으로는 역회귀를 기반으로 하는 방법이 있다. 시간에 따라 변하는 교란이 있다면 시간-업데이트된 교란요인을 포함하는 표준 회귀 모델이 문제가 될 수 있다.</p>	<p>NA / Y / PY / PN / N / NI</p>
<p>1.8. 1.70 Y/PY 인 경우: 이 연구에서 이용할 수 있는 변수에 의해 측정된 교란 영역이 타당하고 신뢰성 있게 보정되었는가?</p>	<p>1.5를 참고하십시오.</p>	<p>NA / Y / PY / PN / N / NI</p>
<p>비밀리위험 판단</p>		
<p>선택 사항: 교란으로 인해 예상되는 비밀리의 범형은 무엇인가?</p>	<p>표 3-5를 보시오.</p> <p>하나 이상의 중요한 교란 영역이 통제되지 않았기 때문에 실제 효과 추정값이 연구에서 추정된 효과보다 크거나 작을 것으로 예상할 수 있는가? 이 질문에 답하려면 다른 연구에서의 전문기적 지식과 결과에 기반해야 하기 때문에 근거가 되는 모든 연구를 검토한 후에야 완성될 수 있다. 측정되지 않은 각 영역의 잠재적 영향과 분석에서 통제되지 않은 모든 중요한 교란 영역이 동일한 방향으로 추정치를 변경할 가능성이 있는지 또는 분석에서 통제되지 않은 하나의 중요한 교란 영역이 지배적인 영향을 미칠 가능성이 있는지 고려해야 한다.</p>	<p>낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음</p> <p>실험 중재 신호 / 비교 중재 신호 / 예측할 수 없음</p>

표 3-5. 교란으로 인한 비뚤림에 대한 비뚤림위험 판단하기

낮은 비뚤림위험 (잘 수행된 무작위 임상시험과 유사함)	교란이 예상되지 않았음
중등도 비뚤림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과 유사하지 않음)	(i) 교란이 예상되었고, 모든 알려진 중요한 교란 영역이 적절하게 측정되고 통제되었음 그리고 (ii) 중요한 영역에 대한 측정의 신뢰도와 타당도는 충분하였으며, 심각한 잔류 교란이 예상되지 않음
높은 비뚤림위험 (몇 가지 중요한 문제가 있음)	(i) 하나 이상의 알려진 중요한 영역이 적절하게 측정되지 않았거나 통제되지 않았음 또는 (ii) 중요한 영역에서의 측정의 신뢰도 또는 타당도가 심각한 잔류 교란을 예상할 수 있을 정도로 낮았음
매우 높은 비뚤림위험 (중재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)	(i) 본질적으로 통제할 수 없는 교란 또는 (ii) 음성 비교중재에 대한 결과 내용으로 미루어 측정되지 않은 교란을 강력하게 시사됨
정보 없음 (이 영역에서 비뚤림위험에 대한 판단을 하기에)	교란이 있는지에 대한 정보가 없음

3.2. 세부 지침: 연구대상자 선택의 비뚤림

3.2.1. 서론

선택 비뚤림은 일부 적격 연구대상자 또는 일부 연구대상자의 초기 추적조사기간, 또는 일부 중재결과 사건들이 제외되어, 목표 임상시험에서 관찰될 수 있는 관련성과는 다르게 중재와 중재결과 간의 관련성에 영향을 미칠 때 발생한다. 1.3절에서 설명했듯이, 선택 비뚤림이라는 용어가 때때로 교란을 의미하기 위해 사용되지만 **이 현상은 교란과 별개이다**.

우리가 “선택 비뚤림”이라는 용어를 사용할 때는 **간접적 문제**(연구에서 제외된 사람들에 대한 일반화 가능성, 적용성 또는 전이성(transferability))가 아닌 **연구 내적인 비뚤림**만을 언급하기 위해 사용한다(Schunemann et al, 2013). 예를 들어, 동반 질환이 없는 개인으

로 연구 대상으로 제한하면 그 결과를 동반 질환이 일반적인 임상 진료에 일반화시키는 것이 제한될 수 있다. 그러나 그것은 동반 질환이 없는 개인에 대한 중재의 추정효과에 영향을 미치지 않는다.

3.2.2. 연구대상자를 선정할 때 비뚤림이 생길 수 있는 경우

연구대상자를 선택적으로 모집한다고 해서 반드시 비뚤림이 발생하는 것은 아니다. 예를 들어, 자격을 갖춘 남성의 절반만 (무작위로) 선택했지만, 자격을 갖춘 모든 여성을 포함하는 연구를 생각해보자. 남성에서의 중재 효과는 모든 남성이 포함된 경우보다 덜 정확하게 추정되지만, 여성에서 중재의 실제 효과가 남성과 동일한 경우, 효과 추정치는 선택에 의해 비뚤림이 발생하지 않는다.

연구대상자의 선택이 중재 및 중재결과와 관련될 때 선택 비뚤림이 발생한다. 예를 들어, 신경관 결합을 예방하기 위한 엽산 보충제에 대한 연구는 그들이 생존하여 출생한 대상으로 제한한 연구의 경우(Hernán et al, 2002), (표본에서 제외된) 사산과 치료적 유산이 중재 및 중재결과와 모두 관련되어 있기 때문에 비뚤림이 발생한다(Velie and Shaw, 1996; Hernán et al., 2002).

같은 이유로, 일부 추적조사 기간이 분석에서 제외될 때 선택 비뚤림이 발생할 수 있다. 예를 들어, 중재의 새로운 사용자가 아닌 기존 사용자가 분석에 포함된 경우 비뚤림위험이 가능하다. 이는 중재 시작 후 얼마 후에 목표 임상시험의 추적조사를 시작하는 것과 유사하다. 그래서 중재를 시작한 후 중재결과를 경험한 일부 대상자들은 제외될 것이다. 이는 **개시 비뚤림(inception bias)** 또는 **시간 지연 비뚤림(lead time bias)**이라고 하는 선택 비뚤림의 유형이다. 무작위 임상시험에서와 같이 연구대상자가 중재의 시작(개시)을 따르지 않는다면, 추적 기간에서 배제되고, 중재 직후 중재결과를 경험한 개인은 분석에서 누락된다. 중요한 문제는 시간이 지남에 따라 중재의 효과가 일정할 것으로 예상할 이유가 없다는 것이다. 따라서, 중재 후 즉시 추적조사를 제외하면 추정된 효과가 높게 또는 낮게 치우칠 수 있다. 추적 조사기간으로 증화된 중재의 효과에 대한 추정치를 보고하는 연구는 중재 시작 이후 시간에 따라 중재 효과가 변하는 정도에 대한 정보를 제공할 수 있다.

예를 들어, 약물을 이미 사용하는 사람들을 기반으로 한 분석은 약물에 내약성이 높은 사람들을 선택하는 경향이 있을 수 있다. “취약자의 감소”는 이미 일어날 것이다. 결과적으로 우리는 중재 그룹에서 부작용의 비율을 과소 추정할 수 있다. 따라서 약물 역학 연구는 종종 지난 12개월 동안 약물 사용에 대한 기록이 없었어야 한다고 명시하고 있다. 예를 들어,

NRSI에서 조사할 때, 새로운 경구 피임 프로게스테론을 복용한 대상자에서 정맥 혈전 색전증의 위험이 뚜렷이 증가하였다(Ray et al, 2003; Suissa et al, 2000).

신약 사용자는 더 오래된 약제의 사용자보다 최근에 치료를 시작했으며, 정맥 혈전 색전증의 위험은 치료 과정에서 초기에 가장 크다. 최신의 방법론적 표준은 이와 같은 비뚤림을 피하기 위해 의료기술의 새로운 사용자 코호트를 확인하고, 치료 시작이 아닌, 치료 결정 날짜부터 추적조사를 시작하는 것의 중요성을 강조한다(Ray et al, 2003; Suissa, 2008).

불멸의 시간 비뚤림(immortal time bias) 관련 비뚤림은 증재결과가 발생할 수 없는 추적기간이 있는 방식으로 증재가 정의될 때 발생한다. 예를 들어, 한 연구는 만성 폐쇄성 폐질환 또는 만성 심부전 환자의 코호트를 대상으로 추적하였으며, 원격 재택 치료를 받은 표준 치료를 받았는지에 따라 두 그룹으로 구분하였다. 그러나 원격 재택 치료를 받으려면 환자가 최초 입원 후 몇 주 동안 생존해야만 했기 때문에 입원과 재택 치료 시작 사이의 시간은 “불멸의 시간”이었다. 이 추적 기간과 해당 기간 동안 발생하는 사망을 제외하면, 원격 재택치료가 사망률을 감소시키는 것으로 결과를 편향시킬 것이다. 목표 임상시험과의 비교는 그러한 비뚤림의 확인을 용이하게 해야 한다. 왜냐하면 임상시험에서 연구대상자는 증재의 구현이 어느 정도 후에 발생하더라도 무작위배정 시점부터 추적될 것이기 때문이다.

3.2.3. 기술 노트 : 선택 비뚤림 보정

이러한 유형의 선택 비뚤림을 보정하기 위한 분석방법과 선택 비뚤림을 예방하는 통계 분석이 있다. 중요한 문제는 의미있는 보정(예: 역확률 가중치를 통한)을 허용하는 측정 변수를 사용할 수 있는지 여부이다. 많은 상황에서 이러한 경우가 없을 것이므로, 이러한 설계 문제는 심각하거나 치명적인 비뚤림위험으로 분류될 수 있다.

3.2.4 연구대상자 선정시 비뚤림에 대한 비뚤림위험 평가

신호 질문과 비뚤림위험 평가는 box 3-5와 표 3-6에 제시하였다.

Box 3-5. ROBINS-I 도구(2단계, 파트 4): 연구대상자 선택의 비뮴림위험

신호 질문	세부 설명(Elaboration)	응답 선택
<p>2.1 종재 시작 후 관찰된 연구대상자의 특성에 기초하여 연구 대상자를 (또는 분석 대상) 선택하였는가? 2.101 N/PN 인 경우: 2.4로 가시오.</p>	<p>이 영역은 종재 시작 후 관찰된 연구대상자의 특성에 기초한 선택에만 관심이 있다. 종재의 시작전에 관찰된 특성에 따라 선택하였다면, 실험 종재와 비교 종재 사이의 불균형을 종재결과에 대한 예측적인 기저 특성에 따라 통제함으로써 해결할 수 있다(기저 상태 고려).</p>	<p>Y / PY / <u>PN</u> / N / NI</p>
<p>2.2 2.101 Y/PY 인 경우: 선택에 영향을 준 종재 후 변수가 종재와 관련이 있을 가능성이 있었는가? 2.3 2.201 Y/PY 인 경우: 선택에 영향을 준 종재 후 변수가 종재결과 또는 종재결과와 원인에 영향을 받을 가능성이 있는가?</p>	<p>선택 비뮴림은 선택이 종재의 영향 또는 종재의 원인과 종재결과 또는 종재결과와 원인과 관련될 때 발생한다. 따라서 연구대상자의 선택이 종재와 종재결과 모두와 관련이 있다면 연구 결과는 선택 비뮴림이 있을 것이다.</p>	<p>NA / Y / PY / <u>PN</u> / N / NI NA / Y / PY / <u>PN</u> / N / NI</p>
<p>2.4 대부분의 연구대상자에서 추적 관찰의 시작 시점과 종재의 시작 시점이 일치하는가?</p>	<p>연구대상자가 종재의 시작부터 추적되지 않는다면, 추적관찰 기간에서 제외될 것이며, 종재 직후 시행된 종재결과는 결측될 것이다. 이 문제는 종재의 신규 사용자보다 기존 사용자가 분석에 포함될 때 발생할 것이다.</p>	<p><u>Y</u> / PY / <u>PN</u> / N / NI</p>
<p>2.5 2.2와 2.301 Y/PY 이거나, 2.401 N/PN 인 경우: 선택 비뮴림을 교정할 수 있는 보정 방법을 사용하였는가?</p>	<p>예를 들어, 역확률 가중치를 사용하여 선택 비뮴림이 제거된 유사 모집단을 생성하거나 혹은 결측된 연구대상자의 분포를 모형화하거나 추적 관찰기간과 결과 발생을 모형화하여 결측자료 분석 방법론을 사용하여 선택 비뮴림을 교정하는 것이 원칙적으로 가능하다. 그러나, 이러한 방법은 거의 사용되지 않으며 이 질문에 대한 대답은 일반적으로 "아니다"가 될 것이다.</p>	<p>NA / <u>Y</u> / PY / <u>PN</u> / N / NI</p>
<p>비뮴림위험 판단</p>	<p>표 3-6을 보시오.</p>	<p>낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음</p>
<p>선택 사항: 연구대상자 선택에 의해 예상되는 비뮴림의 방향은 무엇인가?</p>	<p>비뮴림의 방향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점(null)으로 향하든지(혹은 멀어지는 쪽으로 향하든지) 혹은 종재 중 어느 한쪽으로 유리하게 작용.</p>	<p>실험 종재 선호 / 비교 종재 선호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음</p>

