보건학연구방법론

사전학습 교재 · 플립러닝과 소크라테스식 AI 튜터를 위한 원고

윤진하

연세대학교 의과대학 예방의학교실 · 대학원 보건학과

최종 업데이트 2026-09-07 · 전 11장

차례

머리말 — 이 책을 어떻게 쓰나

왜 이 책인가

역학이 다루는 것은 **인구 집단에서 실제로 일어나는 현상**입니다. 누가 더 많이 앓고, 무엇이 그 차이를 만드는가. 그런데 이 현상은 눈으로 직접 볼 수 없습니다. 우리는 언제나 **연구라는 절차를 통해서만** 그것을 봅니다. 사람을 뽑고, 재고, 분석해 하나의 결론으로 옮깁니다. 논문은 그 현상을 담으려는 **도구**이지, 현상 그 자체가 아닙니다.

문제는 그 절차가 **현실을 그대로 옮기지 못할 때가 있다**는 것입니다. 표본을 잘못 뽑으면(선택 편향), 잴 것을 제대로 재지 못하면, 교란을 걸러내지 못하면, 추적 도중 사람들이 특정한 방향으로 빠져나가면 — 겉으로는 과학의 형식을 완벽히 갖춘 논문이라도 그 숫자는 인구 집단의 진짜 모습과 어긋납니다. **절차만 과학인 것과, 현실을 제대로 담은 과학은 다릅니다.**

그래서 이 책의 목표는 개념을 외우는 것이 아니라, **어떤 연구가 인구 집단의 실제 현상을 제대로 담고 있는지 판단하는 것**입니다. 논문을 의심하는 것이 목적이 아닙니다. 논문 너머의 **사람들에게 실제로 무슨 일이 일어나고 있는가**를, 그 절차가 얼마나 믿을 만하게 보여주는지로 되짚는 것입니다. 지도를 읽되, 그 지도가 땅을 제대로 그렸는지를 함께 묻는 일입니다.

여기에 오늘의 사정이 하나 더 겹칩니다. 생성형 AI에게 물으면 몇 초 만에 유창한 답이 돌아옵니다. 그런데 **유창한 것과 맞는 것은 다릅니다.** 매끄럽게 쓰인 연구 계획, 그럴듯하게 요약된 결론이 실은 편향에 오염되어 있을 때, 그것을 알아보지 못하면 잘못된 정책과 잘못된 임상 판단으로 이어집니다. 발생률과 유병률의 정의를 아는 것과, 그 값이 인구 집단의 현실을 제대로 잰 것인지 판단하는 것은 다른 능력입니다. 앞은 들으면 알지만, 뒤는 스스로 부딪혀야 길러집니다.

이 책의 자세

이 책은 **정답을 떠먹여 주지 않습니다.** 그래서 본문에 답지가 없습니다. 각 장은 개념을 설명한 뒤, 끝에서 여러분에게 두 가지를 돌려줍니다.

- **확인 문제** — 읽었는지 스스로 점검하는 객관식입니다. 답을 싣지 않았습니다. 풀어 보고, 막히면 그 개념으로 돌아가 다시 읽으십시오.
- **스스로 생각해 보기** — 그 장에서 가장 자주 헷갈리는 지점을 파고드는 질문입니다. 정답을 찾는 것이 아니라, 개념을 **자기 말로** 세워 보는 자리입니다.

이 질문들에는 온라인 짝이 있습니다. **yonsei QST(온라인 강의 플랫폼)의 AI 튜터**가 같은 질문을 놓고 여러분과 대화합니다. 튜터는 답을 말해 주지 않습니다. 여러분의 답에 꼬리 질문을 던져, 여러분의 이해가 어디까지 닿았는지를 스스로 발견하게 합니다. **책으로 생각하고, 튜터와 확인하십시오.**

AI를 어떻게 쓸 것인가

이 책과 이 수업은 AI를 금지하지 않습니다. 대신 **어디까지 써도 되는지**를 분명히 합니다.

- **써도 되는 곳** — 브레인스토밍, 자료 탐색, 모르는 용어 찾기, 그리고 이 책의 짝인 AI 튜터
- **직접 해야 하는 곳** — 논문의 약점을 판단하는 일, 시험, 자신의 생각을 적는 성찰
- **밝힐 것** — 어떤 AI를 어디에 썼는지 적습니다. **쓴 것은 문제가 아닙니다. 숨긴 것이 문제입니다.**

AI가 쓴 글은 유창합니다. 이 책은 그 유창함 아래에서 무엇이 맞고 무엇이 틀렸는지를 **알아보는 훈련**입니다. 그 눈이 이 과목이 여러분에게 남기려는 것입니다.

이 책이 서 있는 자리

이 책은 하나의 교재를 요약한 것이 아니라, 여러 원저 연구를 바탕으로 **개념을 다시 세워 쓴 것**입니다. 각 장은 실제 보건학·역학 연구를 사례로 삼아, 그 연구가 무엇을 잘했고 무엇을 조심해야 하는지를 따라갑니다. 인용한 원논문은 그 자리에 밝혀 두었습니다.

이제 첫 장을 펼치십시오. 학기가 끝날 때, 처음 보는 연구를 놓고 **그 결론이 인구 집단의 실제 모습을 제대로 담았는지 — 어디에서, 왜 어긋났는지**를 말할 수 있게 되는 것. 그것이 이 책이 여러분과 함께 가려는 곳입니다.

제1장 · 과학철학과 역학 개론

제1장 과학철학과 역학 개론

> **이 장은 나머지 열 장의 어휘를 정하는 자리입니다.** 여기서 흔들리면 뒤가 계속 흔들립니다.

이 장에서 답할 질문

- 역학은 무엇을 하는 학문이고, 어떤 사고방식 위에 서 있는가
- 좋은 연구 질문은 어떻게 만들어지는가 — 그리고 왜 질문을 잘 짓는 것이 연구의 절반인가
- **발생률과 유병률은 무엇이 다른가** — 이 구분이 5장·6장을 가릅니다
- 예방은 어느 단계에서 이루어지는가

이 장에는 계산이 거의 없습니다. 대신 **말의 정의**가 있습니다. 뒤의 열 장은 전부 이 장의 단어로 쓰여 있으므로, 여기서 "발생률"과 "유병률"을 헷갈린 채 넘어가면 5장에서 그 값이 손에 잡히지 않습니다.

이 책 전체를 관통하는 한 가지

시작하기 전에 이 과목의 뼈대를 한 문장으로 세워 둡니다.

> **우리가 알고 싶은 것은 인구 집단의 실제 현상이고,**
> **연구는 그것을 담으려는 절차일 뿐이다**

역학이 재려는 것은 **사람들로 이루어진 집단에서 실제로 일어나는 일**입니다. 얼마나 많은 사람이 새로 병에 걸리는가, 무엇이 그 차이를 만드는가. 그런데 이 현상은 직접 볼 수 없습니다. 우리는 **표본을 뽑고 · 재고 · 분석하는 절차**를 거쳐서만 그것을 짐작합니다. 논문에 실린 숫자는 그 절차의 결과물이지, 인구 집단의 현실 그 자체가 아닙니다.

그래서 앞으로 어떤 연구를 만나든 물음은 늘 두 층입니다.

1. 이 연구가 **말하는 것**은 무엇인가 (결과·결론)

2. 그 절차가 **인구 집단의 실제 현상을 제대로 담았는가** — 아니면 표집·측정·추적·분석의 어디에서 어긋났는가

두 번째 물음이 이 책의 진짜 주제입니다. 표본을 잘못 뽑거나(선택 편향), 잴 것을 제대로 재지 못하거나, 교란을 걸러내지 못하면, **과학의 형식을 완벽히 갖춘 연구라도 그 숫자는 현실과 어긋납니다.** 이어지는 각 장은 결국 이 한 물음의 서로 다른 얼굴입니다 — 코호트의 추적 탈락(5장), 환자-대조군의 대조군 선정(6장), 표준화가 걷어내는 연령의 왜곡(3장)이 모두 "절차가 현실을 얼마나 믿을 만하게 담는가"를 묻습니다. 이 장의 어휘부터 그 눈으로 읽으십시오.

1.1 역학은 어떤 학문인가

이름이 정의를 담고 있다

역학(epidemiology)은 그리스어 세 낱말에서 왔습니다.

낱말	뜻
epi	~에 대하여, ~사이에서
demos	사람들, 인구
logos	학문, 연구

합치면 **"인구에서 일어나는 것을 연구하는 학문"** 입니다. 현대적으로 다시 쓰면 **"인구집단에서 질병 빈도의 분포(distribution)와 결정요인(determinants)을 연구하는 학문"** 입니다. 이 한 문장에 이 과목 전체가 들어 있습니다.

- **분포** — 누가, 어디서, 언제 더 많이 앓는가. 이것이 2장 기술역학입니다
- **결정요인** — 무엇이 그 차이를 만드는가. 이것이 5장 이후의 분석역학입니다

역학의 단위는 언제나 **개인이 아니라 인구집단**입니다. 임상의는 눈앞의 환자 한 명을 봅니다. 역학자는 "이 집단에서 이 병이 저 집단보다 많은가, 그렇다면 왜인가"를 봅니다. 시선의 단위가 다릅니다.

과학은 어떻게 사고하는가 — 귀납과 연역

현대 역학은 **실증주의(positivism)** 위에 서 있습니다. 관찰 가능한 실재가 있고, 그것을 재고 검증할 수 있다는 관점입니다. 이 관점 안에서 연구는 두 방향으로 움직입니다.

	출발점	하는 일	예
귀납 (induction)	관찰	사례를 모아 일반 규칙을 세운다	여러 지역에서 흡연자에게 폐암이 많다는 관찰들을 모아 "흡연이 폐암을 늘린다"는 가설을 세운다
연역 (deduction)	가설	가설에서 예측을 끌어내 관찰로 검증한다	"흡연이 폐암을 늘린다면, 금연자는 시간이 지나며 위험이 준다"를 예측하고 확인한다

실제 연구는 둘을 오갑니다. 존 스노(John Snow)의 콜레라 연구가 교과서적인 예입니다. 그는 런던의 콜레라 사망이 특정 급수원 주변에 몰려 있다는 것을 **관찰**했고(귀납), 거기서 "물이 콜레라를 옮긴다"는 가설을 세웠습니다. 그리고 **펌프 손잡이를 떼면 발병이 줄 것**이라는 **예측**을 끌어내 확인했습니다(연역). 세균이 발견되기 30년 전의 일입니다.

반증 가능성 — 가설의 자격

> **검증할 수 없는 진술은 과학적 가설이 아닙니다.**

가설의 조건은 참처럼 들리는가가 아니라 **자료로 틀렸음을 보일 수 있는가**입니다. 이것을 반증 가능성(falsifiability)이라 합니다.

- "흡연은 건강에 나쁘다" — 무엇을, 누구에게, 무엇과 비교해 재야 할지 정해져 있지 않습니다. 어떤 자료로도 기각할 수 없으니 가설이 아닙니다
- "하루 20개비 이상 흡연자는 비흡연자보다 10년 내 폐암 발생률이 높다" — 두 집단을 정하고 잴 것을 정했습니다. 자료가 이를 지지하거나 기각할 수 있으니 가설입니다

> **연구 질문을 좁히는 일이 연구의 절반입니다**

이 반증 가능성이 1.2로 이어집니다. 좋은 질문은 곧 **검증 가능한 질문**이기 때문입니다.

통계는 이 안에서 무엇을 하는가

통계는 별도의 과목이 아니라 역학의 도구입니다. 두 가지 일을 합니다. 하나는 **기술(description)** — 자료를 요약해 분포를 보이는 것(2장·3장). 다른 하나는 **추론(inference)** — 표본에서 본 것을 모집단으로 일반화하며 우연의 몫을 가늠하는 것(11장). 이 장에서는 도구를 꺼내지 않고, 그 도구가 답할 **질문**을 짓는 법부터 봅니다.

