Part II 체계적 문헌고찰과 메타분석

AI 산업보건 석사논문

4주차· 5주차

이 Part 에서 답할 질문

- 체계적 문헌고찰은 보통의 종설과 무엇이 다른가
- 메타분석은 여러 연구의 결과를 **어떻게** 하나의 숫자로 합치는가, 그 숫자는 무엇을 뜻하는가
- 무엇이 가장 자주 틀리는가 — 이질성의 해석, 그리고 AI 가 추출한 표

0. 문제로 시작합니다

한 사업장의 보건관리자가 묻습니다. "소음 작업자에게 고혈압이 더 많다는데, 논문마다 숫자가 다릅니다. 어떤 것은 OR 1.2, 어떤 것은 2.5 입니다. 그래서 소음이 고혈압을 얼마나 올린다는 겁니까? 청력보호구 예산을 혈압 관리에도 써야 합니까?"

논문 스무 편을 읽고 "대체로 높인다"고 답하면 그것은 인상입니다. 근거는 **어느 논문을 왜 골랐는지, 각 논문에서 무엇을 어떻게 뽑았는지, 그것을 어떤 규칙으로 합쳤는지**를 남이 따라 할 수 있게 적을 때 생깁니다. 이 Part 는 그 절차를 배우고, 실제로 출판된 메타분석 한 편을 **재현**해 봅니다.

1. 무엇인가 — 1.1 종설과 체계적 문헌고찰

	서술적 종설 (narrative review)	체계적 문헌고찰 (systematic review)
질문	넓고 느슨함	**PECO 로 미리 정한 하나의 질문**
문헌 선택	저자가 아는 것, 고른 것	검색식·포함 기준을 **미리 정하고 전부**
재현 가능성	없음	다른 사람이 같은 절차로 같은 결과
편향 평가	대개 없음	연구마다 비뚤림 위험을 판정
결론	저자의 판단	규칙에 따른 종합 (+ 메타분석)

1. 무엇인가 — 1.1 종설과 체계적 문헌고찰

"체계적"이란

절차를 미리 정해 두었다는 뜻입니다

1. 무엇인가 — 1.2 메타분석

메타분석은 체계적 문헌고찰의 **통계 단계**입니다. 각 연구의 효과 추정치(OR, RR, 평균 차이)를 정밀도로 가중해 하나의 요약 추정치와 신뢰구간을 냅니다. 문헌고찰 없이 메타분석만 할 수도 있지만 그러면 어느 연구를 왜 넣었는지 알 수 없고, 문헌고찰을 하고도 연구가 너무 다르면 메타분석을 하지 않을 수 있습니다.

1. 무엇인가 — 1.3 언제 쓰는가

- 같은 노출–결과 연관을 본 연구가 여럿이고 결과가 갈릴 때 → 종합
- 개별 연구가 작아 효과를 못 잡을 때 → 검정력
- 정책·기준을 만들 때 → 노출기준의 근거는 대개 메타분석과 용량–반응 분석입니다
- **학위논문으로**: 자료 수집이 없어 한 학기에 가능. 대신 검색·선별·추출의 노동이 큽니다