표 3-6. 연구대상자의 선택에서 비뚤림위험 판단하기

판단	연구대상자 선택의 비뚤림
<p>낮은 비뚤림위험 (잘 수행된 무작위 임상시험과 유사함)</p>	<p>(i) 목표 임상시험에 적절할 것 같은 연구대상자가 연구에 포함되었음 그리고 (ii) 각 연구대상자에 대해 추적 조사의 시작과 중재의 시작이 일치하였음</p>
<p>중등도 비뚤림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과 유사하지 않음)</p>	<p>(i) 연구대상자 선택이 중재와 중재결과에 관련이 있을 수 있음 그리고, 선택 비뚤림에 대해 적절한 방법으로 보정하였음 혹은 (ii) 추적 조사의 시작과 중재의 시작이 모든 연구대상자에서 일치하지는 않음 그리고 (a) 이에 해당하는 연구대상자의 비율이 너무 낮아 중요한 비뚤림을 유발할 수 없음 혹은 (b) 선택 비뚤림에 대해 적절한 방법으로 보정됨 혹은 (c) 문헌고찰 저자가 중재 효과에 대한 발생비(rate ratio)가 시간에 따라 일정하다고 확신함</p>
<p>높은 비뚤림위험 (몇 가지 중요한 문제가 있음.)</p>	<p>(i) 연구대상자 선택이 중재와 중재결과에 관련이 있었지만 매우 강력하 지는 않음 그리고 분석에서 보정될 수 없음 혹은 (ii) 추적 조사의 시작과 중재의 시작이 일치하지 않음 그리고 잠재적으로 중요한 추적관찰 시간이 분석에서 제외됨 그리고 발생비(rate ratio)가 시간에 따라 일정하지 않음</p>
<p>매우 높은 비뚤림위험 (중재 효과에 대한 근거로 사용 하기에는 너무 문제가 많음)</p>	<p>(i) 연구대상자 선택은 중재와 중재결과에 매우 밀접하게 관련이 있었음 그리고 분석에서 보정될 수 없음 혹은 (ii) 상당한 양의 추적 관찰 시간의 양이 분석에서 누락된 것으로 보임 그리고 발생비(rate ratio)가 시간에 따라 일정하지 않음</p>
<p>정보 없음 (이 영역에서 비뚤림위험에 대 한 판단을 하기에)</p>	<p>연구대상자 선택이나, 추적 조사의 시작과 중재의 시작이 일치하는지 여 부에 대한 정보가 보고되지 않음</p>

3.3. 세부 지침: 중재 분류의 비뚤림

3.3.1. 서론

중재 상태가 잘못 분류되면 비뚤림이 발생할 수 있다. 이것은 무작위 임상시험같은 실험 연구에서는 거의 문제가 되지 않는다. 왜냐하면 연구자들이 중재를 적극적으로 배정하고 대부분 정확한 기록이 존재하기 때문이다. 그러나 관찰 연구에서는 배정되거나 제공된 중재에 관한 정보를 수집해야 하는 점이 다른 점이다. 가능한 자료 수집 방법은 다음과 같다.

- 환자에 대한 체계적인 평가(임상 진찰, 인터뷰, 진단 검사)
- 행정 또는 기관 데이터베이스(사전에 정해진 구체화된 목적이 없는 전향적인 자료의 기록)
- 의무 기록에서 추출
- 기관 기록 또는 정책 문서(예 : 조직 중재 또는 공중 보건 중재)

3.3.2. 차등적 그리고 비-차등적 오분류

중재 상태의 오분류는 비차등적 또는 차등적일 수 있다. **비차등적 오분류**는 중재결과와 관련이 없다. 예를 들어 담배 연기 경보 설치가 없는 경우와 담배연기 경보 설치를 비교할 때, 담배연기 경보설치 중재군이 불완전하게 기록되어서 일부 사람들이 “경보 없음”에 잘못 배정될 수 있다. 이러한 오분류가 후속 결과와 관련이 없는 경우(예: 화재로 인한 부상의 위험은 담배 연기 경보 설치를 식별하지 못한 이유와 관련이 없다), 오분류는 비 차등적이며 일반적으로 (중재효과가 없거나 중재간에 차이가 없는) 무효점을 향하는 방향으로 중재의 추정 효과를 편향되게 할 것이다.

차등적 오분류는 중재상태의 오분류가 후속 중재결과 또는 중재결과의 위험과 관련될 때 발생한다. 가능한 경우, 후속 중재결과에 대한 지식 없이 중재를 정의하고 분류하는 것이 중요하다. 차등적 오분류의 잘 알려진 사례는 환자-대조군 연구에서의 **회상 비뚤림**이며, 환자-대조군 상태에 대한 지식은 이전 중재의 회상에 영향을 미친다: 일반적으로 환자군은 대조군에 비해 잠재적으로 중요한 사건을 더욱 잘 회상할 가능성이 있다.

중재상태에 대한 정보가 후향적으로 획득되면, 코호트 연구에서 차등적 오분류가 발생할 수 있다. 중재 상태에 대한 정보(또는 정보의 가용성)가 중재결과에 영향을 받는 경우에 발생할 수 있다. 예를 들어 중재결과가 치매인 노인의 코호트 연구 및 연구 시작시 과거 중재 상태에 대한 연구대상자의 회상은 기준에 존재하는 경도인지장애에 의해 영향을 받는다. 대

안으로, 연구진은 연구대상자가 치매를 가질 때 과거 중재 상태를 더 부지런히 탐색할 수 있다. 중재상태의 차등적 오분류도 가능하다. 예를 들어, 아프리카 지역의 어린이 예방 접종 상태에 대한 정보는 가족이 거주하는 가정을 정기적으로 방문할 때 예방 접종 카드를 확인하여 수집되며, 카드가 없으면 예방 접종을 받지 않은 것으로 간주된다. 일부 문화권에서는 어린이가 사망하면 예방 접종 카드가 소실된다. 이러한 어린이의 예방 접종 상태는 예방 접종이 사망률에 미치는 영향에 대한 연구에서 예방 접종을 받지 않은 것으로 분석되면 차등적으로 오분류될 수 있다. 중재 시점에 중재 상태에 관한 정보가 수집되고, 정보가 완전하며, NRSI를 수행하는 사람들이 정보에 접근할 수 있다면 그러한 문제는 피할 수 있다

3.3.3. 중재 분류의 비틀림에 대한 비틀림위험 평가

신호 질문과 비틀림위험 평가는 box 3-6와 표 3-7에 제시하였다.

표 3-7. 중재 분류의 비틀림에 대한 비틀림위험 판단하기

판단	중재 분류의 비틀림
낮은 비틀림위험 (잘 수행된 무작위 임상시험과 유사함)	(i) 중재 상태가 잘 정의됨 그리고 (ii) 중재의 정의는 중재 시 수집된 정보에만 근거함
중등도 비틀림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과 유사하지 않음)	(i) 중재 상태가 잘 정의됨 그리고 (ii) 중재 상태의 배정 시 일부 양상이 후향적으로 결정됨
높은 비틀림위험 (몇 가지 중요한 문제가 있음)	(i) 중재 상태가 잘 정의되지 않음 혹은 (ii) 중재 상태의 배정 시 주요 측면이 중재결과에 대한 지식에 의해 영향을 받을 수 있는 방식으로 결정됨
매우 높은 비틀림위험 (중재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)	(일반적이지 않음) 중재 상태의 오분류의 양이 극도로 많음. (예: 비정상적으로 강한 회상 비틀림)
정보 없음 (이 영역에서 비틀림위험에 대한 판단을 하기에)	중재에 대한 정의나 중재 상태의 정보원에 대한 설명이 보고되지 않음

Box 3-6. ROBINS-I 도구(2단계, 파트 5): 중재분류의 비밀림위험

신호 질문	세부 설명	응답 선택
3.1 중재 군이 명확하게 정의되어 있는가?	중재의 적절한 비교를 위한 전제 조건은 중재가 잘 정의되어 있어야 한다는 것이다. 모호한 정의는 연구대상자를 분류하는데 비밀림을 초래할 수 있다. 개인 수준의 중재의 경우, 각 중재를 받는 개인을 고려하는 기준, 예를 들어 유형, 세팅, 용량, 빈도, 강도, 중재 시기 등이 명확하고 명시적이어야 한다. 인구집단 수준의 중재(예: 대기오염 방지 대책)의 경우, 그 질문은 인구집단을 명확하게 정의하고 있다면, 대답은 '그렇다'일 가능성이 있다.	Y / PY / PN / N / NI
3.2 중재 시작 시 중재군들을 정의하는 데 사용된 정보가 있는가?	일반적으로 시행된 중재에 관한 정보가 후속 결과에 의하여 영향을 받을 수 없는 자료로부터 이용가능한 것이라면, 중재 상태의 처등적 오분류는 거의 일어나지 않을 것이다. 중재 당시 정보를 수집하면 이러한 오분류를 쉽게 피할 수 있다. 인구집단 수준의 중재(예: 대기오염 방지 대책)의 경우 이 질문에 대한 대답은 '그렇다'일 가능성이 높다.	Y / PY / PN / N / NI
3.3 중재 상태의 분류가 중재 결과의 지식 또는 중재결과에 의해서 영향을 받을 수 있는가?	중재 시점에서 정보 수집은 비밀림을 피하기 충분하지 않을 수 있다. NRS의 목적을 위해 자료를 수집하는 방법도 오분류를 피해야 한다.	Y / PY / PN / N / NI
비밀림위험 판단	표 3-7을 보시오.	낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음
선택 사항: 중재결과에 대한 예상되는 비밀림의 영향은 무엇인가?	비밀림의 영향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점으로 향하던지(혹은 멀어지는 쪽으로 향하던지) 혹은 중재 중 한쪽으로 유리할 수 있다.	실험 중재 신호 / 비교 중재 신호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음

3.4. 세부 지침: 의도한 중재에서 이탈되어 나타나는 비뿔림

3.4.1. 서론

배정된 중재를 이탈하여 실험중재군과 비교군에 시행된 치료 사이에 체계적인 차이가 있을 때 발생하는 비뿔림 영역을 고려해야 한다. 이러한 차이는 치료의 추가적인 양상 혹은 제외되지 않은 의도한 치료의 양상을 반영할 것이다.

다음 사항을 구별하는 것이 중요하다.

(a) 적용된 중재에 대한 지식 때문에, 그리고 연구에서 검증되고 있는 가설과 일관성있게 실험 중재와 비교 중재 사이의 차이점으로 발견되길 기대하기 때문에 나타나는 의도한 중재로부터의 이탈. 이러한 이탈은 통상진료의 일부로 볼 수 없다.

(b) 중재 후 통상적인 치료 중에 발생하는 의도한 중재로부터의 이탈(예: 급성 독성으로 인한 약물 투여 중단)

3.4.2. 관심 효과 특성의 중요성

이러한 고려사항이 비뿔림과 관련되어지는 정도는 관심 효과의 특성에 따라 달라진다.

위의 첫 번째 (a) 유형의 이탈은 항상 중요시해야 한다. 예를 들어, 한 연구는 두 개의 서로 다른 영구 심장 박동 장치가 심장 전문의에 의해 삽입된 후 감염률을 비교하였다. 참여 기관에서 예방 항생제를 투여하는 것은 일상적인 일이 아니었다. 눈가림은 불가능했다. 일부 심장전문자들은 기기 A가 기기 B보다 감염률이 높으며 그 결과 기기 B를 받는 환자보다 기기 A를 받는 환자에게 항생제를 투여하는 경우가 더 많을 것이라고 믿었다. 의도한 중재로부터의 이러한 이탈은 일반적인 임상 치료 상황(유형 b)을 반영하지 않는다. 두 기기 사이의 감염률 차이에 대한 심장병 전문의의 기대(유형 a)를 반영한 것이다. 이 연구의 결과는 관심 효과 중 어느 것이든 비뿔림의 위험이 있다.

(b) 유형의 이탈의 중요성은 관심 효과의 성격에 따라 달라진다. 만약 연구 목표가 중재에 대한 할당(또는 시작) 효과의 비뿔림이 없는 추정이라면, (b) 유형의 이탈에서 의도한 중재로부터의 이탈로 인한 비뿔림은 없을 것이다. 구체적으로, 의도한 중재에서 모든 이탈이 통상적 관행이라면 우리는 여전히 중재의 실제 시행과 관계없이 중재에 대한 할당 효과를 평가할 수 있다.

한편, 목표가 중재의 시작 및 준수에 대한 비뚤림이 없는 추정이라면, 목표 “프로토콜”에서의 모든 이탈은 비뚤림을 초래할 것이다. 예를 들어, 한 공개 임상 시험에서 식도암의 최소 침습적 수술과 개복 수술 후의 호흡기 감염률(respiratory tract infection, RTI)을 비교하였다. 공동 중재의 시행에 있어서 중재 그룹 사이에는 두 가지 중요한 차이가 있었다. 첫째, 개복 수술군에는 일측 기계식 폐환기(RTI를 포함한 호흡기 합병증을 증가시키는 것으로 생각됨)가 사용되었고, 반면 최소 침습적 수술군은 양측 폐환기를 받았다. 둘째, 경막외 진통제는 개복 수술 그룹에서 더 자주 사용되었는데, 그것은 전염병을 가진 환자들은 일반적으로 이동성이 떨어지기 때문에 RTI 발병 위험이 증가한다는 것이다. 의도한 중재로부터의 이러한 이탈은 의도한 중재의 시작 및 준수 효과와 관련하여 연구 결과 비뚤림의 위험에 직면하게 된다.