1.2 연구 질문 만들기

연구는 영감이 아니라 과정이다

연구를 "번뜩이는 아이디어 → 완벽한 연구 → 네이처 게재"의 직선으로 상상하기 쉽지만, 실제는 그렇지 않습니다. 좋은 질문은 대개 **단번에 나오지 않습니다.** 기존 연구, 현장의 필요, 예산, 시간이 서로 밀고 당기며 질문이 다듬어집니다. 연구는 직선이 아니라 **순환**입니다. 질문을 던지고, 부분적인 답을 얻고, 그 답이 다음 질문을 낳습니다. 그러므로 첫 질문을 완벽하게 지으려 멈춰 서지 말고, **설계로 옮길 수 있을 만큼 구체적인** 질문을 짓는 것이 목표입니다.

네 칸을 채운다 — P·E·C·O

설계로 옮길 수 있는 질문에는 네 요소가 들어 있습니다.

요소	무엇	예
대상(P)	누구를	40–59세 남성
노출(E)	무엇에 노출된	하루 20개비 이상 흡연
비교(C)	무엇과 비교해	비흡연자
결과(O)	무엇이 달라지는가	10년 내 폐암 발생

네 칸이 채워지면 **어떤 설계를 써야 하는지가 대체로 따라옵니다.**

- 노출을 먼저 재고 결과를 기다린다 → **코호트** (5장)
- 결과에서 출발해 노출을 거꾸로 본다 → **환자-대조군** (6장)
- 연구자가 노출을 직접 배정한다 → **중재연구** (7장)

질문의 문장을 먼저 쓰고 설계를 고르는 것이 아닙니다. **네 칸을 먼저 채우면 설계가 손에 잡힙니다.**

실패하는 연구 질문 세 가지

질문	무엇이 빠졌나	고치면
"스트레스는 건강에 어떤 영향을 주는가?"	**너무 넓다.** 대상도 결과도 정해지지 않았다. 무엇을 재야 하는지 알 수 없다	"교대근무 간호사에서 야간근무 월 8회 이상은 8회 미만보다 1년 내 우울증상 발생이 많은가?"

질문	무엇이 빠졌나	고치면
"미세먼지 경보일에 천식 응급실 방문이 많은가?"	**비교군이 없다.** '많다'는 무엇에 견주어 많은가? 경보가 아닌 날이 있어야 한다	"…경보일은 **비경보일에 비해** 천식 응급실 방문이 많은가?"
"금연 프로그램이 참가자의 삶의 질을 높이는가?"	**결과를 잴 수 없다.** '삶의 질'을 무엇으로 잴지 정해지지 않았다	"…참가자는 비참가자보다 6개월 후 **EQ-5D 점수**가 높은가?"

세 질문 모두 그럴듯하게 들립니다. 그러나 **네 칸 중 하나가 비어 있으면 설계로 옮길 수 없습니다.** 연구 질문을 만들 때는 P·E·C·O 네 칸을 먼저 채우고 문장은 그 다음에 씁니다.

> 주의 — 모든 연구가 반증할 가설을 세우는 것은 아닙니다. "환자들이 이 질환을 어떻게 경험하는가"처럼 **이해**를 목표로 하는 질적 연구도 있습니다. 이때도 질문은 정확해야 하지만, 기각할 가설을 세우는 방식은 아닙니다. 이 과목은 양적 연구를 다루므로 검증 가능한 질문에 집중합니다.

1.3 ★ 발생률과 유병률

이 장에서 가장 중요한 절입니다. 두 지표를 섞으면 5장부터 계속 틀립니다.

왜 율(rate)이 필요한가 — 수만으로는 말할 수 없다

A시에서 작년 폐렴 사망이 200명, B시에서 100명이었다고 합시다. A시가 더 위험한 곳일까요? 알 수 없습니다. A시 인구가 B시의 세 배라면 오히려 A시가 더 안전합니다. **사건의 수는 그것이 나온 인구와 기간을 함께 보아야만 의미를 가집니다.** 그래서 역학은 수가 아니라 **율**로 말합니다. 율은 분자(사건)와 분모(인구·시간)로 이루어집니다. 이 장의 두 율이 발생률과 유병률입니다.

무엇을 세는가

	발생률 (incidence)	유병률 (prevalence)
세는 것	**일정 기간 새로 생긴 사례**	**한 시점에 앓고 있는 사례**
분자	그 기간의 신규 사례 수	그 시점의 총 사례 수

	발생률 (incidence)	유병률 (prevalence)
분모	위험에 처한 사람 · **시간**	조사 시점의 인구
성격	**속도** (얼마나 빨리 생기나)	**스냅숏** (지금 얼마나 있나)
주로 쓰는 곳	원인 연구, 위험 비교	의료 자원 계획, 부담 추정

예를 들어 인구 400,000명 지역에 흡연자가 100,000명이면 흡연 **유병률**은 25%입니다. 한편 어떤 지역에서 한 해 동안 식중독이 새로 500건 생겼고 위험 인구가 50,000명이면 그 해 식중독 **발생률**은 1,000명당 10명입니다. 앞은 "지금 있는 것", 뒤는 "새로 생긴 것"입니다.

욕조로 생각하면

수도꼭지가 물을 채우는 속도가 **발생률**, 욕조에 담긴 물의 양이 **유병률**, 배수구로 빠지는 것이 **회복과 사망**입니다. 이 그림이 다음 사실을 한눈에 설명합니다.

유병률 ≈ 발생률 × 평균 이환기간 (유병률이 대략 10% 미만일 때).

- **요로감염**은 발생률이 높지만(자주 걸림) 며칠이면 낫습니다. 배수구가 넓어 물이 고이지 않으니 어느 시점에 앓고 있는 사람은 의외로 적습니다 — **높은 발생률, 낮은 유병률**
- **조현병**은 발생률이 낮지만(드물게 시작됨) 오래 지속됩니다. 배수구가 좁아 물이 고이니 어느 시점에 앓고 있는 사람의 비율은 상대적으로 높습니다 — **낮은 발생률, 높은 유병률**

그래서 유병률만 보면 안 되는 이유

> 치료가 좋아져 사망이 줄면 **유병률은 오히려 올라갑니다.** 병이 더 많이 생겨서가 아니라 환자들이 **오래 살기 때문**입니다. 배수구가 좁아진 것이지 수도꼭지가 커진 것이 아닙니다.

에이즈 치료가 발전한 뒤 유병률이 올라간 것이 이 예입니다. 새 감염이 줄어도, 환자들이 오래 살면서 어느 시점에 감염 상태로 살아가는 사람은 늘어납니다. **유병률 상승을 "병이 늘었다"로 읽으면 틀립니다.** 원인을 보려면 새로 생기는 것, 즉 **발생률**을 재야 합니다.

발생률의 분모는 사람이 아니다 — 인년(person-year)

추적하는 사람마다 관찰 기간이 다르면(중간에 병이 생기거나, 이사 가거나, 사망하면) 분모를 단순히 "사람 수"로 잡을 수 없습니다. 각자가 **위험에 노출된 시간**을 더한 것이 분모입니다. 이를 인년(person-year)이라 합니다.

```
발생률 = 새로 생긴 사례 수 / 총 관찰 인년
```

예를 들어 30인년을 관찰하는 동안 6건이 생겼다면 발생률은 6/30 = 0.2/인년, 곧 100인년당 20명입니다. 한 사람을 10년 보든 열 사람을 1년씩 보든 분모는 똑같이 10인년입니다. **5장에서 이 계산을 손으로 하게 됩니다.** 여기서는 "분모에 시간이 들어간다"는 것만 잡아 두십시오. 이 한 줄이 5장 코호트와 8장 생존분석을 관통합니다.

1.4 연구 설계의 지도

이 장 이후의 설계들이 어떻게 갈라지는지 한 장의 지도로 먼저 봅니다. 지금 외울 필요는 없습니다. 각 장에 들어갈 때 이 지도의 어디에 있는지만 알면 됩니다.

큰 갈래	설계	한 줄 정의	장
관찰 (연구자가 노출을 바꾸지 않음)	단면연구	한 시점에 노출과 결과를 함께 잰다	4장
	코호트	노출을 먼저 재고 결과를 기다린다	5장
	환자-대조군	결과에서 출발해 노출을 거꾸로 본다	6장
실험 (연구자가 노출을 배정)	중재연구(RCT)	무작위로 노출을 배정하고 결과를 본다	7장

관찰과 실험을 가르는 것은 **누가 노출을 정하는가**입니다. 연구자가 무작위로 배정하면 실험, 세상이 이미 정해 놓은 것을 관찰만 하면 관찰연구입니다. 이 차이가 인과 판단의 힘을 가릅니다. 왜 그런지는 7장에서 다룹니다.

1.5 예방의 개념

예방은 "좋은 정도"가 아니라 **시점**으로 나눕니다.

단계	언제	무엇을 막나	예
1차 예방	발병 **전**	병이 생기는 것 자체	금연, 백신, 나트륨 규제
2차 예방	발병 후 · **증상 전**	병이 진행하는 것	선별검사(암 검진)
3차 예방	발병 **후**	합병증·악화	재활, 합병증 관리

가장 자주 나오는 오해가 **선별검사를 1차 예방으로 넣는 것**입니다. "일찍 찾으니까 예방"이라고 생각하기 쉽지만, 선별검사를 받는 시점에 그 병은 **이미 몸에 있습니다.** 아직 증상이 없을 뿐입니다. 그러므로 선별검사는 발병 자체를 막는 1차가 아니라, 이미 생긴 병을 증상 전에 찾는 **2차 예방**입니다.

10장에서 이 표로 돌아옵니다. 그때는 한 걸음 더 나아가 "일찍 찾는 것이 반드시 좋은가"를 묻습니다. 조기 발견이 늘 이득은 아니라는 것, 그것이 10장의 주제입니다.

정리 — 이 장에서 가져갈 다섯 줄

1. 역학은 **인구집단에서 질병 빈도의 분포와 결정요인**을 연구한다. 단위는 개인이 아니라 집단이다
2. **검증할 수 없는 진술은 과학적 가설이 아니다.** 좋은 질문은 P·E·C·O 네 칸이 채워져 설계로 옮겨진다
3. **발생률은 새로 생긴 것(속도), 유병률은 지금 앓는 것(스냅숏).** 섞으면 뒤가 계속 틀린다
4. **유병률 ≈ 발생률 × 이환기간.** 오래 살면 발생이 그대로여도 유병률은 올라간다 — 유병률 상승을 발생 증가로 읽지 말 것
5. 발생률의 분모는 사람 수가 아니라 **인년**이다. 예방은 좋은 정도가 아니라 **시점**으로 나눈다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 다음 중 **과학적 가설로 쓸 수 있는** 진술은?

- A. 흡연은 사람의 건강에 해롭다
- B. 대기오염은 현대 사회의 큰 문제다
- C. 식습관은 여러 질병과 관련이 깊다
- D. 흡연자는 비흡연자보다 폐암이 많다

2. "40-59세 남성에서 하루 20개비 이상 흡연자는 비흡연자에 비해 10년 내 폐암 발생이 많은가?" 이 연구 질문에서 **비교(comparison)** 에 해당하는 것은?

- A. 40-59세의 남성 집단
- B. 10년 안의 폐암 발생
- C. 흡연하지 않는 사람
- D. 하루 20개비 이상 흡연

3. 발생률(incidence) 이 세는 것은?

- A. 조사 시점에 앓고 있는 사례
- B. 일정 기간 새로 생긴 사례
- C. 특정 질병으로 사망한 사례
- D. 의료기관을 방문한 사례

4. 어떤 질병의 **새로 생기는 환자 수는 그대로**인데 치료가 좋아져 환자들이 이전보다 오래 살게 되었다. 이때 이 질병의 **유병률**은?

- A. 증가한다
- B. 감소한다
- C. 변함없다
- D. 판단불가

5. 자궁경부암 선별검사는 예방의 어느 단계에 해당하는가?