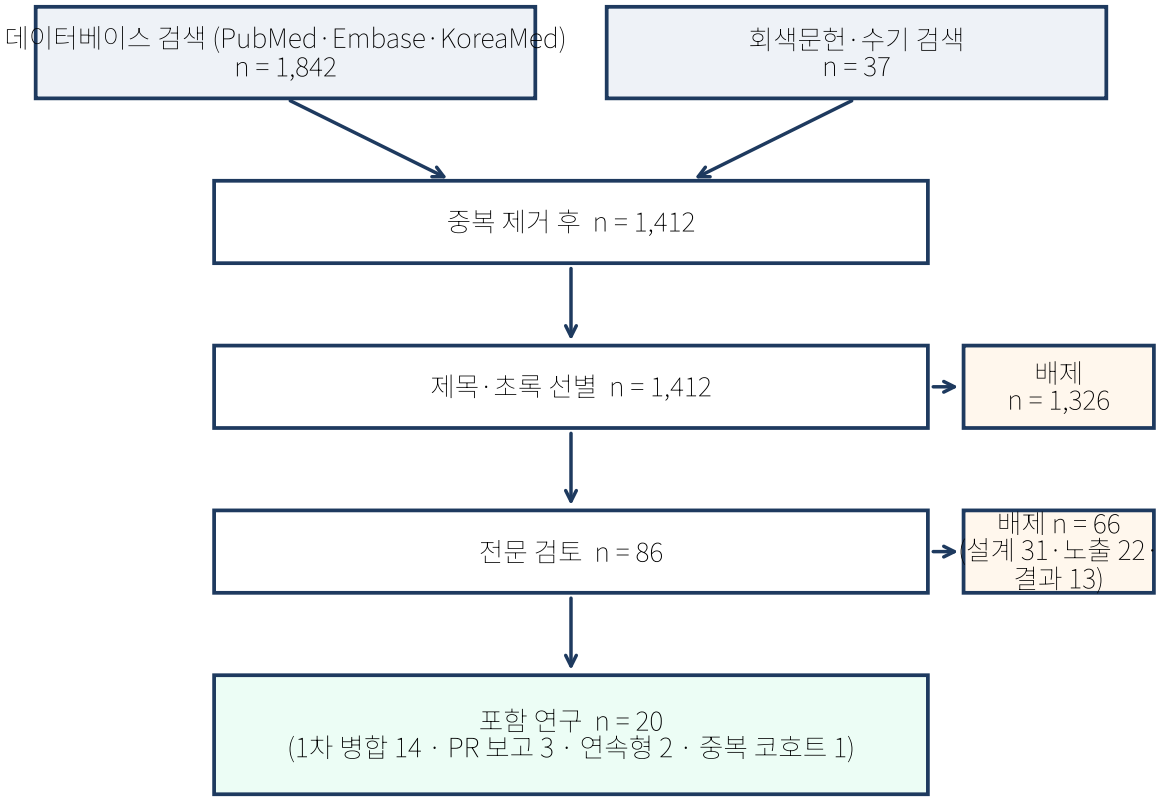
1. 무엇인가 — 1.4 산업보건에서의 예

직업적 소음과 고혈압 · 교대근무와 대사증후군 · 석면과 폐암(용량–반응) · 작업 관련 근골격계 위험요인 · 중재(참여형 인간공학 프로그램)의 효과. 대부분 **관찰연구의 메타분석**이며, 이것은 무작위 시험의 메타분석과 도구(비뚤림 평가)와 해석이 다릅니다.

2. 기본 구조

요소	내용
질문	**PECO** — 대상(P) · 노출(E) · 비교(C) · 결과(O). 예: 근로자(P)에서 8h TWA ≥85 dB(E)는 <80 dB(C) 대비 고혈압(O) 오즈를 높이는가
분석 단위	**연구** 한 편 = 자료 한 행. 변수는 효과 추정치와 그 분산(신뢰구간에서 역산)
설계	포함 연구의 설계(단면·코호트·환자-대조군)가 섞이면 하위군으로 나눔
산출물	PRISMA 흐름도 · 포함 연구 특성 표 · 비뚤림 평가 표 · **forest plot** · funnel plot · 요약 추정치와 예측구간
보고 지침	PRISMA 2020 (27항목). 프로토콜은 PROSPERO 에 사전 등록

2. 기본 구조



선별 단계: AI 보조(민감도 우선) → 배제군 20% 사람 재검토 → 위음성률 보고

3. 하는 법 단계별

단계	하는 일	핵심 규칙
① 프로토콜	PECO · 포함/제외 기준 · DB · 병합 방법 · 하위군을 **미리** 적고 등록	결과를 본 뒤 바꾸면 사후 가설화
② 검색	PubMed·Embase 등 2개 이상 + 회색문헌(정부·기관 보고서). MeSH + 자유어	산업보건은 회색문헌에 **유의하지 않은 결과**가 많다
③ 선별	제목·초록 → 전문. **2인 독립**, 불일치는 합의	잘못 배제한 논문은 영원히 안 보인다 → 애매하면 포함
④ 추출	설계·n·노출 정의·참조군·지표·조/보정 추정치·CI·보정변수 를 **양식**에	2인 독립, 원문 위치 기록
⑤ 비뚤림 평가	노출 관찰연구는 **ROBINS-E** 영역별 판정	총점을 합산하지 않는다
⑥ 병합	log(OR) 로 변환 → 분산의 역수로 가중 → 요약	직업역학은 **변량효과**가 기본