3.4.3. 의도한 중재로부터 이탈의 유형

의도한 중재로부터의 이탈로 인해 발생하는 비뚤림을 **수행 비뚤림(performance bias)**이라고 부르기도 한다. 이는 무작위 임상시험과 NRSI 모두에서 **중재가 시작된 후** 중재군과 대조군 그룹에 제공된 치료 사이의 체계적인 차이가 발생하고, (분석 목적의) 연구대상자가 의도한 중재 그룹의 일부로 지속될 때 발생한다.

기술적 측면: 수행 비뚤림의 위험성에 대한 고려는 교란 요인과 구별된다. 시간 변동 교란 요인에 대해 보정하는 방법은 관심 중재 사이의 변환(ROBINS-I에서는 교란 요인에서 다름)과 의도한 중재로부터의 이탈(이 비뚤림 영역에서 다름) 모두를 보정하는데 사용할 수 있다. 그러한 방법은 수집된 중재 간 변환의 예측변수 또는 이탈에 대한 충분한 데이터에 의존할 수 있다.

무작위 임상시험에서는 연구대상자와 의료서비스 제공자의 **눈가림**을 수행하여 수행 비뚤림을 줄이거나 피할 수 있다. 눈가림은 일반적으로 NRSI에서는 하지 않는다. 따라서 환자와 의료서비스 제공자 모두 일반적으로 시행중인 중재에 대해 알고 있다.

중재 할당에 대한 지식은 **공동 중재**의 가능성(연구된 중재 이외의 중재를 받을 수 있음, 중재 그룹마다 빈도가 다를 수 있음), **이행 충실도** 저하(즉, 임상시험 기간 동안 의료서비스를 제공하는 의료 전문가가 의도한 중재의 일부 또는 전부를 이행하지 않음), 그리고 환자 혹은 연구대상자의 중재 준수에 영향을 미칠 수 있다. 이행이나 준수에서 실패하는 것은 오염되었거나(연구대상자에게 배정되지 않은 다른 중재를 의도치 않게 적용하는 것), 의도한 중재에서 다른 중재로 변환되는 것 또는 중재 미시행까지를 포함한다.

3.4.4. 중재의 시작 및 준수 효과를 평가할 때 의도한 중재로부터의 이탈

연구대상자의 공동 중재, 중재의 이행 및 준수에 대한 고려는 중재의 시작 및 준수의 효과에 관심이 있는 경우에만 평가되어야 한다.

3.4.4.1. 공동 중재에 대한 고려사항

공동 중재는 잠재적으로 중요한 비뚤림 원천이다. 예를 들어, 두 가지 수술 절차 A와 B 중 하나를 받은 환자의 수술 후 감염 비율을 비교하는 관찰 연구를 생각해 보자. 만약 A 중재를 받은 환자에게는 예방적 항생제를 투여하였고, B 중재를 받은 환자에게는 예방적 항생제를 투여하지 않았다면, A를 받은 환자에서 관찰되는 수술 후 감염률이 낮은 것은 수술 절차보다는 예방적 항생제 때문일 수 있으며, 비뚤림이 있을 가능성이 있다. 특정 대상 실험에서 ‘중재 A+항생제 투여군’과 ‘중재 B +항생제 비투여군’을 비교하는 경우는 그렇지 않다. 이와 유사하게, “실용적(pragmatic)” 목표 임상시험은 의학적으로 명시된 바와 같이 항생제의 기회적 사용(opportunistic use)을 허용할 수 있는 반면, A vs B의 단독 효과를 비교하는 “설명적(explanatory)” 임상시험은 항생제 사용의 균형을 목표로 할 수 있다.

공동 중재는 의도한 중재의 일부가 아닌 새로운 중재로 정의된다. 공동 중재의 존재 여부를 결정하기 전에 의도한 중재에 대한 정상적 또는 일반적인 관행을 고려하는 것이 중요하다. 예를 들어 당뇨병에 대한 약물 치료를 할 경우, 약물용량을 늘릴지 아니면 다른 약물로 바꿀지를 고려하기 위해 모니터링을 할 수 있다. 따라서 이러한 투약 조정은 의도한 중재에서 벗어나는 것이 아니다. 마찬가지로, 당뇨병 관리를 목적으로 하는 다른 치료법의 추가도 의도한 중재의 맥락에서 일반적인 임상적 관행의 일부로 사전 규정될 수 있다.

어떤 경우에는 의도한 중재에 대한 프로토콜에서 다른 치료제의 추가가 의사의 재량에 달려 있다고 명시하지만, 그러한 추가가 중재 그룹마다 다를 수 있다. 예를 들어 아스피린 사용자와 비사용자의 위장 궤양 비율을 비교하는 코호트 연구를 생각해 보자. 아스피린을 복용하는 사람들의 출혈을 막기 위해 양성자 펌프 억제제(PPI)를 사용하는 것이 통상적인 치료과정의 일부다. 이들의 사용이 연구 프로토콜에 사전 명시되어 있는 경우, 아스피린 +(필요에 따라) PPI 사용군과 아스피린 비사용군의 비교가 되며, PPI는 공동 중재로 간주되어서는 안 된다. 그러나 이 연구가 아스피린 사용군과 비사용군을 비교하는 것을 목표로 한다면 PPI는 아스피린 그룹에서 아스피린이 위궤양에 미치는 영향을 과소평가하게 되기 때문에 공동 중재로 간주될 수 있다. 마찬가지로 PPI는 설사를 유발하며, 아스피린 사용자에서 설사의 빈도가 높은 것은 아스피린보다는 PPI 공동 중재 때문일 수 있다.

문헌고찰 저자는 중재 그룹마다 다를 수 있고 연구 결과에 영향을 미칠 수 있는 중요한 공동 중재의 목록을 미리 작성해야 한다(2.1.3절 참조). 그런 다음 각 특정 연구의 맥락에서 관리될 가능성이 높은지 여부를 고려해야 한다.

우리는 문헌고찰 저자들이 중요한 공동 중재가 중재 그룹 간에 균형을 이루는지 여부를 고려할 것을 제안한다. 효과적인 공동 중재가 균형을 이루지 못하면 아마도 수행 비뚤림이 존재할 것이다. 그러나 중재 그룹 간에 공동 중재가 균형을 이루고 있어도, 여전히 중재그룹은 관심비교 중재 외에도 관리나 행동에서 차이가 있을 위험이 있다. 이것은 공동 중재가 다른 방식으로 실험군 및 비교군에 상호작용하는 경우에 해당될 것이다(예를 들어, 그것은 한 중재의 효과를 향상시키지만 다른 중재에는 영향을 미치지 않을 수 있다).

3.4.4.2. 의도한 중재 이행의 충실성에 대한 고려사항

중재가 연구 조사자에 의해 의도한 대로 이행되지 않는 이유로는 (i) 조사자에 의한 프로토콜 준수, (ii) 복잡하거나 시술자의 기술에 의존하는 경우 중재에 대한 기술적 문제 및 (iii) 연구 전후 사정의 차이의 문제를 포함한다. 이행의 충실성과 관련된 문제는 중재 그룹 중 하나 또는 어디에서나 발생할 수 있다. 의도한 치료 프로토콜의 적용과 연구 연구대상자의 중재 준수가 서로 영향을 미칠 수 있기 때문에, 항상 분리될 수는 없다. 예를 들어 어떤 중재는 연구 참가자의 준수를 장려하지 않는 방식으로 시행될 수 있다.

중재 수행의 실패 사례로 우울증 및 약물 남용 이력을 가진 청소년을 대상으로 세 가지 복합 중재를 비교한 연구가 있다: (1) 항우울제 A와 인지행동치료(CBT), (2) CBT와 약물 B, (3) CBT만 제공. CBT에 대한 프로토콜을 의사들에게 보냈지만, CBT의 절차와 관련하여 특별한 교육은 없었고, 이로 인해 CBT 중재의 내용과 적용범위에 있어서 잠재적으로 의사들 간 차이가 생기게 되었다. CBT 세션의 내용에 대한 체크리스트도 없었다. 또한 CBT의 경험도 다르고 전문적인 훈련도 다른 의사들이 각각의 중재 그룹 내에서 CBT를 시행하였다. 이러한 문제들은 의도한 중재 수행의 충실성이 결여되어 있는 연구로 인해 비뚤림의 위험이 있다는 것을 의미한다.

문헌고찰 저자들은 적용 방법, 시기, 장소 및 대상과 관련하여 중재의 세부사항을 고려해야 한다. 의도한 중재에 대한 적용 충실도의 부족과 관련된 비뚤림위험을 평가할 때 고려해야 할 몇 가지 주요 특징에는 다음과 같은 평가가 포함된다:

- a) **시행자**: 중재를 관리하는 사람들의 특성(예: 직원 특성, 전문지식 및 훈련 수준, 잠재적 치료 공조) 및 의도한 프로토콜을 조정하기 위한 중재를 시행하는 사람들의 기회

(예: 의사는 부작용의 잠재성 때문에 복용량을 감소시킬 수 있음)

- b) **중재:** 1차 연구의 상황 내에서 제공될 예정이었던 중재군 또는 비교군의 핵심 요소 (예: 중재의 내용); 여기에는 중재의 복잡성(예: 다수의 구성요소들은 준수성에 부정적인 영향을 미칠 것임), 제공 순서, 복용량, 기간, 형식(예: 대면 회의보다 전화 추적 관찰), 치료의 조작적 정의 등이 포함
- c) **맥락:** 의료 환경(예: 공공의료 외래 환자 대 병원 외래 환자), 서비스 구조(예: 관리 의료 또는 공공 자금 지원 프로그램), 지리적 세팅(예: 농촌과 도시), 그리고 중재가 시행되는 문화적 환경과 법적 환경의 특성

이 평가는 연구 조사자가 의도한 대로 중재군과 비교군이 구현되었는지 여부를 결정하는 것을 목적으로 한다. 이는 각 중재 그룹에 대해 평가되어야 한다. 또한 연구설계 및 분석이 의도하지 않은 중재의 적용의 영향을 최소화하기 위해 시도되는 경우에도 고려해야 한다.

환자 모니터링 방법의 차이는 중재의 충실도에 영향을 미칠 수 있다. 예를 들어, 스피로놀락톤(spironolactone) 사용군과 비사용군의 부작용의 효과를 평가하는 코호트 연구에서, 스피로놀락톤이 혈청 칼륨을 증가시킬 수 있다는 것을 인식하는 의사는 혈청 칼륨이 비정상 수준에 도달하기 전에 후속 용량 조절을 통해 혈청 칼륨의 모니터링을 더 자주 시행할 수 있을 것이다.

3.4.4.3. 중재 준수 고려사항

연구대상자들이 의도한 대로 중재를 준수하지 않는다면 비탈림의 위험이 더 높아질 것이다. 준수 부족은 불완전한 준수, 중재 중단, 비교 중재 교차 및 다른 적극적 중재로 전환하는 것을 포함한다. 미준수 가능성은 평가되는 중재의 특성에 따라 달라질 것이다. 예를 들어, 의약품 중재에 대한 불충분한 준수는 빈번할 수 있으며, 한 개인 내에서 중재 간(또는 의약품 복용과 복용 안 함)의 다중 전환이 발생할 수 있다. 그러나 심장 판막이나 관절 보형물과 같은 외과적 중재의 비교에서 시간-변동 미준수에 대한 가능성은 낮다.

이 도구의 사용자는 비교되는 중재안을 고려해야 한다. 한 가지 중재를 받는 사람들이 의도했던 것보다 더 많이 또는 덜 받고, 중재를 중단하거나 다른 중재로 바꿀 가능성이 있는가? 여러 개의 변환될 가능성이 있는가? 각 그룹 내 전반적인 미준수율을 고려하여 이것이 연구 결과에 영향을 줄 수 있는지 결정하는 것이 중요하다. 결과에 영향을 미칠 수 있는 미준수의 임계값은 중재 유형과 연구설계에 따라 다를 것이다.

3.4.4.4. 기술 참고 사항: 의도한 중재로부터의 이탈에 대한 조정

3.1.9절에서는 시간변동 교란 요인에 대해 조정하는 방법을 사용하여 관심 중재 간의 전환을 허용하는 경우 중재 시작 및 준수의 효과를 추정하기 위한 통계적 접근법을 간략하게 설명하였다. 의도한 중재에서 이탈하는 것을 허용하는 경우에도 관련 방법을 사용할 수 있다. 한 가지 접근법은 편차가 발생할 때 추적 관찰을 중단하는 것이다. 예를 들어 역확률 가중치를 통해 그러한 중단에서 발생할 수 있는 비뿔림(기술적으로 이것은 선택 비뿔림의 일종이다)을 피하는 통계적 방법을 사용할 필요가 있다. 예를 들어 선택적 세로토닌-재흡수 억제제(SSRI) A 사용군 대 중재 미시행군의 시작 및 준수 효과를 조사하는 12개월 연구를 생각해 보자. 추적 관찰 기간 중에 일부 환자가 SSRI B로 전환되면 수행 비뿔림이 발생할 위험이 있다. 이는 SSRI B를 받은 후 추적 관찰을 중단하여 처리할 수 있지만, 분석은 SSRI B로 전환하는 환자가 SSRI A에 남아 있는 환자와 체계적으로 다를 가능성을 고려해야 한다.