- A. 1차 예방 — 발병을 막는다
- B. 2차 예방 — 증상 전에 찾는다
- C. 3차 예방 — 합병증을 막는다
- D. 예방이 아닌 진단 행위다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 기본 개념** — 발생률과 유병률은 둘 다 '얼마나 많은가'를 재는 것 같은데, 정확히 무엇을 세는지 서로 어떻게 다를까요? 그리고 왜 이 둘을 굳이 구별해야 할까요?
2. **2단계 · 개념의 실제화** — 이제 발생률과 유병률을 구별할 수 있으니, 평소 궁금했던 보건 문제를 하나 골라 '자료로 답할 수 있는' 연구 질문으로 바꿔 볼까요? 어떤 질문을 만들고 싶으세요?

제2장 · 루틴 데이터와 기술역학

제2장 루틴 데이터와 기술역학

> 이 장은 **연구를 시작하기 전에 이미 있는 자료**로 무엇을 할 수 있는지를 다룹니다.

이 장에서 답할 질문

- 새 조사를 하기 전에, 이미 있는 자료로 어디까지 알 수 있는가
- 사건의 **수**가 아니라 **비율**을 써야 하는 이유
- 치우친 분포를 어떤 값으로 요약해야 하는가
- 집단 수준의 연관을 개인에게 옮길 때 무슨 일이 생기는가

1장에서 좋은 질문을 짓는 법을 배웠습니다. 그런데 값비싼 새 연구를 시작하기 전에, **이미 세상에 쌓여 있는 자료**로 답의 절반이 나오는 경우가 많습니다. 이 장은 그 자료를 읽는 눈을 기릅니다.

2.1 일상적으로 수집되는 보건 정보

연구자가 만들지 않아도 **이미 모이고 있는 자료**가 있습니다. 이를 루틴 자료(routine data)라 합니다. 핵심은 **연구자와 무관하게, 행정·진료의 과정에서 계속 쌓인다**는 점입니다. 이것이 4장의 서베이처럼 가설을 검증하려고 특별히 설계해 모으는 자료와 갈라지는 지점입니다.

출처	무엇
인구 통계	인구 조사, 주민등록
등록 자료	사망 등록, 암 등록, 출생 등록
의료 이용	건강보험 청구, 퇴원 요약
감시 체계	감염병 신고, 예방접종
환경 감시	대기 오염, 수질

- **장점** — 이미 있으므로 빠르고 싸다. 전 인구를 덮는 경우가 많아 표집 오차가 없다
- **한계** — **연구자가 원하는 대로 수집되지 않았다.** 변수의 정의, 완전성(빠진 것), 정확도를 항상 의심하고 확인해야 한다. 원하는 변수가 아예 없을 수도 있다

사망 등록이 만들어지는 과정 — 자료의 질은 어디서 갈리나

사망 원인은 의사가 사망진단서에 적습니다. 그런데 **갑작스럽거나 예상치 못한 사망**은 의사가 바로 쓸 수 없고 **검시관의 조사와 부검**을 거칩니다. 이 절차의 충실도가 통계의 질을 좌우합니다.

> 사망 원인 통계는 진단서를 쓴 의사의 판단, 그 나라의 분류 관행, 부검 시행률에 따라 달라집니다. **같은 병이 나라마다 다른 사인으로 분류될 수 있습니다.** 국가 간 사망률을 비교할 때 늘 물어야 할 것은 "이 두 나라가 같은 방식으로 세었는가"입니다.

이것이 루틴 자료를 쓸 때의 첫 번째 원칙입니다. **숫자를 믿기 전에 그 숫자가 어떻게 만들어졌는지를 먼저 본다.** 자료는 하늘에서 떨어지지 않습니다.

2.2 기술역학 — 사람·장소·시간

기술역학(descriptive epidemiology)은 질병의 **분포**를 봅니다. 세 축으로 나눕니다.

축	무엇을 보나	무엇을 시사하나
사람(person)	연령·성별·직업·사회계층	감수성, 노출 기회
장소(place)	지역·국가·도농	환경 요인, 의료 접근성
시간(time)	추세·계절·유행	원인의 변화, 감염 여부

예를 들어 자살률이 성별·사회경제 계층에 따라 크게 다르다는 것(사람), 어떤 감염병이 특정 지역에 몰린다는 것(장소), 인플루엔자가 겨울에 오른다는 것(시간)은 모두 기술역학의 관찰입니다.

주의할 것은 **이 세 축은 질병 지표(유병률·발생률·사망률)와 층위가 다르다**는 점입니다. 사람·장소·시간은 "어디를 쪼개서 볼 것인가"라는 **변이의 축**이고, 유병률·발생률은 각 칸에서 **재는 값**입니다. "기술역학의 세 축이 무엇인가"라는 물음에 "유병률·발생률·사망률"이라 답하면 축과 지표를 섞은 것입니다.

> **기술역학은 가설을 만드는 단계입니다**
> **검증은 5~7장의 설계로 합니다**

"겨울에 오르고, 특정 지역에 몰리고, 젊은 층에 많다"는 분포는 **원인에 대한 단서**를 줍니다. 그러나 그 자체로 원인을 증명하지는 못합니다. 분포는 질문을 낳고, 설계가 답합니다.

2.3 ★ 수가 아니라 비율

> **사건의 수만으로는 아무것도 비교할 수 없습니다**

A시에서 폐암 사망이 500건, B시에서 200건이라 해도 A시가 더 위험한 것이 아닙니다. **A시 인구가 5배 많다면 오히려 B시가 더 위험합니다.**

```
비율 = 사건 수 / 인구 수    (그리고 언제나 '기간'을 함께 밝힌다)
```

분모를 밝히지 않은 숫자는 정보가 아닙니다. 뉴스에서 "올해 환자가 3천 명"이라 하면 먼저 "몇 명 중에서?"를 물어야 합니다. 이것이 1장에서 본 발생률·유병률이 모두 율(rate)인 이유입니다.

> 여기서 한 걸음 더 나아가면 **3장의 표준화**입니다. 비율을 써도 두 인구의 **연령 구조**가 다르면 여전히 공정한 비교가 아니기 때문입니다. 은퇴자가 많은 도시는 젊은 도시보다 조사망률이 높은 것이 당연합니다. 그것을 "덜 건강한 도시"로 읽지 않으려면 3장이 필요합니다.

분포를 요약하는 법 — 중심과 퍼짐

자료 한 무더기를 두 숫자로 줄이려 합니다. 하나는 **중심**(대푯값), 하나는 **퍼짐**(변동)입니다. 그런데 어떤 값을 쓸지는 **분포의 모양**에 달려 있습니다.

목적	대칭 분포	치우친 분포
중심	평균	**중앙값**

목적	대칭 분포	치우친 분포
퍼짐	표준편차	**사분위수 범위(IQR)**

왜 치우친 자료에 평균을 쓰면 안 되나 — 재원일수 예

한 병동의 환자 9명의 재원일수가 다음과 같다고 합시다.

```
2, 2, 3, 3, 4, 4, 5, 6, 61  (일)
```

여덟 명은 2~6일 만에 퇴원했고, 한 명(중증)이 61일 입원했습니다.

- **평균** = (2+2+3+3+4+4+5+6+61) / 9 = 90 / 9 = **10일**
- **중앙값** = 가운데(5번째) 값 = **4일**

평균 10일은 이 병동의 전형적인 환자를 전혀 대표하지 못합니다. 실제로 열에 여덟은 6일 안에 나갑니다. **단 한 명의 극단값(61일)이 평균을 두 배 넘게 끌어올렸습니다.** 중앙값 4일이 이 자료의 중심을 훨씬 정직하게 말합니다. 소득, 재원일수, 검사수치처럼 오른쪽으로 꼬리가 긴 자료는 거의 언제나 이런 일이 생기므로 중앙값과 IQR로 요약합니다.

두 변수의 관계

- **상관계수 r** — −1에서 +1. 두 연속변수의 **선형** 관계의 방향과 강도
- **결정계수 r^2** — 한 변수의 변동 중 다른 변수로 설명되는 비율

> **상관은 인과가 아닙니다.** 상관이 강해도, 두 변수를 함께 움직이는 제3의 요인(교란)이 있을 수 있습니다. 이 경계가 1장의 반증 가능성, 5장의 Bradford Hill 관점으로 이어집니다.

2.4 ★ 생태학적 오류

집단 수준의 연관을 개인에게 그대로 옮기면 틀릴 수 있습니다. 이 오류에 이름이 붙어 있습니다 — 생태학적 오류(ecological fallacy).

> 국가별로 **평균 지방 섭취량이 높을수록 유방암 발생률이 높다**는 관찰이 있습니다. 그렇다고 **"지방을 많이 먹는 개인이 유방암에 걸린다"** 고 말할 수 없습니다.

왜일까요? 집단 수준의 자료는 **개인 안에서 노출과 결과가 함께 가는지**를 알려주지 않기 때문입니다. 지방을 많이 먹는 나라에서 유방암에 걸린 사람들이 정작 지방을 **적게** 먹은 사람들일 수도 있습니다. 그 나라의 유방암을 실제로 끌어올린 것은 지방이 아니라, 지방 섭취와 함께 가는 다른 요인(출산 시기, 비만, 검진율)일 수도 있습니다. 집단의 평균은 이 안쪽을 보여주지 않습니다.

> **생태학적 연구는 가설을 만드는 데 쓰고**
> **검증은 개인 자료로 합니다**

생태학적 연구가 쓸모없다는 뜻이 아닙니다. 값싸고 빠르게 큰 그림의 가설을 줍니다. 다만 그 가설은 반드시 **개인 단위 설계**로 확인해야 합니다. 집단의 관찰을 개인의 결론으로 건너뛰는 순간 틀립니다.

2.5 역학 연구 설계 개요

이 장은 "이미 있는 자료"와 "가설 만들기"를 다뤘습니다. 다음 장들은 그 가설을 **검증**하는 설계입니다. 아래 표가 5~7장으로 가는 지도입니다.

설계	방향	인과 근거	장
사례 보고·사례군	—	가장 약함. **비교군이 없다**	—
생태학적 연구	집단	약함	2장
단면연구	한 시점	약함. **선후관계 불명**	4장
환자-대조군	결과 → 노출	중간	6장
코호트	노출 → 결과	강함	5장
중재연구(RCT)	배정 → 결과	**가장 강함**	7장

아래로 갈수록 인과 근거가 강해지고, 대개 비용과 시간도 커집니다. 이 사다리에서 자기 연구가 어디에 있는지를 아는 것이 방법론의 기본기입니다.

> 사례 보고의 약점은 딱 하나로 요약됩니다 — **비교할 대상이 없다.** "이 약을 먹은 환자 3명이 좋아졌다"는 관찰은, 그 약을 안 먹은 비슷한 환자들이 어땠는지를 모르면 아무것도 증명하지 못합니다. 가설을 만드는 데는 유용하지만 확증할 수는 없습니다. 비교군의 부재 — 이것이 이 과목 전체를 관통하는 첫 번째 약점입니다.

정리 — 이 장에서 가져갈 다섯 줄

1. 이미 모이는 **루틴 자료**로 많은 것을 할 수 있다. 다만 연구자가 원하는 대로 수집되지 않았으니 정의·완전성·정확도를 먼저 의심한다
2. **사건의 수가 아니라 비율**을 봐야 한다. 분모(와 기간) 없는 숫자는 정보가 아니다
3. 치우친 분포에는 평균이 아니라 **중앙값과 사분위수 범위**를 쓴다. 극단값 하나가 평균을 끌고 간다
4. **생태학적 오류** — 집단의 연관을 개인에게 옮기면 틀릴 수 있다. 생태학적 연구는 가설용, 검증은 개인 자료로
5. 설계마다 인과 근거의 강도가 다르다. **사례 보고의 약점은 비교군이 없다는 것**이다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 일상적으로 수집되는 보건 정보의 출처가 **아닌** 것은?