3. 하는 법 단계별

단계	하는 일	핵심 규칙
⑦ 이질성	Q · I^2 · τ^2 · **예측구간**	I^2 는 크기가 아니라 비율
⑧ 설명	미리 정한 하위군 · 메타회귀(k ≥ 10) · 민감도	사후 하위군은 다중검정
⑨ 출판편향	funnel plot · Egger/Peters (k ≥ 10)	비대칭 = 출판편향이라 단정하지 않음
⑩ 근거 수준	GRADE (관찰연구용 변형)	
⑪ 보고	PRISMA 2020 체크리스트 · 검색식 전문 · 추출 자료 공개	

3. 하는 법 단계별 — 3.1 병합의 원리 — 최소한의 수식

각 연구 i 의 log OR 을 y_i, 그 표준오차를 s_i 라 하면 (신뢰구간에서 $s_i = (\ln U_i - \ln L_i)/3.92$):

- **고정효과** — 모든 연구가 하나의 참값을 본다고 가정. 가중치 $w_i = 1/s_i^2$, 요약 $\hat{\theta} = \sum w_i y_i / \sum w_i$
- **변량효과** — 참값이 연구마다 분포한다고 가정. 연구 간 분산 τ^2 를 추정하고 $w_i^* = 1/(s_i^2 + \tau^2)$ 로 가중

직업역학에서 업종·농도·국가가 다른 연구들이 같은 참값을 볼 일은 거의 없으므로 **변량효과가 기본**입니다. 결과는 로그 척도에서 병합하고 지수를 취해 OR 로 되돌립니다.

3. 하는 법 단계별 — 3.2 이질성 읽기

- $Q = \sum w_i (y_i - \hat{\theta})^2$ — 연구들이 요약값에서 얼마나 흩어졌는가
- $I^2 = (Q - df)/Q$ — 전체 변동 중 연구 간 차이의 **비율**. 연구가 정밀해지면 τ 가 그대로여도 올라갑니다
- **τ** — 연구 간 표준편차, 효과크기와 같은 단위. "연구 간에 OR 이 ×1.06 배 정도 흔들린다"
- **예측구간** — 다음 연구(다음 사업장)에서 나올 값의 범위. 정책 판단에는 신뢰구간보다 정직합니다