그렇지 않으면, 연구대상자가 시작시점 이후 A에서 B로 또는 B에서 A로 변경되지 않는 B에 대한 연속적 중재 대 A로 시작 및 지속적 중재를 받은 효과를 추정하는 것을 목표로 하는 연구를 고려해야 한다. A에 배정된 사람들 중 더 많은 비율이 추적 관찰 기간 동안 동시에 중재 C를 부수적으로 받는 경우, C에 의한 시간 변동 교란 요인이 있다. C에 대한 중재를 예측하는 예후 인자를 시간에 따라 측정한다면, 중재 C의 사용의 불균형으로 인해 야기되는 시간 변동 교란 인자로 인한 비뿔림에 대해 조정하는 방법(3.1.9절 참조)을 사용할 수 있다.

일반적으로 윤리적인 이유로, 추적 관찰 기간 중 연구대상자의 건강 상태에 기초한 중재의 일부 변경을 허용한다. 이런상황은 무작위배정 임상시험에서도 발생한다. 이러한 변경은 연구의 나머지 기간 동안 제공되는 복용량이나 중재 유형이 바뀌는 것일 수 있다. 예를 들어 당뇨병 환자의 혈당을 조절하기 위해 두 가지 약물을 비교하는 연구에서 중재에 대한 반응을 모니터링하기 위해 시간에 따른 혈당 반응(예: HBA1c, 혈압)을 반복적으로 측정한다. 일반적인 치료 과정 중에 이러한 관리가 잘 이뤄지지 않는 경우, 임상 의사는 약물을 변경한다(예를 들어 HBA1c가 너무 높을 경우 약물 복용 용량을 늘린다). 단, 혈당 반응은 주요 결과(예: 심혈관 사망률)에도 영향을 미칠 수 있다. 만약 대상 연구에서 혈당 지수 반응에 따라 복용량을 조정하는 것을 허락하는 프로토콜인 경우, 통계적 보정 방법이 필요하지 않을 것이다. 다만, 초기 약물 복용량에 따른 지속적인 치료 효과를 추정하는 연구에서 혈당 지수에 의한 시간 변동 교란 요인의 문제를 다루기 위해서는 전문적인 통계적 방법(위 참조)이 필요하다.

3.4.5. 의도한 중재로부터 이탈된 비뚤림에 대한 비뚤림위험 평가

신호 질문과 비뚤림위험 평가는 Box 3-7과 표 3-8에 제시되어 있다.

만약 공동 중재가 “균형적” 혹은 중재의 준수 정도가 “높음”으로 판단하는데 사용될 수 있는 기준에 대한 지침이 있다면, 문헌고찰 저자들에게 유용할 것이다. 불행하게도, 우리는 간단한 지침이 일반적으로 적용될 것이라고 믿지 않는다. 즉 공동 중재가 결과에 큰 영향을 미치고, 결과가 발생한 환자가 일반적으로 공동 중재를 받은 환자인 경우, 중요한 공동 중재를 받는 환자의 수가 적더라도 절대적인 차이가 있다면 실질적인 비뚤림을 초래할 것이다.

문헌고찰 팀은 문헌고찰에 포함된 연구 전반에 걸쳐 “균형적” 공동 중재, “성공적” 수행 및 준수 부족에 대한 판단을 일관되게 적용하도록 할 것을 권고한다.

Box 3-7. ROBINS-I 도구(2단계, 파트 6): 의도한 증대에서 이탈되어 나타나는 비밀림위험

신호 질문	세부 설명	응답 선택
본 연구의 목적이 증대에 대한 부정 효과를 평가하는 것이라면 4.1과 4.2에 응답하십시오		
4.1 통상진료에서 예상되는 것 이상으로 의도한 증제로부터 이탈되는 상황이 있었는가?	증제 후 통상진료에서 발생한 이탈(예: 금성 독성으로 인한 약물 투여의 중단)은 의도한 증제의 일부분이므로 증제 부정 효과에 비밀림을 유발하지 않는다. 이탈은 증제군과 비교 증제 간의 차이에 대한 기대로 인해 발생할 수 있다(예를 들어, 연구대상자들이 비교 증제에 배정된 것을 불온하다고 느끼고, 따라서, 적극적 증제 혹은 그러한 요소 혹은 다른 증제를 찾기 때문임). 이러한 이탈은 통상진료의 일부가 아니므로 편향된 효과 추정치를 초래할 수 있다. 그러나, 통상진료에서 개인에 대한 관찰연구에서는 이러한 사항이 예상되지 않는다.	Y / PY / <u>PN</u> / N / NI
4.2 4.101 Y/PY 인 경우: 의도한 증제로부터 이탈된 상황이 증제 군간에 불균형을 초래하고, 증제결과에 영향을 미쳤을 가능성이 있는가?	통상진료를 반영하지 않은 의도한 증제로부터 나타나는 이탈은 증제결과에 영향을 미친다면 중요할 것이나 그렇지 않다면 중요하지 않을 것이다. 또한 두 군에 걸친 이탈의 불균형이 있을 경우에만 비밀림이 발생할 것이다.	NA / Y / PY / <u>PN</u> / N / NI
본 연구의 목적이 증대에 대한 시작 및 준수의 효과를 평가하는 것이라면 4.3~4.6에 응답하십시오.		
4.3 중요한 공동 증제는 증제군간 균형을 이룬다?	계획되지 않은 공동 증제가 증제의 추정 효과를 편향시키는 방식으로 시행될 경우 비밀림의 위험이 더 높아질 것이다. 공동 증제가 증제결과에 영향을 미칠 때는 중요하다. 그렇지 않은 경우에는 중요하지 않다. 비밀림은 증제군 사이에 공동 증제의 불균형이 있을 경우에만 발생할 것이다. 증제결과에 영향을 미칠 가능성이 있고, 이 연구에서 투여되었을 가능성이 있는 사전-명시된 공동 증제를 포함한 공동 증제를 고려해야 한다. 이러한 공동 증제는 증제 군 간에 균형이 맞는지 고려해야 한다.	<u>Y</u> / PY / <u>PN</u> / N / NI
4.4 대부분의 연구대상자에게 증제가 성공적으로 시행되었는가?	예를 들어, 연구 기간동안 보건 의료전문가가 의도한대로 증제가 시행되지 않는다면 비밀림위험이 높을 것이다. 대부분의 연구대상자에게 증제가 성공적으로 수행됐는지 고려해야 한다.	<u>Y</u> / PY / <u>PN</u> / N / NI

<p>4.5 연구대상자는 배정된 중재법을 준수하였는가?</p>	<p>연구대상자들이 의도한대로 중재를 준수하지 않는다면 비몰입위험이 높을 것이다. 미준수에는 불완전한 준수, 중재의 중단, 중재군과 비교 중재의 교차, 다른 적극적인 중재로의 전환 등이 있다. 추적 관찰 기간 동안 배정된 중재를 계속해서 준수한 연구대상자의 비율에 대한 이용가능한 정보를 고려해야 하며, 만약 이 비율이 우려를 일으킬 만큼 충분히 높다면 '아니다', 또는 '아마도 아니다'라고 응답한다. 불완전한 준수가 불가능하도록 단 한번 시행되는 중재 연구에 대해서는 '그렇다'라고 응답한다. 중재 전환된 경우, 중재 전환(중재의 중단 포함) 후 추적 관찰기간 동안은 1)새로운 중재에 해당하는 분석 혹은 2) 최초 중재로 배당하는 분석으로 구분한다. 새로운 중재로 배당하는 분석은 시간-변동 교란효과로 처리하는 것으로 여기서는 더 이상 고려하지 않아야 한다.</p>	<p><u>Y / PY</u> / PN / N / NI</p>
<p>4.6 4.3, 4.4 또는 4.5의 N/PN 인 경우: 중재의 시작 및 준수의 효과를 추정하기 위해 적절한 분석방법을 사용하였는가?</p>	<p>의도한 중재로부터 벗어나지 유형의 이탈을 수정하는 분석을 수행할 수 있다. 적절한 분석 전략의 예로 역확률 기중치나 도구변수 추정 등이 있다. 논문에서 의도한 중재로부터의 이탈에 대한 정보를 보고하지 않고 그러한 분석을 보고하는 것이 가능할 수는 있지만, 정보가 없는 경우 그러한 분석이 적절하다고 판단하기는 어렵다. 이러한 접근법을 사용한 연구를 평가하려면 전문가의 조언이 필요할 수 있다. 만약 한 그룹의 모든 사람이 공동 중재를 받았다면 이를 극복하기 위해 보정을 할 수는 없다.</p>	<p>NA / <u>Y / PY</u> / PN / N / NI</p>
<p>비몰입위험 판단</p>	<p>표 3-8을 보시오.</p>	<p>낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음</p>
<p>선택 사항: 의도한 중재로부터 이탈로 인해 예상되는 비몰입의 방향은 무엇인가?</p>	<p>비몰입의 방향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점으로 향한다면(혹은 멀어지는 쪽으로 향한다면) 혹은 중재 중 어느 한쪽으로 유리하게 작용.</p>	<p>실함 중재 선호 / 비교 중재 선호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음</p>

표 3-8. 의도한 중재에서 이탈되어 나타나는 비뚫림에 대한 비뚫림위험 판단하기

<p>낮은 비뚫림위험 (잘 수행된 무작위 임상시험과 유사함)</p>	<p>중재 배정의 효과: (i) 의도한 중재로부터의 이탈이 일상적인 입장을 반영함 혹은 (ii) 일상적인 입상에서의 이탈이 중재결과에는 영향을 미칠 것 같지 않음</p> <p>중재 시작 및 준수의 효과: 중요한 공동 중재가 중재 군간에 균형을 이루었고, 중재결과에 영향을 줄 수 있는 의도한 중재(적용이나 준수 면에서)로부터 이탈은 없었음</p>
<p>중등도 비뚫림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과는 유사하지 않음)</p>	<p>중재 배정의 효과: 일상적인 입상에서의 이탈이 있으나, 그 중재결과에 미치는 영향이 미미할 것으로 예상됨</p> <p>중재 시작 및 준수의 효과: (i) 의도한 중재로부터의 이탈이 있으나, 그 중재결과에 미치는 영향이 미미할 것으로 예상됨 혹은 (ii) 중요한 공동 중재가 중재 군간에 균형을 이루지 않았거나, 의도한 중재로부터의 이탈이 있어서 중재결과에 영향을 줄 것으로 보임(적용 및 준수 면에서) <i>그리고</i> 중재결과에 대한 영향이 있을 것 같은 이탈을 감안하여, 중재 시작 및 준수의 영향을 추정하는데 분석이 적절하였음(적용, 준수, 그리고 공동 중재 면에서)</p>
<p>높은 비뚫림위험 (몇 가지 중요한 문제가 있음)</p>	<p>중재 배정의 효과: 중재 군간에 불균형이 있었고 중재결과에 영향을 미쳤을 가능성이 있을 만한, 일상적인 입상에서의 이탈이 있었음</p> <p>중재 시작 및 준수의 효과: (i) 중요한 공동 중재가 중재 군간에 균형을 이루지 못하거나, 중재결과에 영향을 미칠 가능성이 있는 의도한 중재로부터의 이탈이 있음(적용 및 준수 면에서) <i>그리고</i> (ii) 중재결과에 대한 영향이 있을 것 같은 이탈을 감안하여, 중재 시작 및 준수의 영향을 추정하는데 분석이 적절하지 않았음(적용, 준수, 그리고 공동 중재 면에서)</p>
<p>매우 높은 비뚫림위험 (중재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)</p>	<p>중재 배정의 효과: 중재 군간에 불균형이 있었고 중재결과에 영향을 미쳤을 것 같을 정도로, 통상진료에서 상당한 이탈이 있었음</p>

	<p>중재 시작 및 준수의 효과:</p> <p>(i) 그룹 전체에서 중요한 공동 중재에서 상당한 불균형이 있거나, 중재결과에 영향을 미칠 가능성이 높은 의도한 중재와 상당한 이탈이 있음(적용 및/혹은 준수 측면에서) 그리고</p> <p>(ii) 중재결과에 대한 영향이 있을 것 같은 이탈을 감안하여, 중재 시작 및 준수의 영향을 추정하는데 분석이 적절하지 않았음(적용, 준수, 그리고 공동 중재 면에서)</p>
<p>정보 없음 (이 영역에서 비뚤림위험에 대한 판단을 하기에)</p>	<p>의도한 중재로부터 이탈이 있었는지에 대해 보고된 정보가 없음</p>

3.5. 세부 지침: 결측치로 인한 비뚤림

3.5.1. 서론

결측 자료는 탈락(추적 관찰 손실), 예약 부도, 불완전한 자료 수집 및 연구책임자에 의해 연구대상자가 분석에서 제외되는 경우를 통해 발생할 수 있다. NRSI에서 기저 특성(시행된 중재 또는 기저 교란변수를 포함), 중재결과 측정, 분석과 관련된 기타 변수 또는 이들의 조합에 대한 데이터가 누락될 수 있다. 결측 자료로 인한 비뚤림을 고려하는 일반적인 규칙은 우리가 수집하고자 하는 자료를 이용한 분석이 비뚤림이 없는 효과 추정치를 생성한다고 가정하여 결측 자료에 의해 발생할 수 있는 비뚤림에만 집중해야 한다는 것이다.