- A. 인구 통계 자료 (예 — 인구 조사)
- B. 등록 자료 (예 — 사망·암 등록)
- C. 가설 검증용으로 고안된 조사 자료
- D. 환경 감시 자료 (예 — 대기 오염)

2. **기술역학**에서 건강의 변이를 볼 때 쓰는 **세 축**은?

- A. 나이 · 소득 · 교육

- B. 건강 · 질병 · 사망
- C. 유병 · 발생 · 사망률
- D. 사람 · 장소 · 시간

3. 사망률을 해석할 때 **사건의 수만으로는 부족하고** 인구 크기를 고려한 **비율**을 써야 하는 이유는?

- A. 비율은 어떤 자료에서든 정규 분포를 따르기 때문이다
- B. 인구 크기가 다르면 사건 수의 순위가 뒤집힐 수 있다
- C. 비율을 써야 통계적 유의성을 계산할 수 있기 때문이다
- D. 비율은 기간 유병률을 재는 데에만 쓸 수 있기 때문이다

4. 분포가 심하게 **치우쳐 있을 때** 중심 위치를 가장 잘 나타내고 극단값에 덜 민감한 것은?

- A. 중앙값
- B. 평균
- C. 최빈값
- D. 표준편차

5. **인구 집단 수준**에서 노출 요인과 건강 결과의 상관을 조사하는 연구 설계는?

- A. 코호트 연구
- B. 환자-대조군 연구
- C. 생태학적 연구
- D. 중재 연구

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 자료를 읽는 눈** — A지역의 암 사망자가 B지역보다 두 배 많다는 기사를 봤습니다. A지역이 더 위험한 곳이라고 결론지어도 될까요? 무엇을 더 알아야 할까요?
2. **2단계 · 자료의 한계** — 나라별로 지방 섭취량이 많은 나라일수록 유방암이 많다는 그래프가 있습니다. 그럼 '지방을 많이 먹는 사람'이 유방암에 걸린다고 말할 수 있을까요?

제3장 · 표준화

제3장 표준화

> **계산이 나옵니다. 표를 눈으로만 읽지 말고 3.3의 예제를 한 번 손으로 따라 해 보십시오.**

이 장에서 답할 질문

- 조율(crude rate)이 왜 두 지역을 공정하게 비교하지 못하는가
- **간접 표준화(SMR)와 직접 표준화는 무엇이 다른가** — 이 장의 핵심
- SMR 120은 무슨 뜻이며, 왜 다른 지역의 SMR과 직접 비교하면 안 되는가

3.1 왜 표준화가 필요한가

2장에서 **수 대신 비율**을 써야 한다고 했습니다. 그런데 비율로도 부족합니다.

> 은퇴자가 많이 사는 A지역의 **조사망률**이 젊은 노동자가 많은 B지역보다 높습니다. A지역이 더 건강하지 않은 곳일까요? **아닙니다. 나이가 많으면 원래 더 많이 죽습니다.**

> **조율의 차이는 건강의 차이일 수도 있고**
> **단지 연령 구조의 차이일 수도 있습니다**

조사망률(crude death rate)은 그 지역에서 죽은 사람 수를 전체 인구로 나눈 값입니다. 그런데 사망은 나이에 크게 좌우되므로, 노인이 많은 지역은 **건강 수준과 무관하게** 조사망률이 높게 나옵니다. 두 지역의 건강을 견주려면 **연령 구조의 차이를 먼저 걷어내야** 합니다. 그 일을 하는 것이 표준화입니다. 표준화는 **"두 지역의 연령 구조가 같았다면 어땠을까"** 를 계산합니다.

> 여기서 걷어내는 연령의 효과가 곧 5·6장에서 다룰 **교란(confounding)** 입니다. 연령은 사망과도 관련되고 지역과도 관련되므로 전형적인 교란요인입니다. 표준화는 **분석 단계에서 교란을 다루는 가장 오래된 방법**입니다.

3.2 ★ 간접 표준화 — SMR

아이디어

> **"표준인구의 연령별 사망률이 우리 지역에 그대로 적용된다면 몇 명이 죽었을까?"** 그 **기대 사망자 수**와 **실제 관찰된 사망자 수**를 비교합니다.

$$\text{SMR} = \frac{\text{관찰 사망자 수 (O)}}{\text{기대 사망자 수 (E)}} \times 100$$

계산 순서

단계	하는 일
1	**표준인구**의 연령별 사망률을 가져온다
2	각 연령대에서 **표준 사망률 × 우리 지역 인구** = 기대 사망자 수
3	연령대별 기대값을 **모두 더한다** → E
4	실제 사망자 수 O를 E로 나누고 100을 곱한다

해석

SMR	뜻
100	표준인구와 같다
120	**기대보다 20% 많이 죽었다**

SMR	뜻
85	기대보다 15% 적게 죽었다

간접 표준화를 쓰는 때

> **우리 지역의 연령별 사망률을 모를 때 씁니다**

필요한 것은 **우리 지역의 연령별 인구 수**와 **총 사망자 수**뿐입니다. 연령대별로 각각 몇 명이 죽었는지는 몰라도 됩니다. 그래서 **작은 지역·소수 직업군**처럼 연령별 사망자 수가 너무 적어 연령별 비율이 들쭉날쭉 불안정한 경우에 간접 표준화를 씁니다.

3.3 ★ 직접 표준화

아이디어

> **"우리 지역의 연령별 사망률이 표준인구에 적용된다면 사망률이 얼마일까?"** 방향이 간접 표준화와 **정반대**입니다.

$$\text{표준화 사망률} = \frac{\Sigma\,(\text{우리 지역 연령별 사망률} \times \text{표준인구 연령별 인구})}{\text{표준인구 총 인구}}$$

한 자료로 끝까지 계산해 보기

두 지역을 봅니다. 계산을 단순하게 하려고 연령을 **청년·노년** 둘로만 나눴습니다.

	A지역(은퇴 도시) 인구	A 사망	A 연령별률	B지역(대학가) 인구	B 사망	B 연령별률
청년	10,000	20	**2/1,000**	40,000	80	**2/1,000**
노년	40,000	800	**20/1,000**	10,000	200	**20/1,000**

	A지역(은퇴 도시) 인구	A 사망	A 연령별률	B지역(대학가) 인구	B 사망	B 연령별률
합	50,000	820		50,000	280	

먼저 **조사망률**을 봅니다.

- A: 820 / 50,000 = **16.4/1,000**
- B: 280 / 50,000 = **5.6/1,000**

조율만 보면 A가 B보다 세 배 가까이 나쁩니다. 그런데 표의 **연령별 사망률을 보십시오.** 청년 2/1,000, 노년 20/1,000 — **두 지역이 모든 연령에서 완전히 똑같습니다.** A가 나빠 보인 것은 오로지 노인이 많기 때문입니다. 이제 표준화가 이 착시를 걷어내는지 확인합니다.

직접 표준화 (표준인구를 청년 25,000·노년 25,000, 합 50,000으로 잡음):

- A: (2/1,000 × 25,000) + (20/1,000 × 25,000) = 50 + 500 = 550 → 550 / 50,000 = **11.0/1,000**
- B: (2/1,000 × 25,000) + (20/1,000 × 25,000) = 50 + 500 = 550 → **11.0/1,000**

같습니다. 직접 표준화는 두 지역이 사실 똑같이 건강하다는 것을 정확히 말해 줍니다. 결과가 율(rate)이므로 A와 B를 **직접 견줄 수 있습니다.**

간접 표준화(SMR) 도 해 봅니다(표준 연령별률을 청년 3/1,000·노년 15/1,000로 잡음):

- A 기대 E = (3/1,000 × 10,000) + (15/1,000 × 40,000) = 30 + 600 = 630. 관찰 O = 820 → SMR = 820/630 × 100 ≈ **130**
- B 기대 E = (3/1,000 × 40,000) + (15/1,000 × 10,000) = 120 + 150 = 270. 관찰 O = 280 → SMR = 280/270 × 100 ≈ **104**

여기서 결정적인 것을 봅니다. 두 지역의 연령별 사망률이 **완전히 같은데도** SMR은 130과 104로 다릅니다. 왜냐하면 SMR은 각 지역의 **자기 연령 구조**를 표준률과 곱해 계산하기 때문입니다. 그러므로 **A의 130과 B의 104를 "A가 더 나쁘다"로 비교하면 틀립니다.** 각 SMR은 오직 자기 기대치(100)와만 비교할 수 있습니다. 같은 자료를 직접 표준화로 보면 둘은 똑같았습니다.

두 방법의 대조

	간접 표준화 (SMR)	직접 표준화
빌려오는 것	**표준인구의 사망률**	**표준인구의 인구 구조**

	간접 표준화 (SMR)	직접 표준화
우리 지역에서 필요한 것	연령별 **인구**, 총 사망자 수	연령별 **사망률**
결과	비(ratio), 보통 ×100	**율(rate)**
지역 간 직접 비교	**하면 안 된다**	**할 수 있다**
쓰는 때	사건 수가 적을 때	연령별 율을 알 때

> **간접은 율을 빌려오고**
> **직접은 인구를 빌려옵니다**

이 한 줄만 기억하면 나머지는 따라옵니다.

3.4 연령 말고 다른 요인

표준화는 연령 전용이 아닙니다. **성별·사회계층·인종**으로도 할 수 있습니다. 원리는 같습니다 — **비교하려는 집단들의 구성을 같게 맞춰 놓고 본다.**

다만 표준화는 **표준화한 그 요인만** 걷어냅니다. 연령으로 표준화했다면 **흡연·소득 같은 다른 교란은 그대로 남아 있습니다.** 연령을 맞췄다고 두 지역이 모든 면에서 비교 가능해진 것이 아닙니다.

> 그래서 5장 이후에는 여러 교란을 **동시에** 다루는 다변량 분석(회귀)으로 넘어갑니다. 표준화는 교란을 다루는 가장 오래되고 직관적인 방법이지만, 한 번에 한둘의 요인만 감당합니다.

정리 — 이 장에서 가져갈 다섯 줄

1. **조율의 차이는 건강의 차이가 아닐 수 있다.** 연령 구조가 다르면 비교가 성립하지 않는다
2. **간접(SMR)은 표준인구의 율을 빌려온다.** SMR = O/E × 100
3. **직접은 표준인구의 인구 구조를 빌려온다.** 결과가 율이므로 지역 간 비교가 된다

4. **SMR 120 = 기대보다 20% 많이 죽었다.** 각 SMR은 100과만 비교하고, 다른 지역의 SMR과 직접 비교하지 않는다
5. 표준화는 **표준화한 요인만** 걷어낸다. 나머지 교란은 남으므로 다변량 분석이 필요하다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 표준화를 하는 주된 목적은?

- A. 코호트 연구의 통계적 검정력을 높이기 위해서다
- B. 환자-대조군 연구의 오즈비를 정확히 구하기 위해서다
- C. 집단 간 인구 구조(연령 분포)의 차이를 보정하기 위해서다
- D. 표본 추출 오류를 없애 가설 검증을 쉽게 하기 위해서다

2. 관찰 사망자 수를 기대 사망자 수로 나눈 뒤 100을 곱해 얻는 지표는?

- A. 누적 발생률
- B. 연령별 사망률
- C. 표준화 사망비
- D. 상대 위험도

3. SMR의 해석으로 옳은 것은?

- A. SMR = 100 이면 인덱스 인구와 표준 인구의 사망 수준이 같다
- B. SMR > 100 이면 인덱스 인구의 사망 수준이 표준보다 낮다
- C. SMR < 100 이면 인덱스 인구의 사망 수준이 표준보다 높다
- D. SMR = 1 이면 인덱스 인구와 표준 인구의 사망 수준이 같다

4. 직접 표준화와 간접 표준화의 **근본적인 차이**는?

- A. 직접은 사망률만, 간접은 유병률만 다룬다는 점이다
- B. 간접은 표준 인구의 율을 인덱스 인구의 수에 적용한다
- C. 직접은 연령을 반드시 쓰지만 간접은 그렇지 않다는 점이다

- D. 간접이 인과 관계 판단에 더 강력한 근거를 준다는 점이다

5. 간접 표준화에 대한 설명으로 옳지 **않은** 것은?