3. 하는 법 단계별 — 3.2 이질성 읽기

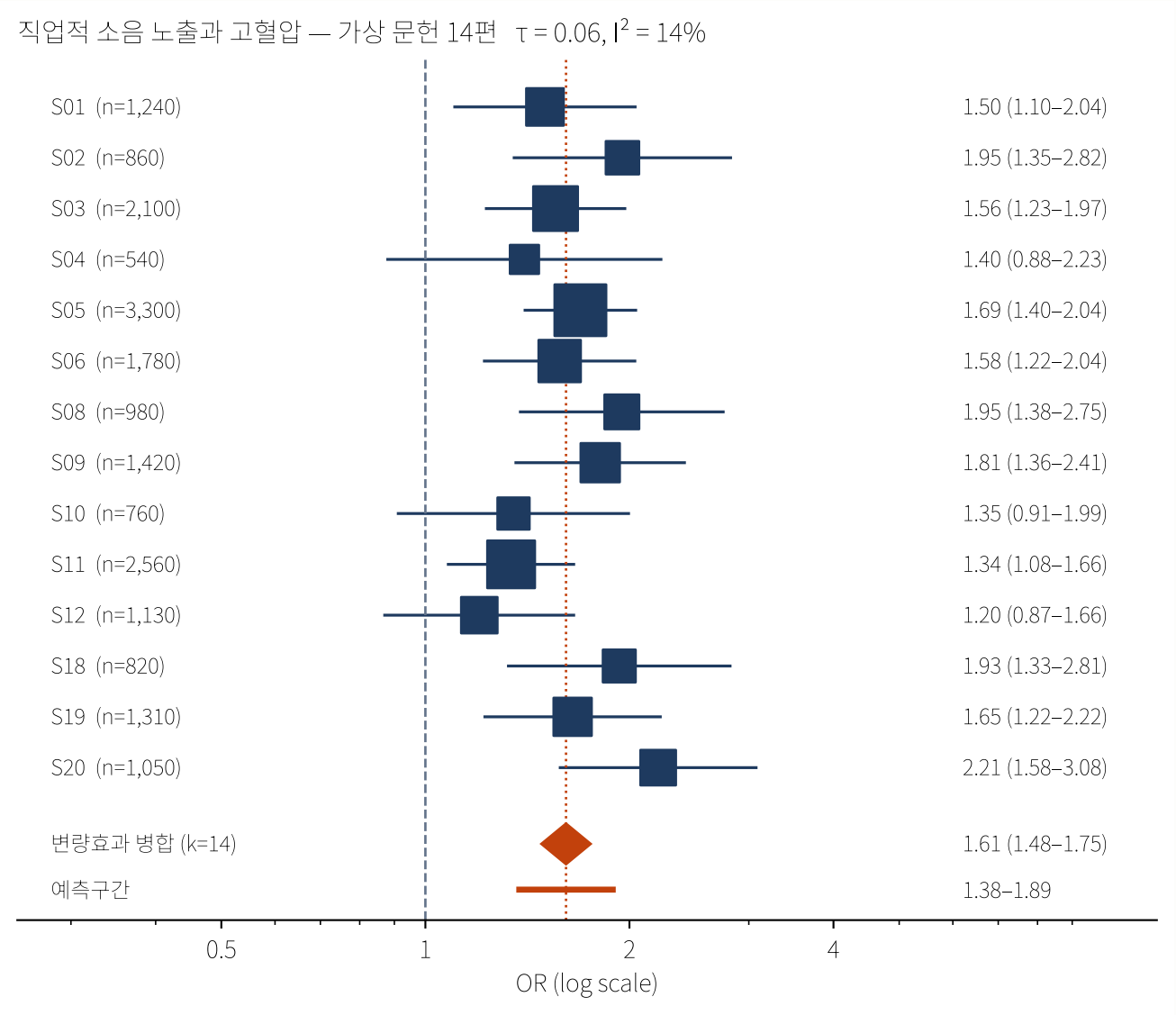
I^2 는 이질성의 크기가 아니라 비율입니다

τ 와 예측구간을 보고합니다

4. 시범 — 끝까지 한 번

가상 문헌 20편(직업적 소음 ≥85 dB 와 고혈압, `합성자료/studies/`)으로 ①~⑦ 을 돌린 결과입니다. 20편 중 이분형 OR 을 보고한 14편이 1차 병합에 들어갑니다. 유병비(PR)를 보고한 3편은 따로 병합하고, 10 dB 당 연속형 2편은 용량–반응용으로 두고, 같은 코호트를 두 번 보고한 1편은 뺐습니다.

4. 시범 — 끝까지 한 번


직업적 소음 노출과 고혈압 — 가상 문헌 14편 τ = 0.06, I² = 14%
S01 (n=1,240)
1.50 (1.10–2.04)
S02 (n=860)
1.95 (1.35–2.82)
S03 (n=2,100)
1.56 (1.23–1.97)
S04 (n=540)
1.40 (0.88–2.23)
S05 (n=3,300)
1.69 (1.40–2.04)
S06 (n=1,780)
1.58 (1.22–2.04)
S08 (n=980)
1.95 (1.38–2.75)
S09 (n=1,420)
1.81 (1.36–2.41)
S10 (n=760)
1.35 (0.91–1.99)
S11 (n=2,560)
1.34 (1.08–1.66)
S12 (n=1,130)
1.20 (0.87–1.66)
S18 (n=820)
1.93 (1.33–2.81)
S19 (n=1,310)
1.65 (1.22–2.22)
S20 (n=1,050)
2.21 (1.58–3.08)
변량효과 병합 (k=14)
1.61 (1.48–1.75)
예측구간
1.38–1.89
0.5
1
2
4
OR (log scale)