결측 중재결과 자료로 인한 비뚤림위험을 고려하는 출발점은 특히 중재에 대한 배정(또는 시작)과 중재 시작 및 준수(2.2.2절 참조)를 구별하는 것과 관련하여 관심있는 비교대상의 특성을 명확히 하는 것이다. 예를 들어, “완전한” 자료 세트는 선별검사를 제공받았는지 여부에 따라 비교하는 경우와 선별검사를 시행하였는지 여부에 따라 비교하는 경우는 다를 수 있다. 따라서 결측 자료의 정의도 다를 것이다. 따라서 결측 중재결과 데이터를 고려하기 위해서는 처음부터 연구 샘플을 명확하게 정의하는 것이 중요하다. 이것은 목표 무작위 임상 시험의 고려를 통해 달성될 수 있다.

3.5.2. 차등적 누락(Differential missingness)

결측 자료에 대한 구체적인 고려사항은 무작위 임상시험에서 확립된 것과 무작위 임상시험에 대한 기존 Cochrane RoB 도구에 설명된 것을 광범위하게 따른다. 중재 그룹 간의 결측 자료의 차이와 함께 데이터가 누락된 이유가 중요하다. (i) 결측 자료의 비율과 (ii) 결측 자료의 이유가 중재 그룹 간에 유사하다면, 일반적으로 효과 추정치에는 제한된 비뚤림만 있을 것이므로 비뚤림의 위험이 낮거나 중간인 것으로 간주될 것이다(2.4절 참조). 결측 자료의 비율이 증가함에 따라 중재에 대한 반응성의 차이가 비뚤림 가능성에 대한 우려를 증가시킬 수 있다. (i) 통상적으로 보고된 데이터로부터 설정할 수 있지만, (ii) 일반적으로 문헌고찰 저자의 판단이다. 이러한 점을 고려할 때, 중재 그룹 전체의 결측 비율의 균형은 비뚤림의 위험에 대한 중등도 수준의 확신 정도를 제공한다.

3.5.3. 유해효과

의도하지 않은 효과가 나타날 때, 중요한 고려사항은 문헌고찰 저자가 추적 관찰 기간 동안 부작용이 연구대상자 개개인에게 적지 않은 비율로 만연할 수 있는 상황을 체계적으로 배제하지 않았다는 것에 납득하는지 여부이다. 예를 들어, 노년층이 더 많이 탈락(혹은 예약 부도)하고 또한 더 많은 부작용이 있을 경우, 부작용의 많은 부분이 분석에서 누락될 수 있다. 노년층이 다른 사람보다 어떤 중재 그룹에서 탈락할 가능성이 더 높으면 비뚤림이 생기겠지만, 이것은 반드시 비뚤림을 유발하지는 않을 것이다. 예를 들어, 이것은 인지력 감퇴를 방지하기 위해 운동과 크로스워드 퍼즐을 비교할 때 그런 현상이 발생할 수 있다.

3.5.4. 결측 자료로 인한 비뚤림에 대한 비뚤림위험 평가

신호 질문과 비뚤림 평가의 위험은 box 3-8과 표 3-9에 제시되어 있다. 우리는 결측 자료의 범위에 대한 지침을 제공하여 연구결과가 비뚤림위험이 보통이거나 높다는 결론을 이끌어 낼 수 있다면 문헌고찰 저자들이 매우 유용할 것이라는 것을 우리는 알고 있다. 예를 들어, 일부 지침에서 추적 관찰 기간의 80% 미만을 완전성 기준으로 임계값을 사용하여 왔다. 불행하게도, 우리는 하나의 임계값을 의미 있게 정의할 수 있다고 믿지 않는다. 예를 들어, 95%의 완전한 중재결과 데이터에 기초한 연구결과는 중재결과가 드물게 발생하고 중재결과 데이터가 누락된 이유가 중재 그룹과 밀접하게 관련되어 있다면, 여전히 높은 비뚤림위험에 처할 수 있다.

Box 3-8. ROBINS-I 도구(2단계, 파트 7): 결측치로 인한 비틀림위험

신호 질문	세부 설명	응답 선택
5.1 모든 혹은 거의 모든 연구대상자의 종재 결과 자료를 사용할 수 있는가?	"거의 모든"은 "연구 결과를 신뢰할 만큼 충분하다"로 해석되어야 하며, 적절한 비율은 맥락에 따라 달라질 수 있다. 어떤 상황에서는 두 종재 군 간 상당히 공통적으로 관심 사건이 발생한다면, 연구대상자의 95% (혹은 가능한 90%) 자료를 이용할 수 있는 것으로 충분할 수 있다. 이것의 한가지 측면은 문헌고찰 저자가 해당 연구의 정확한 분석 계획을 찾으려고 이상적으로 시도한다는 것이다.	Y / PY / PN / N / NI
5.2 종재 상태에 대한 결측 자료로 인해 연구 대상자가 제외되었는가?	종재 상태가 누락되면 문제가 될 수 있다. 이를 위해서 의도한 연구 표본이 명확해야 하나, 실제로는 그렇지 않을 수 있다.	Y / PY / PN / N / NI
5.3 분석에 필요한 다른 변수에 대한 결측자료로 인해 연구대상자가 제외되었는가?	이 질문은 특히 분석에서 통제된 교란요인에 대한 정보가 누락되었기 때문에 분석으로부터 제외된 연구대상자와 관련이 있다.	Y / PY / PN / N / NI
5.4 5.10I PN/N 혹은 5.2나 5.30I Y/PY/NI 경우: 결측된 연구대상자의 비율과 원인이 중재 기간 유사한가?	이 질문은 (i) 결측된 관찰값의 비율의 차이 혹은 (ii) 결측된 관찰값에 대한 원인의 차이가 다루고 있는 질문에 대한 우리의 답변 능력에 실질적으로 영향을 줄 수 있는 여부를 도출하는 것을 목적으로 한다. "비슷하다"는 우연에 의해 기대되는 것만큼 종재 군 간 약간의 불일치한 것을 의미한다.	NA / Y / PY / PN / N / NI
5.5 5.10I PN/N 혹은 5.2나 5.30I Y/PY/NI 경우: 결측치에 관계없이 연구 결과가 견고(robust)하다는 근거가 있는가?	견고성(robustness)에 대한 근거는 결측치를 다른 분석 방법과 조사자에 의해 수행된 민감도 분석 혹은 때때로 체계적 문헌고찰 수행자에 의해 시행된 추가 분석 여부에서 따라 나타날 수 있다. 분석에 사용된 기법이 명확하고 타당한지를 평가하는 것이 중요하다. 이를 위해서는 연구 내용에 대한 지식과 통계 전문지식이 필요하다. 예를 들어, 다중 대체법(multiple imputation)과 같은 통계 방법의 사용이 적절한 답변을 보장하는 것은 아니다. 문헌고찰 저자는 비교를 위해 나이브한(원료된 사례) 분석을 찾아야 하며, 원료된 사례와 다중 대체법 기반의 결과가 다른 경우, 분석 방법의 타당성에 대해 주의 깊은 평가가 이어져야 한다.	NA / Y / PY / PN / N / NI
비틀림위험 판단	표 3-9를 보시오.	낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음
선택 사항: 결측치에 의해 예상되는 비틀림의 방향은 무엇인가?	비틀림의 방향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점으로 향한다면(혹은 멀어지는 쪽으로 향한다면) 혹은 종재 중 어느 한쪽으로 유리하게 작용.	실험 종재 선호 / 비교 종재 선호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음

표 3-9. 결측치로 인한 비뚫림의 비뚫림위험 판단하기

<p>낮은 비뚫림위험 (잘 수행된 무작위 임상시험과 유사함)</p>	<p>(i) 자료가 합리적으로 완료됨 혹은 (ii) 결측된 연구대상자의 비율이나 이유가 중재군간에 비슷하였음 혹은 (iii) 분석을 통해 결측치를 해결하였고 비뚫림위험을 제거한 것으로 보임</p>
<p>중등도 비뚫림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과는 유사하지 않음)</p>	<p>(i) 결측된 연구대상자의 비율과 이유가 군간에 약간 다름 그리고 (ii) 분석으로 결측치에서 나타나는 비뚫림위험을 제거하였을 것 같지 않음</p>
<p>높은 비뚫림위험 (몇 가지 중요한 문제가 있음)</p>	<p>(i) 결측된 연구대상자의 비율이 중재 간에 상당히 큼 혹은 결측의 이유가 중재 간에 상당히 다름 그리고 (ii) 분석으로 결측치로 인한 비뚫림위험을 제거할 것 같지 않음 혹은 결측치가 분석에서 적절하지 않게 다루어짐 혹은 결측치가 적절한 분석을 통해 비뚫림위험을 제거할 수 없는 특성을 가지고 있음</p>
<p>매우 높은 비뚫림위험 (중재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)</p>	<p>(i) (일반적이지 않음) 결측치가 있는 연구대상자의 중재 사이에 치명적인 차이점이 있음 그리고 (ii) 결측치가 적절한 분석을 통해서 다루어지지 않거나 다룰 수가 없음</p>
<p>정보 없음 (이 영역에서 비뚫림위험에 대한 판단을 하기에)</p>	<p>결측치나 결측되었을 가능성이 있는지에 대해 보고된 정보가 없음</p>

3.6. 세부 지침: 중재결과 측정 비뚤림

3.6.1. 서론

중재결과가 잘못 분류되거나 잘못 측정될 경우 비뚤림이 발생할 수 있다. 가능한 자료 수집 방법은 다음과 같다:

- 환자에 대한 체계적인 평가(임상 진찰, 인터뷰, 진단 검사)
- 행정 혹은 기관 데이터베이스(사전 정의된 목적이 없는 데이터의 전향적 기록)
- 의무 기록에서 추출
- 기관 기록 또는 정책 문서(예: 관리기관 혹은 공공 보건 중재결과 자료)

3.6.2. 차등적, 비차등적 측정 오류

중재결과의 오분류 혹은 측정 오류는 차등적일 수도 비차등적일 수도 있다. 비차등적 측정 오류는 받은 중재와는 관련이 없다. 이러한 것은 체계적일 수 있다(예를 들어, 모든 연구 대상자에서 혈압 측정이 지속적으로 5단위가 높을 경우) - 이러한 경우, 정밀도에 영향을 미치거나 비뚤림을 일으키지 않을 수 있다. 또는 무작위적일 수 있다(예를 들어, 혈압 측정값이 중재 혹은 중재결과에 의존하지 않는 방식으로 너무 높고 때로는 너무 낮을 수 있다) - 이 경우 비뚤림은 일으키지 않고 정밀도에 영향을 미칠 수 있다.

차등적 측정 오류는 중재 상태와 관련된 측정 오류이다. 이것은 중재-중재결과 관계의 비뚤림을 가져올 것이다. 이를 흔히 **중재결과 확인 비뚤림(detection bias)**이라고 한다. 중재결과 확인 비뚤림이 발생할 수 있는 상황의 예는 (i) 중재결과 평가자가 중재 상태를 인지하는 경우(특히 중재결과가 주관적일 때), (ii) 다른 중재 그룹의 중재결과를 평가하기 위해 상이한 방법(혹은 관찰 강도)을 사용하는 경우, (iii) 측정 오류가 중재 상태(혹은 중재-중재결과 관계의 교란 요인)와 관련이 있는 경우이다.

중재결과 평가자의 눈가림은 중재 그룹 간의 측정에서 체계적인 차이를 방지하는 것을 목적으로 한다. 그러나 현실적인 이유로 눈가림이 불가능하거나 수행되지 않는 경우가 많다. 그것은 또한 무작위 임상시험보다 NRSI에서 훨씬 더 빈도가 낮다.