- A. 표준인구의 범주별 율을 인덱스 인구에 적용해 기대값을 구한다
- B. 인덱스 인구의 범주별 율을 몰라도 계산할 수 있다는 장점이 있다
- C. 서로 다른 집단의 SMR을 직접 비교할 수 없다는 한계가 있다
- D. 서로 다른 집단의 SMR을 직접 비교할 수 있다는 장점이 있다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 왜 표준화가 필요한가** — 은퇴자가 많이 사는 도시의 조사망률이 대학가 도시보다 높게 나왔습니다. 이 도시가 더 건강하지 못한 곳이라고 봐도 될까요?
2. **2단계 · SMR을 읽는 법** — A지역 SMR이 120, B지역이 110입니다. A가 B보다 10% 더 나쁘다고 비교해도 될까요? SMR은 원래 무엇과 비교하라고 만든 값일까요?

제4장 · 서베이

제4장 서베이

> 이 장은 **자료를 직접 만들어 내는** 첫 장입니다. 2~3장은 있는 자료를 썼습니다.

이 장에서 답할 질문

- 표본을 어떻게 뽑아야 모집단을 대표하는가
- **표본이 크면 편향이 줄어드는가** — 이 장에서 가장 많이 틀리는 지점
- 재는 도구가 좋다는 것은 무슨 뜻인가

> **이 장을 관통하는 물음** — 서베이는 인구 집단을 재려고 그중 일부(표본)만 뽑는다. 그러므로 모든 것은 결국 하나로 모인다: **이 표본이 인구 집단을 대표하는가.** 대표하지 못하면, 아무리 크고 정밀한 조사라도 인구의 현실이 아니라 뽑힌 사람들의 현실만 재게 된다.

4.1 서베이는 무엇을 하는 연구인가

서베이는 **한 시점**에 모집단의 상태를 재는 연구입니다. 2장의 분류로는 **단면연구**입니다.

할 수 있는 것	할 수 없는 것
유병률 추정	**선후관계 판단**
요구도 파악	인과 추론
변수 간 연관 관찰	

> 단면연구에서 노출과 결과를 **같이** 재기 때문에 어느 쪽이 먼저인지 알 수 없습니다. 이것이 6장 환

> 자-대조군, 5장 코호트로 넘어가는 이유입니다.

4.2 ★ 표집 방법

확률 표집

방법	하는 일	특징
단순 무작위	모두에게 같은 확률	기준. 명부가 필요
계통	k번째마다 뽑는다	간편. **명부에 주기성이 있으면 위험**
층화	층으로 나눈 뒤 각 층에서 뽑는다	**소수 집단을 확실히 포함**. 정밀도 향상
집락	집단(학교·마을)째로 뽑는다	흩어진 모집단에 실용적. **정밀도는 손해**

층화와 집락은 정반대다

둘 다 모집단을 나눈다는 점은 같지만 방향이 반대입니다.

	층화	집락
나누는 기준	층 안이 **비슷**하도록	집락 안이 **모집단을 닮도록**
뽑는 방식	**모든 층에서** 뽑는다	**일부 집락만** 뽑는다
정밀도	**올라간다**	**내려간다**
목적	정확성	**비용·실행 가능성**

> **층화는 정밀도를 위해**
> **집락은 비용을 위해 씁니다**

표집틀

표집틀은 모집단의 명부입니다. 여기에 없는 사람은 처음부터 뽑힐 수 없습니다.

> 전화번호부로 표집틀을 만들면 **전화가 없는 사람은 영원히 빠집니다.** 이 사람들이 관심 있는 특성에서 다르다면 **아무리 잘 뽑아도 편향이 남습니다.**

4.3 ★ 표본 크기와 편향 — 가장 많이 틀리는 지점

> **표본을 키우면 우연오차가 줄고**
> **편향은 그대로 남습니다**

	우연오차	편향
원인	표집의 무작위성	**잘못된 절차**
방향	무작위	**한쪽으로 치우침**
표본을 키우면	**줄어든다**	**그대로다**
대책	표본 크기, 신뢰구간	**설계 단계에서 막는다**

> 표집틀이 잘못된 조사를 10배로 키우면 **틀린 답을 더 좁은 신뢰구간으로** 얻게 됩니다. **더 확신을 갖고 틀리게 됩니다.**

이 구분은 **11장(우연·확률)** 과 **5·6장(편향)** 으로 각각 이어집니다.

신뢰구간

표본에서 얻은 추정치에는 **우연에 의한 폭**이 있습니다. 그 폭을 나타낸 것이 신뢰구간입니다.

- **표본이 커지면** 신뢰구간이 좁아진다
- 신뢰구간의 폭은 표본 크기의 **제곱근**에 반비례한다 — **4배로 늘려야 절반**이 된다

표본이 4배여야 오차가 절반 — 숫자로

유병률(비율 p)을 추정할 때 95% 신뢰구간의 한쪽 폭(오차 한계)은 대략 다음과 같습니다.

```
오차 한계 ≈ 1.96 × √( p(1−p) / n )
```

가장 불리한 경우인 p = 0.5로 놓고 표본 크기만 바꿔 봅니다.

표본 n	오차 한계
100	약 **±10%p**
400	약 **±5%p**
1,600	약 **±2.5%p**

오차를 ±10%p에서 ±5%p로, 즉 **절반**으로 줄이려면 표본을 100에서 **400으로 네 배** 늘려야 합니다. 다시 절반(±2.5%p)으로 줄이려면 또 네 배가 필요합니다. 분모에 √n 이 있기 때문입니다. 그래서 "표본을 두 배로 늘리면 오차가 절반"이라는 직관은 틀립니다. **정밀도는 비싸게 삽니다.**

> 📌 주의 — 이 계산은 **우연오차(정밀도)** 에 대한 것입니다. 표본을 아무리 키워도 아래 4.3 앞부분의 **편향**은 줄지 않습니다. 큰 표본은 틀린 답을 더 좁은 구간으로 줄 뿐입니다.

4.4 응답률

응답하지 않은 사람이 응답한 사람과 다르면 결과가 치우칩니다 — 무응답 편향입니다.

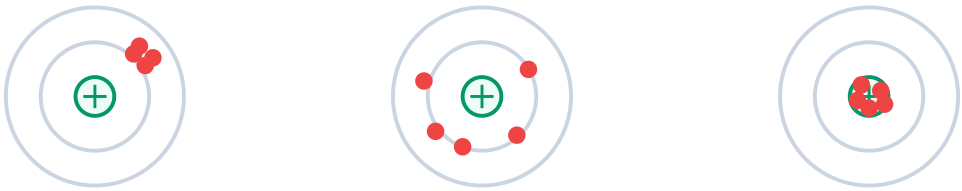
> 흡연 조사에서 흡연자가 덜 응답하면 흡연율이 **과소 추정**됩니다.

응답률을 높이는 방법: 사전 안내, 재접촉, 설문 단축, 응답 편의 제공. **응답률 자체보다 응답자와 무응답자가 다른지가 중요합니다.**

4.5 ★ 측정 — 타당도와 신뢰도

	묻는 것	실패하면
타당도 (validity)	**재려는 것을 재고 있는가**	체계적으로 틀린다
신뢰도 (reliability)	**다시 재도 같은 값이 나오는가**	값이 흔들린다

과녁 비유



신뢰도 O · 타당도 X
모였지만 과녁을 벗어남

신뢰도 X · 타당도 O
가운데 근처지만 흩어짐

둘 다 O
모였고 가운데다

> **신뢰도가 높아도 타당도는 없을 수 있습니다**
> **그 반대는 성립하기 어렵습니다**

눈금이 3kg 틀어진 저울은 **재현성은 완벽하지만 타당하지 않습니다. 신뢰도는 타당도의 필요조건이지 충분조건이 아닙니다.**

신뢰도의 종류

- **검사-재검사** — 같은 사람에게 시간을 두고 두 번
- **관찰자 내 / 관찰자 간** — 같은 사람이 두 번 / 다른 사람이 각각
- **내적 일관성** — 같은 것을 재는 문항들이 서로 일치하는가

4.6 자료의 종류

종류	예	요약
명목	혈액형, 성별	빈도·비율
순서	통증 등급, 만족도	**중앙값**. 간격이 같지 않다

종류	예	요약
이산	자녀 수	평균 또는 중앙값
연속	혈압, 체중	평균·표준편차 (대칭일 때)

순서형은 순서만 있고 간격이 같지 않습니다. "매우 불만 1 – 매우 만족 5"에서 1→2와 4→5가 같은 크기라는 보장이 없습니다. 그래서 평균보다 **중앙값**이 안전합니다.

정리 — 이 장에서 가져갈 다섯 줄

1. 서베이는 단면연구다. 유병률은 알 수 있지만 **선후관계는 알 수 없다**
2. **층화는 정밀도를 위해, 집락은 비용을 위해** 쓴다
3. **표집틀에 없는 사람은 뽑힐 수 없다.** 여기서 생긴 편향은 표본을 키워도 남는다
4. **표본을 키우면 우연오차만 줄어든다.** 편향은 설계 단계에서 막아야 한다
5. **타당도는 맞는 것을 재는가, 신뢰도는 다시 재도 같은가.** 신뢰도는 타당도의 필요조건일 뿐이다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 모집단의 **모든 개인이 표본에 포함될 확률이 같다**는 원칙을 가진 표집 방법은?

- A. 단순 무작위 추출
- B. 체계적 표본 추출
- C. 할당 표본 추출
- D. 편의 표본 추출

2. 빈칸에 들어갈 말을 순서대로 고르시오.

층화 추출은 같은 크기의 단순 무작위 표본에 비해 정밀도를 **(A)** 시키므로 **(B)** 표본을 쓸 수 있다. 집락 추출은 정밀도를 **(C)** 시키므로 **(D)** 표본이 필요하다.

- A. (A)감소 — (B)작은 — (C)증가 — (D)큰
- B. (A)증가 — (B)큰 — (C)감소 — (D)작은
- C. (A)증가 — (B)작은 — (C)감소 — (D)큰
- D. (A)감소 — (B)큰 — (C)증가 — (D)작은

3. 표집틀(sampling frame) 이 반드시 갖춰야 하는 두 가지 기능은?

- A. 연령별 사망률 자료와 성별 분포 자료를 담아야 한다
- B. 표본을 뽑을 정보를 주고 뽑힌 사람에게 연락이 되어야 한다
- C. 인구 이동률을 포함하고 코호트 분석을 지원해야 한다
- D. 표준화된 자료를 담고 95% 신뢰 구간을 제공해야 한다

4. 허용 오차 범위를 **절반으로 줄이려면** 필요한 표본 크기는 어떻게 되는가?

- A. 2배로 증가한다
- B. 절반으로 준다
- C. 변하지 않는다
- D. 4배로 증가한다

5. 타당도(validity) 와 **신뢰도(reliability)** 를 가장 잘 설명한 것은?

- A. 타당도는 무작위 오류를, 신뢰도는 체계적 오류를 재는 것이다
- B. 타당도는 참값에 가까운 정확성, 신뢰도는 측정의 일관성이다
- C. 타당도는 95% 신뢰 구간을, 신뢰도는 표준 오차를 뜻한다
- D. 타당도는 작성 소요 시간을, 신뢰도는 응답률을 재는 것이다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 표집** — 설문 응답자를 '길에서 편하게 만난 사람'으로 모으는 것과 '명부에서 무작위로 뽑는 것'은 결과를 해석할 때 무엇이 달라질까요?
2. **2단계 · 표본 크기와 측정** — 설문 표본을 10배로 늘리면 결과가 그만큼 더 정확하고 믿을 만해질까요? 표본을 아무리 키워도 줄지 않는 오차가 있을까요?