4. 시범 — 끝까지 한 번

읽는 법	이 그림에서
상자 위치 = 연구의 OR, 가로선 = 95% CI	대부분 1 의 오른쪽. 두 편은 CI 가 1 을 걸침
상자 크기 = 가중치 (n 이 크면 큼)	S05 (n=3,300) 이 가장 큼
마름모 = 요약 추정치와 CI	**OR 1.61 (1.48–1.75)**
τ · I^2	τ 0.06, I^2 14% — 연구 간 차이가 작다
예측구간	1.38–1.89. 다음 사업장에서도 1 을 넘을 것으로 예상

방법 절에는 이렇게 적습니다 — "이분형 노출을 보고한 14편을 변량효과(DerSimonian–Laird) 모형으로 병합하였다. 요약 OR 은 1.61 (95% CI 1.48–1.75), τ 0.06, I^2 14%, 95% 예측구간 1.38–1.89 였다." 결과 문장은 표와 그림에서 그대로 옮깁니다.

5. 주의점과 흔한 오류

오류	무엇이 잘못인가	대신
"I^2 = 84% 로 이질성이 커서 해석에 주의"	I^2 는 비율. 정밀한 연구가 많으면 올라감	τ 와 예측구간으로 서술, 하위군으로 **설명**
이질성이 크면 병합을 포기	이질성은 정보	미리 정한 하위군·메타회귀
고정효과를 주 결과로	연구 간 차이를 무시 → CI 과소	변량효과
조 OR 과 보정 OR 이 병기된 표에서 조 OR 을 씀	교란 미보정 값이 섞임	보정 추정치, 무엇을 보정했는지 기록
참조군이 노출군인 논문을 그대로	방향 반전	역수
PR·RR 을 OR 로	지표가 다름	별도 병합 또는 변환 근거

5. 주의점과 흔한 오류

오류	무엇이 잘못인가	대신
10 dB 당 값을 이분형과 섞음	노출 형태가 다름	용량–반응 분석
같은 코호트 두 편을 둘 다	이중 계수	최근·최장 추적 하나
그림만 있는 논문의 값을 "읽음"	근거 없음	저자 문의, 없으면 제외 표기
질 점수를 합산해 순위	다른 종류의 문제를 상쇄	영역별 판정 그대로
학술지만 검색	출판편향 그대로	회색문헌

5. 주의점과 흔한 오류 — AI 가 추출하면 오류가 한 방향으로 납니다

사람이 20편에서 값을 옮기면 실수는 흩어집니다. AI 는 조·보정 병기 표에서 **일관되게 조 OR 을**, 참조 반전 논문에서 **일관되게 역수를 안** 취하고, 그림만 있는 논문에서 **숫자를 지어냅니다**. 가상 20편의 AI 추출본(`합성자료/ai_extraction_v1.csv`)이 그렇게 만들어져 있습니다.

	사람 손 추출 14편	AI 추출본 20편 그대로
요약 OR	1.61 (1.48–1.75)	1.44 (1.28–1.62)
$\tau \cdot I^2$	0.06 · 14%	**0.24 · 84%**
예측구간	1.38–1.89	**0.86–2.41**

무작위로 4편만 검수하면 조 OR 오류(5편)를 만날 확률이 72%, 그림 창작(1편)은 20% 입니다. 한 건을 만났으면 **그 보고 형식의 나머지 편을 전부** 다시 봅니다 — 표적 검수.

5. 주의점과 흔한 오류 — AI 가 추출하면 오류가 한 방향으로 납니다

추출 오류는 요약값보다

이질성에 먼저 나타납니다

6. AI 와 함께 — 위임 가능 과업과 검증

> 자율성 수준 L0(쓰지 않음) · L1(보조 — 사람이 전부 검토) · L2(초안 — 전면 재검토) · L3(위임 — 표본 검수) · L4(자동 — 테스트 통과 시 채택). 정의는 Part 0 §2.2, 약어는 `00_공통/용어집.md`.