신호 질문에는 데이터 수집 방법의 비교 가능성 및 측정 오류가 중재 상태와 관련되었는지 여부에 대한 고려가 포함된다. 데이터 수집 방법이 매우 잘 표준화되어 있다면 비뚤림의 위험이 더 낮아질 수 있다. 또한 중재 그룹 전반에 걸친 조사의 강도를 고려하는 것도 중요

하다. 예를 들어, 코르티코스테로이드 약물이 제2형 당뇨병의 위험 증가와 관련이 있는지를 평가하는 연구에서, 제2형 당뇨병의 진단은 정기적인 혈액검사에 기초할 수 있다. 만약 코르티코스테로이드를 복용하는 환자들이 코르티코스테로이드의 비사용자들보다 더 빈번하게 혈액검사를 받는다면, 당뇨병이 발견될 가능성이 높아져 비뚤림이 발생하게 된다. 합리적으로 객관적인 중재결과 측정에 대한 표준화된 평가임에도 불구하고 결과 확인 비뚤림의 또 다른 예는 혈압 측정을 위한 표준적인 크기의 커프를 사용하는 것과 관련이 있으며, 이는 비만 환자의 실제 혈압을 과대평가할 수 있다. 만약 중재가 체질량과 관련이 있는 경우, 측정 오차는 비뚤림을 유발할 것이며, 중재결과가 중재 상태에 대해 눈가림으로 측정되더라도 이러한 비뚤림이 존재할 수 있다.

3.6.3. 중재결과 측정에서 비뚤림에 대한 비뚤림위험 평가

신호 질문과 비뚤림위험 평가는 Box 3-4와 Box 3-9, 표 3-10에 제시하였다.

Box 3-9. ROBINS-I 도구(2단계, 파트 8): 중재결과 측정 비틀림위험

신호 질문	세부 설명	응답 선택
6.1 중재결과 측정이 제공받은 중재의 지식에 의해 영향을 받을 수 있는가?	일부 중재결과 측정에는 평가자 판단을 무시할 만한 것들이 있다. 예를 들어 모든 원인 사망률 또는 반복 불가능한 자 동화된 실험실 지표 같은 것들이다. 이러한 중재결과의 측정으로 인한 비틀림위험은 낮을 것으로 예상된다.	Y / PY / <u>PN</u> / N / NI
6.2 중재결과 평가자는 연구대상자들이 받은 중재를 알고 있는가?	만약 중재결과 평가자가 중재 상태에 능가림이 되어 있다고 한다면, 이 질문에 대한 대답은 '아니다'일 것이다. 다른 상황에서, 임상 연구자들에 의해 적극적으로 능가림을 시행하지 않았다 하더라도, 중재결과 평가자들이 연구대상자들이 받은 중재를 인지하지 못할 수도 있을 것이다. 이런 경우 이 질문에 대한 답은 '아니다'가 될 것이다. 예를 들어 설문지와 같이 중재결과를 자가 보고하는 연구에서, 중재결과 평가자는 연구대상자가 된다. 관찰 연구에서 연구대상자가 직접 중재결과를 보고할 때 이 질문에 대한 답은 일반적으로 '그렇다'가 될 것이다.	Y / PY / <u>PN</u> / N / NI
6.3 중재결과 평가 방법은 중재군 간에 유사했는가?	유사한 평가 방법(즉, 자료 수집)에는 동일한 중재결과 추출 방법과 임계치, 동일 시점, 동일 정의 및 동일 측정치 같은 것들이 있다.	<u>Y</u> / PY / PN / N / NI
6.4 제공받은 중재와 관련된 중재결과 측정 시 체계적 오류가 있었는가?	이 질문은 중재결과와 치등적 오분류에 관한 것이다. 중재결과 측정에 체계적 오류가 있는 경우, 중재 혹은 중재-중재 결과 관계의 교란 변수와 관련이 있다면 비틀림이 발생할 수 있다. 이는 일반적으로 중재결과 평가자가 제공받은 중재를 인지하고 있거나 혹은 중재결과 평가 방법의 비교 불가능성(non-comparability)에 기인하지만, 이러한 통제에도 불구하고 발생하는 치등적 오분류의 예가 있다.	Y / PY / <u>PN</u> / N / NI
비틀림위험 판단	표 3-10을 보시오.	낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음
선택 사항: 중재결과 측정에 의해 예상되는 비틀림의 방향은 무엇인가?	비틀림의 방향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점으로 향하든지 (혹은 멀어지는 쪽으로 향하든지) 혹은 중재 중 어느 한쪽으로 유리하게 작용.	실험 중재 선호 / 비교 중재 선호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음

표 3-10. 중재결과 측정 비뚫림의 비뚫림위험 판단하기

<p>낮은 비뚫림위험 (잘 수행된 무작위 임상시험과 유사함)</p>	<p>(i) 중재결과 평가 방법이 군간에 유사함 그리고 (ii) 중재결과 측정이 연구대상자가 받은 중재에 대한 지식에 의해 영향을 받지 않았거나(즉 객관적임), 중재결과 평가자가 연구대상자가 받은 중재를 알지 못했음 그리고 (iii) 중재결과 측정에서의 오차가 중재 상태와 관련이 없음</p>
<p>중등도 비뚫림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과는 유사하지 않음)</p>	<p>(i) 중재결과 평가 방법이 중재 군간에 유사함 그리고 (ii) 중재결과 측정이 연구대상자가 받은 중재에 대한 지식에 의해 영향을 최소로만 받았음 그리고 (iii) 중재결과 측정에서의 오차가 중재 상태와 최소한으로만 관계가 있음</p>
<p>높은 비뚫림위험 (몇 가지 중요한 문제가 있음.)</p>	<p>(i) 중재결과 측정 방법이 중재 군간에 유사하지 않음 혹은 (ii) 중재결과 측정이 주관적임(즉, 연구대상자가 받는 중재에 대한 지식에 의한 영향에 취약함) 그리고 연구대상자가 어떤 중재를 받는지 알고 있는 평가자에 의해 중재결과가 평가됨 혹은 (iii) 중재결과 측정에서의 오차가 중재 상태와 관련이 있음</p>
<p>매우 높은 비뚫림위험 (중재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)</p>	<p>중재결과 평가하는 방법이 중재군 간에 합리적으로 비교할 수 없을 정도로 달랐음</p>
<p>정보 없음 (이 영역에서 비뚫림위험에 대한 판단을 하기에)</p>	<p>중재결과 평가 방법에 대해 보고된 정보가 없음</p>

3.7. 세부 지침: 보고된 연구결과 선택의 비뚤림

3.7.1. 서론

이 장에서 우리는 다음과 같이 정의한다: **중재결과 영역**이란 측정 방법에 관계없이 관심결과 변수의 실제 상태나 종착점(예: 우울증의 여부 혹은 중증도)임, **중재결과 측정**이란 연구대상자에 대한 구체적 측정(예: 치료 6주 후 해밀턴 등급 척도를 사용하여 우울증 측정)임, **중재결과 분석**이란 한 개 이상의 중재결과 측정값을 분석을 통하여 얻은 구체적 연구결과(예: 중재군과 대조군 사이의 기저시점으로부터 6주 후 해밀턴 등급 척도 점수에서 평균 변화의 차이)

3.7.2 비뚤림위험(RoB) 도구에서 중재결과 비보고 비뚤림 제외

임상 시험에서 선택적 보고는 주로 측정되고 분석된 중재결과 영역의 보고 실패 또는 부분 보고로 설명되어 왔다(Kirkham 등, 2010). 중재결과 보고 비뚤림(Outcome reporting bias, ORB)은 중재결과 영역에서 중재군과 관련된 방향, 규모 혹은 통계적 유의성에 기초하여 보고되지 않거나 부분적으로 보고될 때 발생한다. 체계적 문헌고찰에 포함된 하나 이상의 연구에서 이러한 비뚤림의 존재는 체계적 문헌고찰에 의해 보고된 치료 효과 추정치를 편향되게 한다.(대개 효과의 크기를 과장하는 방향으로).

제안된 RoB 도구의 새로운 구조에 따르면 이러한 종류의 선택적 중재결과 보고는 출판 비뚤림과 유사하게 간주한다. 따라서, RoB 도구의 일부가 아닌 다른 메커니즘(예: 연구결과 요약의 GRADE 평가의 일부로서)을 사용하여 이러한 종류의 선택적 중재결과 보고를 평가하는 것이 제안된다. 이것은 현재 무작위 임상시험을 위한 코크란 RoB 도구와 두드러지게 차이가 나는 부분이다. 따라서 우리는 중재결과 영역의 선택적 비보고(또는 불충분한 보고)에 대한 신호 질문을 이 장에 포함시키지 않는다.

이러한 종류의 선택적 중재결과 보고를 고려하기 위해 Kirkham 등(2010)이 제시한 체계를 권고한다.

3.7.3. 합성에 영향을 미치는 연구결과의 선택적 보고

여기서는 **완전하게 보고된 연구결과**, 즉 메타분석(또는 다른 합성 방법)에 추정치를 포함할 수 있도록 충분히 보고된 연구결과의 선택적 보고를 고려할 것이다. 이 영역은 (i) 중재결

과 영역 내에서 평가된 복수의 측정값에서 **특정 중재결과 측정치의 선택적 보고**; (ii) 특정 중재결과 측정의 다중 분석치에서 **특정 분석의 선택적 보고**; 그리고, (iii) **연구대상자의 하위군에 대한 선택적 보고**를 모두 말한다. 이러한 유형의 선택적 보고에서는 ROBINS-I 도구에 고려된 다른 비뚤림 영역과 동일한 방법으로 개별 1차 연구의 효과 추정치를 비뚤림위험이 발생하게 한다. 선택적 보고가 중재 효과 추정치의 방향, 크기 또는 통계적 유의성에 기반하는 경우 비뚤림으로 이어질 것이다.

선택적 중재결과 보고는 중재결과 영역에 대한 다중 중재결과 중 특정 측정치가 선택되었을 때 발생한다. 예로는 다수의 측정도구(예: 통증 척도)를 사용하고, 가장 유리한 결과만 보고한다든지, 다른 항목별 측정이 가능할 때 계측기에 대해 가장 유리한 하위 척도(또는 하위 척도의 하위군(subset)만 보고하거나, 중재결과가 측정된 시점 중 하나 또는 일부만 보고하는 것이다.

선택적 분석 보고는 다수의 방법으로 추정된 중재 효과에서 연구결과를 선택할 때 발생한다. 예를 들어, 기저시점에 맞게 조정된 변화된 점수와 중재 후 점수 모두 분석을 수행하였거나; 변형되었거나 혹은 변형되지 않은 특정 측정값에 대한 다수의 분석을 수행하였거나; 잠재적 교란변수(혹은 잠재적 교란변수의 다른 집합에 대한 보정)에 대한 보정 혹은 보정을 하지 않은 특정 측정값에 대한 다수의 분석을 수행하는 경우; 결측 자료를 고려하기 위한 방법을 사용하였거나 혹은 사용하지 않은 경우와 다른 방법을 사용하여 특정 측정값에 대한 다수의 분석을 수행한 경우; 다른 임계값을 가진 연속형 자료를 범주형 자료의 중재결과로 변환하거나; 하나의 중재결과 영역에 대해 다수의 복합 중재결과가 분석되었으나 복합 결과 중 하나(혹은 하나의 하위군)의 결과만 보고한 경우가 해당된다.(일반적이지 않은 복합 중재 결과에 대한 효과 추정치의 보고는 이와 같은 선택적 보고의 근거일 수 있다.)

대형 코호트 연구로부터 하위군(subgroup) 선택: 분석에 사용된 코호트는 더 흥미로운 발견에 기초하여 데이터를 이용할 수 있는 더 대규모 코호트에서 선택되었을 수 있다. 일반적이지 않은 방법으로 정의된 하위군(예: 복용량이나 복용 빈도에 의한 하위군의 비정상적 분류)은 그러한 선택적 보고의 근거를 제공할 수 있다.

위해와 이익에 대한 효과 추정치의 선택적 보고에 기초하는 동기(및 비뚤림의 방향)는 다를 수 있지만, 선택적 보고는 위해와 이익 모두에 대해 발생할 수 있다. 선택적 보고는 일반적으로 뉴스 가치가 있거나 출판 가치가 있을 정도로 주목할 만한 발견에 대한 욕망에서 발생하며, 이전의 근거(또는 사전 가설)가 뒷받침되거나 모순되는 경우에 그러할 수 있다.

이러한 유형의 선택적 보고는 균집화 여부에 관계없이 모든 코호트 연구설계에 적용된다. 선택적 보고는 탐색적 목적을 가진 연구에서 발생할 가능성이 더 높는데, 이는 그 성격상

그러한 연구는 종종 복수 중재 또는 복수 중재결과 사이의 많은 연관성을 검사하기 때문이다.

그러나 조사된 모든 연관성을 완전히 보고한 탐색적 연구는 선택적 보고의 위험에 처하지 않을 것이다. 즉, 탐색적 성격이 문제가 아니라 선택적 보고가 문제이다.

3.7.4. 선택적 보고의 근거

논문은 선택적 보고의 근거를 너무 많은 방법으로 하여 목록화하기가 어렵다. 비뚤림위험을 낮게 평가하기 위해서는 프로토콜 또는 통계 분석 계획에 명시된 중재결과 측정과 실제 분석이 일치해야 한다.