제5장 · 코호트 연구

제5장 코호트 연구

> 수업에서는 이 내용을 다시 강의하지 않습니다. 여러분이 틀린 것만 다룹니다.

이 장에서 답할 질문

- 왜 굳이 사람을 몇 년씩 추적하는가? 더 빠른 설계를 두고 코호트를 고르는 이유는 무엇인가
- 추적 중에 사람이 빠져나가면 결과는 어느 쪽으로 틀어지는가
- 연관성을 보았을 때, 그것을 인과라고 부르려면 무엇이 더 필요한가

> **이 장을 관통하는 물음** — 코호트가 재려는 것은 인구 집단의 발생이다. 그런데 추적 도중 특정한 사람들이 특정한 방향으로 빠져나가면, 그 발생률은 인구의 현실이 아니라 **끝까지 남은 사람들의 현실**이 된다. 5.4의 추적 탈락이 이 장에서 가장 중요한 이유다.

5.1 왜 코호트 연구를 하는가

코호트 연구의 정의는 한 문장이면 끝난다. **노출을 먼저 재고, 그다음에 일어나는 일을 지켜본다.** 이 순서가 설계에 박혀 있다는 것이 이 장 전체를 지배한다.

이 순서 때문에 얻는 것:

- **발생률을 직접 계산할 수 있다.** 분모(위험에 처한 사람과 시간)를 알기 때문이다. 환자-대조군은 이걸 못 한다 — 그래서 RR 대신 OR을 쓴다
- **시간적 선후관계가 확보된다.** Bradford Hill 9기준 중 유일한 필요조건을 설계가 자동으로 만족시킨다
- 한 노출에 대해 **여러 결과**를 동시에 볼 수 있다

대가로 치르는 것: 시간, 비용, 그리고 **추적 탈락**. 5.4가 이 대가를 다룬다.

이 장의 관통 사례 — 영국 지역 심장 연구(BRHS)

이 장은 하나의 연구를 끝까지 따라간다. **British Regional Heart Study(BRHS)** 다.

- 영국 24개 도시에서 40–59세 남성 **7,735명**을 모집해, 기초 설문으로 흡연·운동·혈압 등을 측정하고 이후 수년간 사망과 발병을 추적했다
- 원래 심혈관질환을 보려고 설계했으나 같은 코호트로 암 발생도 분석했다. 이 코호트로 얻은 암 발생 분석과 흡연·허혈성심질환 분석을 이 장에서 함께 인용한다

한 연구를 5.1부터 5.7까지 계속 참조하는 이유가 있다. 설계·측정·추적·분석·인과판단은 **서로 다른 다섯 과목이 아니라 한 연구의 다섯 국면**이다. 장마다 다른 연구를 예로 들면 그게 보이지 않는다.

추적 탈락을 다루는 5.4에서만 BRHS 로 부족해 **브라질 Pelotas 출생 코호트**를 함께 쓴다. 탈락률을 특성별로 쪼개어 보고한 드문 연구이기 때문이다.

5.2 표본 확보

BRHS 는 **7,735명**으로 출발했다. 코호트치고 특별히 큰 표본이 아니다.

코호트가 큰 표본을 요구하는 이유는 정밀도 때문만이 아니라 **기다려야 하기 때문**이다. 사건이 저절로 일어날 때까지 기다리는 설계이므로, 표본이 작으면 암처럼 비교적 흔한 결과조차 충분한 사례 수가 쌓이는 데 감당하기 어려운 시간이 걸린다.

> **혼동 주의** — "허용 오차를 절반으로 줄이려면 표본이 4배 필요하다"($n \propto 1/\varepsilon^2$)는 관계는 **단면조사에서 비율을 추정할 때**의 공식이고, 이 책에서는 **4장(서베이)** 에서 다룬다. 코호트의 표본 크기는 "얼마나 정밀하게 추정할까"가 아니라 **"목표한 상대위험도를 얼마의 검정력으로 잡아낼까"** 로 정해진다 — 5.6에서 다룬다.

5.3 측정

무작위 오차와 체계적 오차를 가른다.

	무작위 오차	체계적 오차(편향)
표본을 늘리면	줄어든다	**줄지 않는다**

	무작위 오차	체계적 오차(편향)
효과	정밀도 저하	추정치 자체가 치우침
대책	표본 크기, 반복 측정	표준화된 절차, 눈가림, 기기 교정

표본을 늘려도 편향은 사라지지 않는다 — 이 한 줄이 5.3의 전부다.

혈압 측정이 보여주는 것

혈압 측정은 간단해 보이지만 분해하면 오차가 끼어들 자리가 많다 — 커프 크기, 팔의 높이, 휴식 시간, 청진 시점, 그리고 **읽는 사람**.

측정 오차는 유형별로 나뉘는데, 그중 **관찰자 오차**가 이 장의 핵심이다. 관찰자 사이의 체계적 차이는 우연이 아니라 **편향**이고, 앞의 표대로 표본을 늘려도 사라지지 않는다.

대책은 세 층이다.

1. 절차 표준화와 측정자 훈련
2. 수행 점검 — 같은 대상을 두 사람이 재서 비교
3. **측정값을 측정자에게 숨기는 장비**

세 번째가 인상적이다. BRHS 당시 연구 현장에서는 측정이 끝날 때까지 수치를 조작자에게 보여주지 않는 혈압계가 흔히 쓰였다. **사람이 조심하는 것으로는 부족하다고 보고 장비로 차단한 것**이다. 이 발상이 7장 중재연구의 눈가림(blinding)으로 이어지고, 6장에서는 조사자 편향을 막는 절차로 다시 나온다.

5.4 추적

추적 탈락은 표본이 줄어드는 문제가 **아니다.** 방향을 가진 편향이다.

핵심 판별 질문 두 개:

1. 탈락이 **노출**과 상관 있는가?
2. 탈락이 **결과**와 상관 있는가?

둘 다 "아니오"라면 검정력만 잃는다. 하나라도 "예"라면 RR이 특정 방향으로 끌려간다. 어느 방향인지는 외우는 것이 아니라 2×2 표를 그려서 매번 따져야 한다.

숫자로 보기 — 같은 500명이 빠져도 결과가 다르다

BRHS 의 흡연과 허혈성심질환(IHD) 자료다 (Shaper et al. 1985).

흡연	사례	인원	위험도
비흡연	18	1,819	0.99%
현재 흡연	108	3,185	3.39%

> RR = 3.39 / 0.99 = **3.42**

이제 현재 흡연군에서 **500명이 추적에서 빠졌다**고 하자. 두 경우로 나눠 계산한다.

경우 A — 탈락이 결과와 무관할 때 (무작위 탈락)

빠진 500명 중 사례는 그 집단의 평균만큼, 약 17명이다.

```
흡연군 = (108 − 17) / (3,185 − 500) = 91 / 2,685 = 3.39%
RR = 3.39 / 0.99 = 3.42        ← 변하지 않는다
```

표본이 줄었으니 신뢰구간은 넓어진다. 그러나 **추정치 자체는 제자리다.**

경우 B — 탈락이 결과와 상관될 때 (차별 탈락)

빠진 500명이 이미 증상이 심한 사람들이었다면 그중 사례가 훨씬 많다. 30명이라고 하자.

```
흡연군 = (108 − 30) / (3,185 − 500) = 78 / 2,685 = 2.91%
RR = 2.91 / 0.99 = 2.94        ← 참값 3.42보다 14% 작다
```

노출군에서 일어날 사건을 놓쳤으므로 분자가 줄고, RR은 **1 쪽으로 끌려간다.**

> **이 두 경우의 차이가 5.4의 전부다.** 무작위 탈락은 정밀도만 갉아먹는다. 그러나 **노출·결과와 상관된 탈락은 방향을 가진 편향**이다. 그래서 논문에서 탈락률 숫자 하나만 보고 넘어가면 안 되고, **누가 빠졌는지**를 봐야 한다.

실제로 이걸 보고한 연구가 Pelotas 출생 코호트다. 원 코호트 5,249명 중 10–12세에 87.5%를 추적했는데, **추적률이 특성별로 달랐다.**

층화 변수	추적률 차이
성별	차이 없음 (p = 0.18)
출생체중	차이 없음 (p = 0.16)
가족 소득	**차이 있음 (p < 0.001)**
모의 교육	**차이 있음 (p < 0.001)**
임신 전 체질량지수	**차이 있음 (p = 0.004)**

"전체 87.5%"라는 한 숫자만으로는 이 차이가 보이지 않는다. 층화해서 보고했기에 드러난 것이다.

그리고 **방향이 직관과 반대다.** 빠진 쪽은 가난한 집이 아니라 부유한 집이었다 — 가족 소득이 최저임금 1배 이하인 집은 88.3%가 추적된 반면, 6.1–10배인 집은 79.9%에 그쳤다.

가족 소득 (최저임금 배수)	추적률
≤ 1	88.3%
3.1–6.0	88.9%
6.1–10.0	**79.9%**
> 10.0	82.6%

누가 빠졌는지는 넘겨짚을 수 없고 표를 봐야 안다. 이것이 5.4의 실무적 결론이다. 이 논문 전체는 워크숍에서 부검한다.

5.5 결과의 제시와 분석

5.5.1 발생률의 분모는 사람이 아니라 사람×시간이다

추적 기간이 사람마다 다르면 — 그리고 실제 코호트에서는 항상 다르다 — 분모는 인원수가 아니라 **인년(person-year)** 이다.

```
발생률 = 새로 발생한 사건 수 / 총 관찰 인년
```

500명을 5년 추적했다고 분모가 2,500인년인 것이 아니다. 중도 탈락자는 빠져나간 시점까지만, 사망자는 사망 시점까지만 센다. 이 계산을 손으로 한 번 해보지 않으면 반드시 틀린다.

5.5.2 상대위험도

```
RR = 노출군의 발생률 / 비노출군의 발생률
```

RR	의미
> 1.0	노출군의 위험이 더 높다
= 1.0	차이 없다
< 1.0	보호 효과

해석할 때 자주 틀리는 지점: RR = 0.68 은 "위험이 68% 낮다"가 아니라 **"32% 낮다"** 이다. 위험 감소분은 1 − RR 이다.

> 예: 격렬한 운동군의 암 발생률 5.7, 거의 운동하지 않는 군 8.4. RR = 5.7 / 8.4 = **0.68** → 운동군의 암 위험이 32% 낮다.

5.5.3 신뢰구간을 먼저 보고 p값을 나중에 본다

점추정치(RR = 2.8)만 보면 강한 효과처럼 보인다. 그러나 95% 신뢰구간이 0.9–8.7 이라면 그 연구는 **아무것도 배제하지 못했다.** 보호 효과부터 8배 위험까지 전부 자료와 양립한다.

"통계적으로 유의하지 않다"는 "효과가 없다"가 아니라 "판단할 근거가 부족하다"이다. 이 구분이 이 장에서 가장 자주 무너지는 곳이고, 수업 중 투표에서 그 지점을 다룬다.

5.5.4 어떤 검정을 쓰는가

계산은 11장에서 한다. 여기서는 **선택 기준**만 잡는다.

비교하려는 것	자료 형태	검정
노출 수준과 발병 여부의 연관	범주형 × 범주형	**카이제곱 검정**
두 독립 집단의 평균	연속형	**두 표본 t검정**

비교하려는 것	자료 형태	검정
세 집단 이상의 평균	연속형	분산분석 (11장)
두 연속형 변수의 관계	연속형 × 연속형	상관·단순선형회귀 (5.8)

t검정에는 조건이 붙는다 — 연속형 자료, 무작위 표본, **독립**(한 사람이 두 표본에 동시에 들어가지 않는다), 근사적 정규성, 그리고 두 집단의 표준편차가 비슷할 것. 흔히 쓰는 실무 기준은 **큰 쪽 SD ÷ 작은 쪽 SD < 2** 다.

> 이 조건 중 **독립**이 6장에서 깨진다. 짝지은 자료는 짝을 반영하는 검정을 써야 한다.