과업 (GAIDeT 문헌고찰 영역)	수준	무엇을 맡기나	틀리면 어떻게 알아채나
검색식 초안	L2	PECO → 블록별 MeSH·자유어	MeSH Browser 전수 실존 확인 · 핵심 논문 5편 포착 검사
제목·초록 선별	L3	한 건씩 JSON, "애매하면 포함"	배제군 20% 사람 재검토 → 위음성률
자료 추출	L3	원문 인용 병기, 판단 대신 **플래그**(조/보정·참조 반전·연속형·비OR·그림만·중복)	2인 손 추출과 대조 κ → 불일치 형식 전수 재검토
논리 검사 코드	L4	CI 가 점추정치를 포함하는가, 라벨	코드가 잡는 것은 형식 오류뿐임을 안다
비뚤림 평가 초안	L1–L2	영역별 근거 문장 찾기	판정은 2인
			참값을 심은 가상 10편으로 복원 · 이종

6. AI 와 함께 — 위임 가능 과업과 검증

과업 (GAIDeT 문헌고찰 영역)	수준	무엇을 맡기나	틀리면 어떻게 알아채나
이질성·결과 서술 초안	L2	τ·PI 기반 문장	"I^2 크다" 로 끝나는 문장 삭제
해석·GRADE	L0	—	사람

프롬프트 원문은 `AI실습/01~05`. 순서가 전부입니다 — **사람 2인 손 추출 → 합의 → 그 뒤에 AI 추출본**. 반대로 하면 사람 추출이 AI 에 끌려갑니다.

7. 실습 — 출판된 메타분석의 재현

원 논문은 **Bolm-Audorff U, et al. Occupational Noise and Hypertension Risk: A Systematic Review and Meta-Analysis. IJERPH 2020;17:6281** (오픈액세스)입니다. 포함 연구 24편, 노출 >80 dB(A) vs ≤80, 변량효과, 정의별 하위군, 용량–반응(10 dB·년당), 정정(2026)까지 있어 재현 실습에 맞습니다. 원 논문 PDF·포함 연구 목록·확보된 PDF 는 `원논문/`, 나머지 유료 논문은 교수가 수업에서 배포합니다. 출판된 forest plot(그림 2·3)이 정답입니다.

회차	하는 일	주는 것 (발판)
4주	원 논문 해부(정답 카드 봉인) → 병합 계획 커밋 → 검증 짝 2인 독립 손 추출 → 합의	추출 양식 · 계획서 양식
5주	AI 추출(플래그 스키마) → 대조·κ → 표적 검수 → 논리 검사 → 가상 10편 사전 검증 → 병합 → **원 논문과 대조**	병합 코드 예 (`합성자료/verify.R`)
제출	P2 결과 보고 II — 추출표·forest·τ·PI·원 논문과의 차이와 원인 분류	루브릭 `30_실습.md`

8. 되돌아보기

- 오늘 잡은 오류 한 가지 — AI 가 어느 형식에서 어떻게 틀렸는가
- 원 논문과 내 재현이 다른 곳 — 내 추출인가, 저자의 선택인가, 원 논문의 오류인가
- 다음 Part 로 가져갈 것: **"한 건을 만났으면 그 형식 전체를 다시 본다"** 는 습관은 설문 자료 정제에서도 같습니다

참고문헌

- Page et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. doi:10.1136/bmj.n71
- DerSimonian et al. (1986). Meta-analysis in clinical trials. doi:10.1016/0197-2456(86)90046-2
- Higgins et al. (2002). Quantifying heterogeneity in a meta-analysis. doi:10.1002/sim.1186
- IntHout et al. (2016). Plea for routinely presenting prediction intervals in meta-analysis. doi:10.1136/bmjopen-2015-010247
- Higgins et al. (2024). A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). doi:10.1016/j.envint.2024.108602

(실존 확인: `reference/refs_verified.csv`)

가져갈 다섯 줄

1. 체계적 문헌고찰은 **절차를 미리 정한** 고찰이고, 메타분석은 그 통계 단계다
2. 효과는 log 척도에서 정밀도로 가중해 합친다. 직업역학은 **변량효과**가 기본이다
3. **I^2 는 비율이다.** τ 와 예측구간을 보고하고, 이질성은 설명한다
4. AI 추출 오류는 **일관되며 이질성에 먼저 나타난다.** 무작위 검수에 표적 검수를 더한다
5. 출판된 forest plot 을 재현해 보면 추출·병합의 어디가 흔들리는지 보인다