선택적 보고가 심각한 문제가 아닐 수 있다는 간접적인 근거는, 보고된 중재결과 측정과 분석과 사전 계획 사이의 일관성(일치만큼 강력한 요구사항은 아님) 혹은 논문에서 내부적으로 연구방법(Methods)과 연구결과(Results)에 걸쳐 일치하는 명확하게 정의된 중재결과 측정 및 분석에서 얻을 수 있으며, 외부적으로 그 연구를 보고한 다른 논문과 일치하는 것에서 얻을 수 있다. 비뚤림위험이 중등도 수준이라면, 다수의 분석에서 보고된 분석이 선택되었다는 제시가 없어야 하며, 연구결과에 기초한 분석 및 보고를 위한 코호트 또는 하위군 선택에 대한 표시가 없어야 한다.

중재결과 측정, 분석 또는 분석 코호트의 불일치(예: 적합한 연구대상자의 코호트의 크기와 분석된 코호트 크기 사이의 큰 차이)는 특히 모든 보고된 연구결과가 통계적으로 유의한 경우 선택적 보고의 비뚤림위험이 심각할 수 있음을 나타낼 것이다. 일부 상황에서는 다수의 분석으로부터 선택적 보고의 위험을 증가시키는데, 예를 들어, 이러한 위험을 최소화하기 위한 전략(예: 다중 회귀 모델에 공변량 포함 기준)을 제시하지 않고, 기저시점에서의 예측 변수의 상당한 불균형이 발생하는 것이 해당된다.

선택적 보고에 대한 직접적인 근거 또는 강한 의심(매우 높은 비뚤림위험을 나타냄, 아래 참조)은 때때로 논문의 본문에서 찾을 수 있다. 경보를 발생해야 하는 문장의 예는 다음과 같다: (a) “중재결과 X에 대한 연구결과 [체계적 문헌고찰 중재결과 영역 D와 관련된]는 중재결과 Y [동일한 체계적 문헌고찰 중재결과 영역 D와 관련된]에 비해 더 좋았다”; (b) 연속형 변수를 이분화하고/분류하기 위한 다양한 임계값 기준이 시도되었다; (c) 변화 점수 또한 분석되었지만 그 효과가 통계적으로 유의하지 않았기 때문에 보고하지 않았다. 매우 높은 비뚤림의 판단을 유발하는 특정 문구는 프리 텍스트 상자에 기록해야 한다.

3.7.5 보고된 결과의 선택에서 비뚫림에 대한 비뚫림위험 평가

신호 질문과 비뚫림위험 평가는 Box 3-10와 표 3-11에 제시하였다.

표 3-11. 보고된 연구결과 선택 비뚫림의 비뚫림위험 판정

<p>낮은 비뚫림위험 (잘 수행된 무작위 임상시험과 유사함)</p>	<p>모든 보고된 연구결과가 모든 의도한 증재결과, 분석 및 하위 코호트에 해당한다는 명확한 근거가 있음(보통 사전에 등록된 프로토콜 또는 통계분석 계획을 통해)</p>
<p>중등도 비뚫림위험 (비무작위 연구에 적합하지만 잘 수행된 무작위 임상시험과는 유사하지 않음)</p>	<p>(i) 증재결과 측정과 분석이 사전 계획과 일치함 혹은 명확하게 정의되고 내부적, 외부적으로 일관성이 있음 그리고 (ii) 여러 분석 중에서 보고된 분석을 선택한다는 표시가 없음 그리고 (iii) 연구결과에 기초한 분석 및 보고를 위한 코호트 또는 하위군을 선택한다는 표시가 없음</p>
<p>높은 비뚫림위험 (몇 가지 중요한 문제가 있음.)</p>	<p>(i) 증재결과가 논문의 방법 부분과 결과 부분, 혹은 해당 연구의 다른 논문에서 다르게 정의되어 있음 혹은 (ii) 여러 분석 중에서 선택적으로 보고된 높은 위험이 있음 혹은 (iii) 코호트나 하위군이 분석을 위해 더 큰 연구에서 선택되었으며, 연구결과에 기초해서 보고된 것으로 보임</p>
<p>매우 높은 비뚫림위험 (증재 효과에 대한 근거로 사용하기에는 너무 문제가 많음)</p>	<p>(i) 연구결과를 선택적으로 보고하였다는 근거나 강한 의심이 있음 그리고 (ii) 보고된 연구 결과와는 상당히 다른 보고되지 않은 연구 결과가 있는 것 같음</p>
<p>정보 없음 (이 영역에서 비뚫림위험에 대한 판단을 하기에)</p>	<p>판단을 하기에에는 정보가 너무 없음(예: 해당 연구에 대해서 초록만 이용할 수 있는 경우)</p>

Box 3-10. ROBINS-I 도구(2단계, 파트 9): 보고된 연구결과 선택의 비뿔림위험

신호 질문	세부 설명	응답 선택
보고된 효과 추정치는 다음과 같은 연구 결과에 기초하여 선택되었을 가능성이 있는가?		
7.1 중재결과 영역 내에서 여러 개의 중재결과 측정	특정 중재결과 영역의 경우, 서로 다른 추정치에 대해 다수의 효과 추정치를 생성할 수 있다. 다수의 추정치가 이루어졌지만 하나 또는 일부만 보고된 경우, 연구결과에 따른 선택적 보고의 위험이 있다.	Y / PY / <u>PN</u> / N / NI
7.2 중재-중재결과 관계에서의 다중 분석	효과 분석에 대한 비무작위 연구로부터 사용된 자료의 한계 때문에(교란의 통제 필요, 상당한 결측치 등), 통계 분석자들은 이러한 제한점을 해결하기 위해 다른 분석 방법을 수행할 수 있다. 예를 들어, 비보정 vs 보정 모형; 최종값 vs 기저 상태로부터 변화량 vs 공분산 분석; 변수 변환; 다양한 역할을 가진 연속형 변수를 범주형 변수로 변환; 다양한 공변량 집합에 의한 보정; 결측치를 처리하기 위한 다른 전략 등이 그것이다. 복수의 방법을 적용하면 특정 중재결과 측정에 대한 복수의 효과 추정치가 생성된다. 만약 통계분석자들이 적용할 방법이 사전에 정의하지 않고, 다수의 추정치가 생성됨에도 불구하고 하나 혹은 하위집단 결과만 보고한 경우, 결과에 기초한 선택적 보고의 위험이 있다.	Y / PY / <u>PN</u> / N / NI
7.3 다른 하위군	일상적인 자료원으로부터 종종 사용할 수 있는 큰 코호트의 경우 여러 하위군의 경우 여러 추정치를 생성하거나 원래 코호트의 다양한 비율을 단순히 생략할 수 있다. 다수의 추정치가 생성되지만 하나 또는 일부만 보고되는 경우 연구결과에 따른 선택적 보고의 위험이 있다.	Y / PY / <u>PN</u> / N / NI
비뿔림위험 판단	표 3-11을 보시오.	낮음 / 중등도 / 높음 / 매우 높음 / 정보 없음
선택 사항: 보고된 연구결과와 선택에 의해 예상되는 비뿔림의 방향은 무엇인가?	비뿔림의 방향을 예측하는 것이 가능하다면, 이를 기술하는 것이 도움이 될 것이다. 그 방향이 무효점으로 향한다면(축은 멀어지는 쪽으로 향한다면) 혹은 중재 중 어느 한쪽으로 유리하게 작용.	실현 중재 선호 / 비교 중재 선호 / 무효점으로 향함 / 무효점에서 멀어짐 / 예측할 수 없음

References

1. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2016 (Published online 29 September).
2. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med*. 2002;21(11):1513-24.
3. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-12.
4. Savović J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med*. 2012;157(6):426-38.
5. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377(14):1391-8.
6. Insight Start Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med*. 2015;373(9):795-807.
7. Lodi S, Sharma S, Lundgren JD, Phillips AN, Cole SR, Logan R, et al. The per-protocol effect of immediate versus deferred antiretroviral therapy initiation. *AIDS*. 2016;30(17):2659-63.
8. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325(7365):652-4.
9. Menerit CL. *Clinical Trials – Design, Conduct, and Analysis*. Second Edition: Oxford University Press; 2012.
10. Piantadosi S. *Clinical Trials: A Methodologic perspective*. Second Edition: Wiley; 2005.
11. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319(7211):670-4.
12. Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials*. 2007;4(4):350-6.
13. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs: a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118.
14. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.
15. Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: what is the right question? *Clin Trials*. 2014;11(1):28-37.
16. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA*. 2014;312(6):606-15.

17. Savovic J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane Reviews: the ROBES meta-epidemiologic study. *Am J Epidemiol*. 2017.
18. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-12.
19. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-6.
20. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med*. 2002;136(3):254-9.
21. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359(9305):515-9.
22. Schulz KF, Grimes DA. *The Lancet Handbook of Essential Concepts in Clinical Research*. Edinburgh, UK: Elsevier; 2006.
23. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet*. 2002;359(9310):966-70.
24. Berger VW, Ivanova A, Knoll MD. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat Med*. 2003;22(19):3017-28.
25. Berger VW. Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biometrical journal Biometrische Zeitschrift*. 2005;47(2):119-27; discussion 28-39.
26. Brown S, Thorpe H, Hawkins K, Brown J. Minimization--reducing predictability for multi-centre trials whilst retaining balance within centre. *Stat Med*. 2005;24(24):3715-27.
27. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? *BMJ*. 2016;355:i5663.
28. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274(18):1456-8.
29. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-6.
30. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *The Cochrane database of systematic reviews*. 2012;11:MR000030.
31. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol*. 2007;36(4):847-57.
32. Corbett MS, Higgins JPT, Woolacott NF. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. *Research synthesis methods*. 2014;5(1):79-85.
33. Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hrobjartsson A, Savovic J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological

- Studies. PLoS One. 2016;11(7):e0159267.
34. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002;287(22):2973-82.
 35. Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. *PLoS One*. 2014;9(2):e88008.
 36. Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. *J Clin Epidemiol*. 2013;66(11):1271-80.
 37. Chapin RB, Williams DC, Adair RF. Diabetes control improved when inner-city patients received graphic feedback about glycosylated hemoglobin levels. *J Gen Intern Med*. 2003;18(2):120-4.
 38. Saugstad OD, Rootwelt T, Aalen O. Resuscitation of asphyxiated newborn infants with room air or oxygen: an international controlled trial: the Resair 2 study. *Pediatrics*. 1998;102(1):e1.
 39. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614-8.
 40. Bellomo R, Chapman M, Finfer S, Hickling K, Myburgh J. Low-dose dopamine in patients with early renal dysfunction: a placebo-controlled randomised trial. Australian and New Zealand Intensive Care Society (ANZICS) Clinical Trials Group. *Lancet*. 2000;356(9248):2139-43.
 41. Smilde TJ, van Wissen S, Wollersheim H, Trip MD, Kastelein JJ, Stalenhoef AF. Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *Lancet*. 2001;357(9256):577-81.
 42. de Gaetano G. Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice. Collaborative Group of the Primary Prevention Project. *Lancet*. 2001;357(9250):89-95.
 43. Brightling CE, Monteiro W, Ward R, Parker D, Morgan MD, Wardlaw AJ, et al. Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet*. 2000;356(9240):1480-5.
 44. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274(18):1456-8.
 45. Altman DG, Bland JM. How to randomize. *BMJ*. 1999;319:703-4.
 46. Shrier I, Platt RW, Steele RJ. Re: "Variable selection for propensity score models". *Am J Epidemiol*. 2007;166(2):238-9.
 47. Senn S. Seven myths of randomisation in clinical trials. *Stat Med*. 2013;32(9):1439-50.
 48. Fu R, Vandermeer BW, Shamliyan TA, O'Neil ME, Yazdi F, Fox SH, et al. AHRQ Methods for Effective Health Care: Handling Continuous Outcomes in Quantitative Synthesis. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville (MD): Agency for

- Healthcare Research and Quality (US): 2008.
49. Wright IS, Marple CD, Beck DF. Report of the Committee for the Evaluation of Anticoagulants in the Treatment of Coronary Thrombosis with Myocardial Infarction; a progress report on the statistical analysis of the first 800 cases studied by this committee. *Am Heart J.* 1948;36(6):801-15.
 50. Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *J Clin Epidemiol.* 2007;60(12):1229-33.
 51. Carlisle JB. Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia.* 2017;72(8):944-52.
 52. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ.* 2016;355:i4919.
 53. Hernan MA, Scharfstein D. Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat. *Ann Intern Med.* 2018;168(7):515-6.
 54. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology.* (in press).
 55. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* 2005;365(9453):82-93.
 56. Worsley SD, Oude Rengerink K, Irving E, Lejeune S, Mol K, Collier S, et al. Series: Pragmatic trials and real world evidence: Paper 2. Setting, sites, and investigator selection. *J Clin Epidemiol.* 2017;88:14-20.
 57. Madsen MV, Gotzsche PC, Hrobjartsson A. Acupuncture treatment for pain: systematic review of randomised clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups. *BMJ.* 2009;338:a3115.
 58. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med.* 2006;3(10):e425.
 59. Jensen JS, Bielefeldt AO, Hrobjartsson A. Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. *J Clin Epidemiol.* 2017;87:35-46.
 60. Bello S, Moustgaard H, Hrobjartsson A. Unreported formal assessment of unblinding occurred in 4 of 10 randomized clinical trials, unreported loss of blinding in 1 of 10 trials. *J Clin Epidemiol.* 2017;81:42-50.
 61. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ.* 2004;328(7437):432.
 62. Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Contemp Clin*