5.6 표본 크기

5.2와 중복이 아니다. 5.2는 *왜 커야 하는가*, 5.6은 *그래서 몇 명인가*를 다룬다.

코호트의 표본 크기는 다섯 가지를 정하면 계산된다.

정해야 할 것	예시값
유의수준 (양측)	0.05
검정력 (1 − β)	80%
비노출군 : 노출군 비	1 : 1
비노출군의 결과 발생률	5%
잡아내려는 최소 상대위험도	1.5

이 값으로 계산하면 **약 3,100명**이 필요하다.

핵심은 마지막 줄이다. **"얼마나 작은 효과까지 잡아낼 것인가"를 연구자가 먼저 정해야 한다.** 이것은 통계 문제가 아니라 보건학적 판단이고, 통계 자문에게 넘길 수 없는 부분이 정확히 여기다.

	실제로 효과가 없다	실제로 효과가 있다
효과가 있다고 결론	제1종 오류 (α)	옳음

	실제로 효과가 없다	실제로 효과가 있다
효과가 없다고 결론	옳음	제2종 오류 (β)

검정력 = 1 − β = **실제로 있는 효과를 놓치지 않을 확률.**

5.7 인과관계 평가

5.7.1 Bradford Hill 관점

1965년 Austin Bradford Hill 이 제시한 9가지. **체크리스트가 아니다.** 9개를 다 만족해야 인과인 것도 아니고, 하나씩 점수를 매기는 것도 아니다.

다만 **시간적 선후관계만은 다르다.** 이것은 지지 근거가 아니라 **논리적 필요조건**이다. 원인이 결과보다 나중에 있을 수는 없다. 단면연구 결과로 인과를 주장하는 논문이 왜 틀렸는지는 여기서 갈린다.

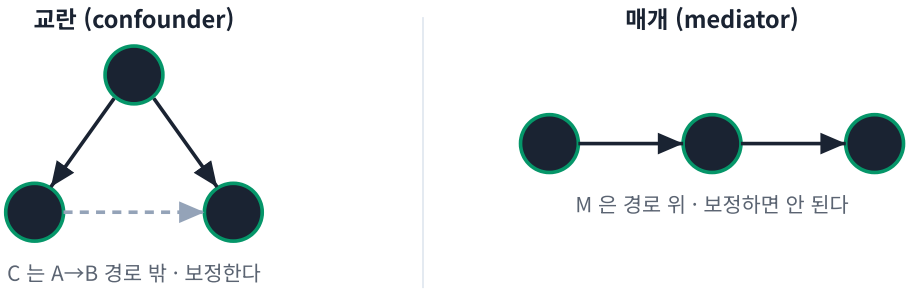
나머지 8개(연관성의 강도, 일관성, 특이성, 생물학적 경사, 실험적 증거, 개연성, 정합성, 비유)는 **있으면 인과를 지지하지만 없다고 인과가 부정되지 않는다.** 강도가 약해도 인과일 수 있고, 경사가 없어도 역치 효과일 수 있다.

5.7.2 교란

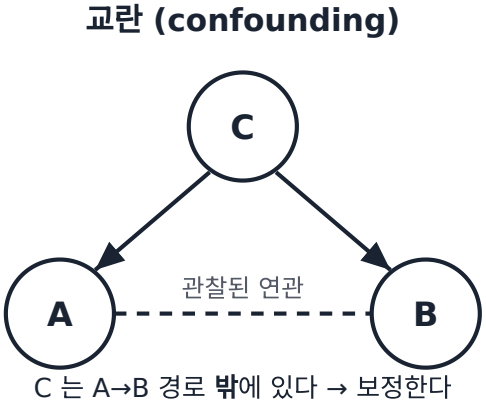
교란변수의 세 조건:

1. **노출과 연관되어 있다**
2. **결과의 예측 변수다**
3. **노출과 결과를 잇는 인과 경로 위에 있지 않다** ← 여기가 판별점

세 번째가 이 장에서 가장 중요한 한 줄이다. 경로 위에 있는 변수는 교란이 아니라 **매개변수 (mediator)** 이고, 매개변수를 보정하면 **노출의 효과를 스스로 지운다.**



흡연 → COHb 상승 → 저체중아 에서 COHb 는 매개다. COHb 를 보정해 흡연의 RR 이 1로 수렴했다면 그것은 "흡연은 무해하다"가 아니라 "흡연이 **어떤 경로로** 해를 끼치는지 보인 것"이다. 수업 중 투표가 정확히 이 함정을 겨눈다.




```
<line x1="330" y1="40" x2="330" y2="190" stroke-width="1" opacity=".3"/>

<!-- 오른쪽: 매개 -->
<text x="500" y="24" text-anchor="middle" font-size="15" font-weight="700" stroke="non
<circle cx="404" cy="115" r="26" fill="none" stroke-width="2"/>
<text x="404" y="121" text-anchor="middle" font-size="16" font-weight="700" stroke="no
<circle cx="500" cy="115" r="26" fill="none" stroke-width="2"/>
<text x="500" y="121" text-anchor="middle" font-size="16" font-weight="700" stroke="no
<circle cx="596" cy="115" r="26" fill="none" stroke-width="2"/>
<text x="596" y="121" text-anchor="middle" font-size="16" font-weight="700" stroke="no
<line x1="430" y1="115" x2="468" y2="115" stroke-width="2" marker-end="url(#ar)"/>
<line x1="526" y1="115" x2="564" y2="115" stroke-width="2" marker-end="url(#ar)"/>
<text x="500" y="200" text-anchor="middle" font-size="12.5" stroke="none">M 은 경로 <ts
```

세 번째 조건이 유일한 판별점이다. 앞의 두 조건(노출과 연관 · 결과의 예측변수)은 **교란과 매개가 똑같이 충족한다.** 그래서 눈으로는 구분되지 않고, 화살표를 그려야 갈린다.

5.8–5.9 회귀분석

회귀는 **여러 교란을 동시에 보정하는 도구**다. §5.7.2에서 교란을 하나씩 층화해 다루는 방법을 봤다면, 회귀는 그것을 여러 변수에 대해 한 번에 한다. 이 장의 층화·표준화와 같은 목적(교란 통제)을, 더 일반적인 형태로 하는 것이다.

계수를 어떻게 읽는가(β·OR·HR)는 **8장에서 Cox 회귀와 함께** 다룬다. 선형회귀·로지스틱 회귀·Cox 회귀가 exp(계수)라는 같은 골격을 공유하므로, 세 장에 흩지 않고 한자리에서 정리하는 편이 낫기 때

문이다.

정리 — 이 장에서 가져갈 다섯 줄

1. 코호트는 노출을 먼저 재고 결과를 기다린다. 그래서 발생률과 RR을 직접 구할 수 있다
2. 발생률의 분모는 사람 수가 아니라 **인년**이다
3. RR = 0.68 은 위험이 **32%** 낮다는 뜻이다
4. 추적 탈락은 표본 손실이 아니라 **방향을 가진 편향**이다
5. 인과 경로 위의 변수를 보정하면 **효과를 지운다.** 교란과 매개를 가르는 것이 3번 조건이다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 코호트 연구를 단면연구·환자-대조군 연구와 구별 짓는 핵심 특징은?

- A. 노출을 먼저 측정하고 이후 발생하는 결과를 추적한다
- B. 질병이 있는 집단과 없는 집단을 먼저 나누어 과거를 조사한다
- C. 한 시점에서 노출과 질병을 동시에 측정한다
- D. 연구자가 노출 여부를 무작위로 배정한다

2. 500명을 5년간 추적했으나 중도 탈락과 사망으로 총 관찰 기간은 2,000인년(person-year)이었다. 이 기간에 새로 발생한 심근경색은 40건이다. 발생률은?

- A. 연간 1,000명당 8.0건
- B. 연간 1,000명당 20.0건
- C. 연간 1,000명당 16.0건
- D. 연간 1,000명당 80.0건

3. 격렬한 운동군의 암 발생률이 인년당 5.7, 거의 운동하지 않는 군이 8.4였다. 운동군을 노출군으로 볼 때 상대위험도(RR)와 그 해석은?

- A. RR = 1.47 이며 운동군의 위험이 47% 높다
- B. RR = 2.70 이며 운동군의 위험이 170% 높다
- C. RR = 0.68 이며 운동군의 위험이 32% 낮다
- D. RR = 0.32 이며 운동군의 위험이 68% 낮다

4. Bradford Hill 의 인과성 판단 기준 중, 충족되지 않으면 인과관계를 주장할 수 없는 **필요조건**에 해당하는 것은?

- A. 연관성의 강도 — 상대위험도가 충분히 커야 한다
- B. 생물학적 경사 — 노출이 클수록 위험이 커져야 한다
- C. 일관성 — 여러 연구에서 같은 결과가 나와야 한다
- D. 시간적 선후관계 — 원인이 결과보다 먼저여야 한다

5. 어떤 변수가 교란변수(confounder)로 인정되기 위한 세 조건에 **해당하지 않는** 것은?

- A. 노출 요인과 통계적으로 연관되어 있어야 한다
- B. 노출로 인해 새로 생긴 결과가 아니어야 한다
- C. 결과의 독립적인 예측 변수여야 한다
- D. 노출과 결과의 인과 경로 위에 있어야 한다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 코호트의 뼈대** — 코호트 연구에서는 왜 사람 수가 아니라 '인년(person-year)'으로 나누어 발생률을 구할까요? 사람마다 추적 기간이 다르면 무슨 일이 생기나요?
2. **2단계 · 추적과 편향** — 5년 추적 중 20%가 연락이 끊겼습니다. 그냥 표본이 줄어든 것뿐일까요, 아니면 결과가 특정 방향으로 틀어질 수도 있을까요?
3. **3단계 · 연관에서 인과로** — 흡연과 폐암의 연관을 관찰했습니다. 이것을 '인과'라고 부르려면 무엇이 더 있어야 할까요? 그리고 보정은 많이 할수록 좋은 걸까요?

제6장 · 환자-대조군 연구

제6장 환자-대조군 연구

> **5장을 먼저 읽으셔야 합니다.** 이 장은 코호트와의 대비로만 이해됩니다.

이 장에서 답할 질문

- 결과에서 거꾸로 출발하는데도 왜 인과를 논할 수 있는가
- 대조군은 어디서 뽑아야 하는가. 그 기준은 왜 "같은 병원"이 아닌가
- 오즈비를 "위험이 N배"라고 옮겨도 되는 때는 언제인가

> **이 장을 관통하는 물음** — 이 설계의 성패는 한 가지에 달렸다: **대조군이 사례가 나온 그 인구 집단을 대표하는가.** 대조군을 엉뚱한 곳에서 뽑으면 오즈비는 커 보이든 작아 보이든 그 인구의 현실을 담지 못한다. 절차가 현실을 비추지 못하는 전형이다.

6.1 왜 환자-대조군 연구를 하는가

코호트가 노출에서 출발한다면, 환자-대조군은 **결과에서 출발해 거꾸로 노출을 본다.**

	코호트	환자-대조군
출발점	노출	**결과**
얻는 지표	발생률, **RR**	**OR** (발생률 없음)
드문 질병	매우 비효율	**최적**
여러 결과	가능	불가 (결과가 고정)
회상 편향	자유로움	**취약**

	코호트	환자-대조군
비용·기간	크다	작다

이 표의 두 번째 줄이 이 장 전체를 지배한다. **분모(위험에 처한 사람과 시간)를 모르므로 발생률을 구할 수 없고, 따라서 RR도 구할 수 없다.** 대신 쓰는 것이 오즈비다.

이 장의 관통 사례 — 뉴질랜드 사산 연구

오클랜드 지역의 후기 사산(late stillbirth) 환자-대조군 연구를 끝까지 따라간다.