- Trials. 2005;26(1):25-37.
63. Hrobjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol.* 2007;36(3):654-63.
 64. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol.* 2007;36(3):664-5.
 65. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA.* 2001;285(15):2000-3.
 66. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol.* 2005;58(12):1220-6.
 67. Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials.* 2006;3(4):360-5.
 68. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol.* 2002;55(8):787-90.
 69. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Bmj.* 2010;340:c332.
 70. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA.* 2006;296(20):2441-50.
 71. De Bruyne B, Fearon WF, Pijls NH, Barbato E, Tonino P, Piroth Z, et al. Fractional flow reserve-guided PCI for stable coronary artery disease. *N Engl J Med.* 2014;371(13):1208-17.
 72. Biere SS, van Berge Henegouwen MI, Maas KW, Bonavina L, Rosman C, Garcia JR, et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multicentre, open-label, randomised controlled trial. *Lancet.* 2012;379(9829):1887-92.
 73. St Peter SD, Adibe OO, Juang D, Sharp SW, Garey CL, Laituri CA, et al. Single incision versus standard 3-port laparoscopic appendectomy: a prospective randomized trial. *Ann Surg.* 2011;254(4):586-90.
 74. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med.* 2012;157(6):429-38.
 75. Hrobjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *Int J Epidemiol.* 2014;43(4):1272-83.
 76. Nuesch E, Reichenbach S, Trelle S, Rutjes AW, Liewald K, Sterchi R, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum.* 2009;61(12):1633-41.
 77. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in

- new drug applications. *BMJ*. 2003;326(7400):1171-3.
78. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol*. 2007;60(7):663-9.
79. Abraha I, Cherubini A, Cozzolino F, De Florio R, Luchetta ML, Rimland JM, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. *BMJ*. 2015;350:h2445.
80. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol*. 2005;34(1):79-87.
81. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*. 2010;340:c2697.
82. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med*. 2015;12(7):e1001855.
83. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010.
84. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319(7211):670-4.
85. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019.
86. Higgins JPT, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials*. 2008;5(3):225-39.
87. Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med*. 2011;154(2):113-7.
88. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Second Edition. Hoboken, NJ: Wiley; 2002.
89. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367(14):1355-60.
90. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-95.
91. Robins J, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. *Journal of the American Statistical Association*. 1995;90:106-21.
92. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation*. 1981;64(4):669-73.
93. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987.
94. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.

95. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res.* 2016;962280216683570.
96. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials.* 2012;9(4):396-407.
97. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *Pharmacoeconomics.* 2018.
98. Akl EA, Kahale LA, Agoritsas T, Brignardello-Petersen R, Busse JW, Carrasco-Labra A, et al. Handling trial participants with missing outcome data when conducting a meta-analysis: a systematic survey of proposed approaches. *Syst Rev.* 2015;4:98.
99. Akl EA, Johnston BC, Alonso-Coello P, Neumann I, Ebrahim S, Briel M, et al. Addressing dichotomous data for participants excluded from trial analysis: a guide for systematic reviewers. *PLoS One.* 2013;8(2):e57132.
100. Ebrahim S, Akl EA, Mustafa RA, Sun X, Walter SD, Heels-Ansdell D, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol.* 2013;66(9):1014-21 e1.
101. Ebrahim S, Johnston BC, Akl EA, Mustafa RA, Sun X, Walter SD, et al. Addressing continuous data measured with different instruments for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol.* 2014;67(5):560-70.
102. Mavridis D, White IR, Higgins JP, Cipriani A, Salanti G. Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis. *Stat Med.* 2015;34(5):721-41.
103. Akl EA, Kahale LA, Ebrahim S, Alonso-Coello P, Schunemann HJ, Guyatt GH. Three challenges described for identifying participants with missing data in trials reports, and potential solutions suggested to systematic reviewers. *J Clin Epidemiol.* 2016;76:147-54.
104. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials.* 2004;1(4):368-76.
105. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med.* 2001;135(11):982-9.
106. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med.* 2007;26(14):2745-58.
107. Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Burgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ.* 2009;339:b3244.
108. Rothman KJ, Greenland S, Lash TL. Validity in epidemiological studies. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology.* 3rd ed. Philadelphia (PA):

- Lippincott Williams & Wilkins: 2008. p. 128-47.
109. Clemens JD, Chuong JJ, Feinstein AR. The BCG controversy. A methodological and statistical reappraisal. *JAMA*. 1983;249(17):2362-9.
 110. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PE, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis*. 2014;58(4):470-80.
 111. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1-2):51-63.
 112. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and quality of life outcomes. 2006;4:79.
 113. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*. 2012;344:e1119.
 114. Engebretsen K, Grotle M, Bautz-Holter E, Ekeberg OM, Juel NG, Brox JI. Supervised exercises compared with radial extracorporeal shock-wave therapy for subacromial shoulder pain: 1-year results of a single-blind randomized controlled trial. *Phys Ther*. 2011;91(1):37-47.
 115. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365.
 116. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Systematic reviews*. 2016;5(1):108.
 117. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy*. 2012;17(1):1-2.
 118. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med*. 2017;14(11):e1002447.
 119. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev*. 2016;5:11.
 120. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument selection using the OMERACT Filter 2.1: the OMERACT methodology. *J Rheumatol*. 2019.
 121. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JPT, Sterne JAC, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med*. 2014;11(6):e1001666.
 122. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine* 2006; 119: 166. e7-166.e16.
 123. Cook TD, Campbell DT, Peracchio L. Quasi experimentation. In: Dunnette MD, Hough LM (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 491-567). Palo Alto, CA, US: Consulting Psychologists Press, 1990.

124. Davey Smith G, Phillips AN, Neaton JD. Smoking as “independent” risk factor for suicide: illustration of an artifact from observational epidemiology? *Lancet* 1992; 340: 709-12.
125. Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Annals of Internal Medicine* 1983;99:544-50.
126. Hernán MA. With great data comes great responsibility. Publishing comparative effectiveness research in Epidemiology [editorial]. *Epidemiology* 2011; 22: 290-291.
127. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-25.
128. Hernán MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;155(2):176-84.
129. Higgins JPT, Altman DG, Gøtzsche PC, Jüi P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JAC, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
130. Institute of Medicine. Ethical and Scientific Issues in Studying the Safety of Approved Drugs. Washington, DC: The National Academies Press, 2012.
131. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *International Journal of Epidemiology* 2006; 35: 337-344. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010; 340: c365. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010; 21: 383-8.
132. Magid DJ, Shetterly SM, Margolis KL, Tavel HM, O’Connor PJ, Selby JV, Ho PM. Comparative effectiveness of angiotensin-converting enzyme inhibitors versus beta-blocker as second-line therapy for hypertension. *Circulation: Cardiovascular Quality and Outcomes* 2010; 3: 453-458.
133. Majeed AW, Troy G, Nicholl JP, et al. Randomized, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet* 1996; 347: 989-94.
134. McMahon AJ, Russell IT, Baxter JN, et al. Laparoscopic versus minilaparotomy cholecystectomy: a randomised trial. *Lancet* 1994; 343: 135-8.
135. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *American Journal of Epidemiology* 2003;158(9):915-20.
136. Schünemann HJ, Tugwell P, Reeves BC, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Research Synthesis Methods* 2013; 4: 49-62.
137. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. New York: Oxford University Press, 2003.

138. Suissa S, Spitzer WO, Rainville B, et al, Recurrent use of newer oral contraceptives and the risk of venous thromboembolism. *Human Reproduction* 2000;15(4):817-21.
139. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008;167(4):492-9.
140. Rothman KJ and Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins 1998.
141. Velie EM, Shaw GM. Impact of prenatal diagnosis and elective termination on prevalence and risk estimates of neural tube defects in California, 1989-1991. *American Journal of Epidemiology* 1996;144(5):473-9.
142. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011; 155: 529-36.

부록: 용어 해설

이 섹션에서는 ROBIS에서 용어들이 어떻게 사용되었는지에 대해 정의와 함께 ROBIS에서 사용된 몇 가지 용어에 대한 개괄을 제시한다.

용어	정의
Bias (비뚤림)	연구 결과를 왜곡하는 문헌고찰 (또는 일차 연구)의 설계, 수행 또는 분석의 체계적인 결함 또는 제한점
Data collection (자료 수집)	일차 연구 보고서(report) 및 기타 출처로부터 자료를 수집하는 과정. 때로는 자료 추출이라고도 한다. 추가적인 출처로부터 정보를 수집하는 것(예: 저자와 접촉하는 것) 뿐만 아니라 일차 연구 보고서에서 자료를 추출하는 것을 포함할 수 있다.
Databases (데이터베이스)	연구 보고서를 색인하는 서지 데이터베이스
Electronic sources (전자 출처)	체계적 문헌고찰에서 포함하기 위한 연구의 보고서를 찾기 위해 사용된 온라인 출처(서지 데이터베이스 제외)
Eligibility criteria (적격 기준)	체계적 문헌고찰에서 포함하거나 제외해야 하는지를 결정하는데 사용되는 기준. 또한 알려진 포함/제외 기준
Guideline (진료지침)	의사 및 환자가 특정 임상 상황에 적합한 의료 서비스에 대한 결정을 내릴 수 있도록 돕기 위해 체계적으로 작성된 진술서. 이용할 수 있는 가장 좋은 근거에 기초해야 한다. 이는 진료지침 개발자들이 일반적으로 가이드라인(guidance)에 기존의 체계적 문헌고찰을 포함시키고 평가해야 한다는 것을 의미한다.
Heterogeneity (이질성)	연구 간 변동성. 이질성은 인구집단, 중재, 비교중재, 방법론적 질 또는 연구설계의 측면에서 연구들 간의 차이를 설명하는 일반적인 의미로 사용될 수 있다. 통계적 이질성은 표본 변이(우연)로 인해 연구들 사이의 효과 추정치가 예상한 것보다 더 많이 다를 때 발생한다.
Meta-analysis (메타 분석)	여러 연구 결과를 결합하기 위한 통계적 방법. 이는 여러 연구에서 얻은 효과 추정치의 가중 평균을 계산하여 이루어진다. 메타 분석은 종종 체계적 문헌고찰의 일부로 수행되지만 항상 가능하거나 적절한 것은 아니다.
Overview (Overview)	체계적 문헌고찰의 문헌고찰

용어	정의
Overview authors (Overview 저자)	Overview의 저자
Readers (독자)	체계적 문헌고찰을 읽는 사람
Relevance (타당성, 관련성)	관심 있는 질문과 체계적 문헌고찰이 일치하는 범위
Reports (보고서)	연구 방법, 연구 결과 및 결론의 요약. 발표되거나 발표되지 않을 수 있다. 단일 연구는 여러 개의 다른 보고서에 보고될 수 있으며, 일부 단일 보고서는 여러 개의 연구를 보고할 수 있다.
Review (문헌고찰)	체계적 문헌고찰의 짧은 용어. 비 체계적 문헌고찰을 의미하는데 사용될 수도 있지만, 이는 ROBIS의 맥락에서의 의미는 아니다.
Review authors (문헌고찰 저자)	체계적 문헌고찰의 저자. 때로는 단순히 'reviewer'라고 언급하기도 한다.
Search strategy (검색 전략)	서지 데이터베이스를 검색하는데 사용되는 정확한 용어와 그 조합
Statistical significance (통계적 유의성)	어떤 문장(귀무 가설)이 참이라는 점을 감안할 때 관측된 결과나 더 극단적인 결과가 나올 확률은 우연에 의해 발생한다. 통계적 가설 검정은 연구 결과가 귀무가설의 기각으로 이어질 수 있는지 여부를 결정하는데 사용된다. 5%로 귀무가설의 기각에 대한 임계값을 설정하는 경우가 많지만, 다른 값을 선택할 수도 있다. 따라서 p값이 0.05(또는 정해진 수준) 미만인 것으로 확인되면 그 결과는 통계적으로 유의한 것으로 간주되고 귀무 가설은 기각된다.
Studies (연구)	체계적 문헌고찰에 포함된 일차 연구. 단일 연구가 여러 보고서에 포함될 수 있다.
Summary estimate (요약 추정치)	메타 분석을 사용하여 여러 연구에서 도출된 효과의 단일 추정치
Synthesis (합성)	체계적 문헌고찰에 포함된 주요 연구 결과를 요약하는데 사용된, 양적 및 질적 분석
Systematic review (체계적 문헌고찰)	연구의 근거를 검토하고 요약하기 위한 체계적인 접근법. 특정 연구 질문을 다루는 모든 이용 가능한 근거를 찾고, 평가 및 요약하기 위해 정의된 구조를 따른다. 메타 분석을 포함할 수 있지만 체계적 문헌고찰의 요구 사항은 아니다.
Target question (목표질문)	평가자들이 overview나 진료지침에서 답하려고 하는 질문