- 3년에 걸쳐 후기 사산을 겪고 참여에 동의한 여성 **155명**을 사례로 모았다
- 임신 중 **수면 자세**가 사산 위험과 관련되는지를 물었다
- 사례 1명에 대조군 1명을 짝지어 **180쌍**으로 분석했다

이 연구를 고른 이유는 **6장의 논점이 전부 들어 있기** 때문이다 — 드문 결과, 후향적 노출 수집, 대조군 선정의 어려움, 짝짓기와 그 분석, 그리고 회상 편향 논쟁.

마지막이 특히 값지다. 이 논문이 나온 뒤 **덴마크 전국 출생 코호트** 연구진이 "우리 자료에서는 수면 자세와 사산의 연관이 보이지 않는다, 뉴질랜드 결과는 회상 편향일 수 있다"고 반박했다(Stacey et al. 2011). 덴마크 연구가 그 반박을 할 수 있었던 근거는 하나였다 — **자기네 자료는 전향적이라 회상 편향에서 자유롭다**는 것.

5장과 6장의 차이가 교과서 속 표가 아니라 실제 논쟁으로 드러난 사례다. 6장을 읽는 내내 이 반박을 염두에 두십시오.

6.2 설계의 핵심 요소

6.2.1 사례 선정

사례 정의를 먼저 못 박는다. 정의가 흔들리면 뒤의 모든 것이 흔들린다.

6.2.2 대조군 선정 — 이 장에서 가장 자주 틀리는 곳

원칙은 하나다. **사례가 발생한 것과 같은 모집단에서 뽑는다.**

판별 질문도 하나다:

> **"이 대조군이 그 병에 걸렸다면, 이 연구의 사례로 잡혔을까?"**

"아니오"라면 잘못된 대조군이다. "같은 병원에서 뽑았으니 조건이 같다"는 흔한 오해인데, 기준은 병원이 아니라 모집단이다. 병원 대조군은 입원 사유가 노출과 얽혀 있을 때 특히 위험하다 (수업 중 투표에서 다루는 함정이다).

6.2.3 노출 평가와 편향

편향	무엇	어디서 생기나
회상 편향	사례·대조군의 기억 정확도가 다르다	대상자
조사자 편향	조사자가 사례 여부를 알고 캔다	조사자
기록 편향	두 군의 기록 품질이 다르다	자료원

회상 편향은 **방향이 정해져 있지 않다.** 환자가 더 잘 기억해 OR이 부풀 수도 있지만, 노출이 낙인이 되는 주제(음주·약물)에서는 사례군이 오히려 축소 보고해 반대로 간다. **편향의 이름을 대는 것으로 끝내지 말고 매번 방향을 따져야 한다.**

눈가림(blinding)으로 막을 수 있는 것

셋 중 조사자 편향은 절차로 줄일 수 있다.

1. **조사자가 대상자의 사례·대조군 여부를 모르게 한다.** 원칙은 이것이지만 현실에서는 어렵거나 불가능한 경우가 많다 — 사산을 겪은 여성과 그렇지 않은 여성은 면담 몇 분이면 드러난다
2. **조사자와 대상자 모두가 연구 가설을 모르게 한다.** 적어도 여러 위험요인 중 무엇이 주 가설인지 모르게 한다. 이건 대개 가능하다

두 번째가 현실적인 방어선이다. 관통 사례가 실제로 그렇게 했다.

> *"Recall bias was reduced as far as possible in this study by using a **structured interview** and by ensuring that **participants were not aware of the study hypotheses being tested**."*

사례 여부는 숨길 수 없었지만 **가설은 숨겼다.** 완전한 눈가림이 안 되는 설계에서 무엇을 지킬 수 있는지를 보여주는 예다. **7장 중재연구**에서 눈가림이 왜 설계의 심장이 되는지, 그리고 관찰연구에서는 왜 그만큼 쓰기 어려운지를 여기서 미리 볼 수 있다.

6.3 분석 — 오즈비

6.3.1 오즈와 오즈비

오즈 = 노출될 확률 / 노출되지 않을 확률 = `p / (1−p)` 확률 0.3 이면 오즈는 0.3/0.7 ≈ 0.43 이다. 확률과 오즈는 다른 수다.

오즈비 = 사례군의 노출 오즈 / 대조군의 노출 오즈. 2×2 표에서는 교차곱으로 바로 나온다.

```
            노출 O    노출 X
사례          a         b          OR = (a×d) / (b×c)
대조          c         d
```

6.3.2 ★ 오즈비를 위험비로 읽어도 되는 때

결과가 드물 때만이다. 대략 유병률 10% 미만이 기준이다.

흔한 결과에서는 OR이 RR을 **과장한다.**

```
RR ≈ OR / (1 − p₀ + p₀·OR)
```

유병률 **40%** 인 결과에서 OR = 4.0 이면

```
RR ≈ 4.0 / (1 − 0.40 + 0.40 × 4.0) = 4.0 / 2.2 ≈ 1.8
```

"4배"가 아니라 "약 1.8배"다. 언론 보도와 초록이 가장 흔하게 무너지는 지점이다. **유병률이 낮을수록 OR과 RR이 가까워진다** — 직접 다른 값을 넣어 확인해 보십시오.

> **이 장의 관통 사례는 이 조건을 만족한다.** 뉴질랜드 연구가 보고한 후기 사산 유병률은 **1,000분만 3.09건(0.31%)** 이다. 이렇게 드문 결과에서는 OR을 위험비처럼 읽어도 무방하다. 조건을 따져 보고 쓰는 것과, 따지지 않고 쓰는 것은 다르다.

6.3.3 짝짓기와 그 분석

짝짓기는 설계 단계의 교란 통제 수단이다. 두 가지를 반드시 지킨다.

1. **짝지었으면 짝을 반영해 분석한다.** McNemar 검정, 조건부 로지스틱 회귀. 짝을 무시하고 단순 2×2로 분석하면 추정치가 **OR = 1 쪽으로 끌려가 효과를 과소추정한다**
2. **노출과 밀접한 변수로는 짝짓지 않는다.** 과잉짝짓기다. 교란을 통제한 것이 아니라 보려던 노출 차이를 지운 것이 된다

짝지은 자료에서 정보를 주는 것은 **불일치 짝(discordant pair)뿐**이다. 둘 다 노출됐거나 둘 다 노출되지 않은 짝은 OR 추정에 기여하지 않는다. 수업 중 투표가 이 지점을 겨눈다.

짝 단위로 보면 표가 달라진다

아래는 **짝짓기 분석을 설명하기 위한 1:1 예시 자료**다. 뉴질랜드 연구의 실제 표가 아니다 — 그 연구는 사례 1명에 대조군 2명(1:2)이라 2×2 표로 요약되지 않는다. 개념을 먼저 1:1로 익힌 뒤 1:2로 확장하는 것이 교재의 순서다.

사람이 아니라 짝이 단위라는 점이 핵심이다.

	대조군 노출	대조군 비노출	계
사례 노출	12	**42**	54
사례 비노출	**7**	119	126
계	19	161	180

굵게 표시한 두 칸이 **불일치 짝(discordant pair)** 이다 — 짝 안에서 노출이 갈린 경우.

```
짝짓기 OR = 42 / 7 = 6.00
```

나머지 12 + 119 = **131쌍, 전체의 73%는 계산에 들어가지 않는다.** 둘 다 노출됐거나 둘 다 노출되지 않은 짝은 "노출이 결과를 가르는가"에 아무 말도 하지 못하기 때문이다.

검정도 짝 단위로 한다 — **McNemar 검정**.

```
χ² = (42 − 7)² / (42 + 7) = 1,225 / 49 = 25.00      p < 0.0005
```

짝을 무시하면 무슨 일이 생기나

같은 180쌍을 짝에서 풀어 사람 단위로 흩뜨리면 이렇게 된다.

	사례	대조군
노출	54	19
비노출	126	161
계	180	180

```
짝 무시 OR = (54 × 161) / (126 × 19) = 8,694 / 2,394 = 3.63
```

6.00 이 3.63 으로 내려앉았다. 자료는 한 건도 바뀌지 않았고 분석 방법만 바뀌었다.

> 짝짓기를 무시하면 오즈비 추정치가 **귀무가설(OR=1) 쪽으로 끌려가 효과를 희석한다.**
>
> **방향이 정해져 있다는 것이 실무적으로 중요하다.** 논문에서 "짝지어 모집했다"는 문장을 봤는데 분석이 단순 2×2 라면, 보고된 OR은 **참값보다 작다**고 읽어야 한다.

실제 논문은 어떻게 했나

뉴질랜드 연구는 사례 155명에 재태주수를 맞춘 대조군 310명(1:2)을 짝지었고, 분석에서 짝을 그대로 반영했다. 논문의 표현은 이렇다.

> *"We used standard **conditional** regressions for matched case–control studies using the 'proc logistic' procedure with the **'strata' statement to control for matching**."*

1:2 이상에서는 2×2 표로 요약할 수 없고 조건부 로지스틱 회귀로 넘어간다(6.3.4). **설계에서 짝지었으면 분석에서도 짝을 유지한다** — 규칙은 1:1이든 1:2든 같다.

보정 후 결과는 이렇다 (마지막 밤 수면자세, 왼쪽 옆으로 자는 것을 기준으로).

수면 자세	보정 오즈비 (95% CI)
왼쪽 옆	1.00 (기준)

수면 자세	보정 오즈비 (95% CI)
오른쪽 옆	1.74 (0.98 – 3.01)
바로 누움	**2.54 (1.04 - 6.18)**
그 외	2.32 (1.28 – 4.19)

> 오른쪽 옆의 신뢰구간이 **0.98-3.01 로 아슬아슬하게 1을 포함한다.** "유의하지 않다"와 "효과가 없다"는 다른 말이라는 것을 실제 논문에서 볼 수 있는 자리다.

6.3.4 로지스틱 회귀

여러 교란을 동시에 보정해야 할 때 쓴다. 여기서는 **어느 것을 쓰는가**만 가른다.

	언제 쓰나	짝을 반영하나
비조건부(unconditional) 로지스틱 회귀	짝짓기를 하지 않은 설계	아니오
조건부(conditional) 로지스틱 회귀	**짝짓기를 한 설계**	예

규칙은 앞 절과 똑같다 — **짝지었으면 조건부로 간다.** 뉴질랜드 연구도 단순 짝짓기 분석을 한 뒤 조건부 로지스틱 회귀로 다변량 보정을 했다.

계수 해석(β·OR·HR을 어떻게 읽는가)은 **8장에서 Cox 회귀와 함께** 다룬다.

정리 — 이 장에서 가져갈 다섯 줄

1. 결과에서 출발하므로 발생률을 모른다. 그래서 RR이 아니라 **OR**이다
2. 대조군의 기준은 병원이 아니라 **모집단**이다 — "그 병에 걸렸다면 사례로 잡혔을까?"
3. 회상 편향은 **방향이 정해져 있지 않다.** 매번 따진다
4. OR ≈ RR 은 **결과가 드물 때만** 성립한다. 흔하면 OR이 과장한다
5. **짝지었으면 짝을 반영해 분석한다.** 무시하면 효과를 과소추정한다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 환자-대조군 연구가 코호트 연구보다 유리한 상황은?

- A. 질병이 매우 드물고 잠복기가 길 때
- B. 한 노출의 여러 결과를 함께 볼 때
- C. 노출의 발생률을 직접 구해야 할 때
- D. 노출을 무작위로 배정할 수 있을 때

2. 대조군 선정에서 가장 중요한 원칙은?

- A. 사례군과 성별·연령 분포가 같아야 한다
- B. 사례와 같은 모집단에서 뽑아야 한다
- C. 사례군과 같은 병원에 입원해 있어야 한다
- D. 사례군보다 인원이 반드시 많아야 한다

3. 폐암 환자 200명 중 흡연자가 160명, 대조군 200명 중 흡연자가 80명이었다. 오즈비는?

- A. 2.0
- B. 4.0
- C. 6.0
- D. 0.17

4. 회상 편향(recall bias)이 환자-대조군 연구에서 특히 문제가 되는 이유는?

- A. 조사자가 대상자의 사례 여부를 알기 때문이다
- B. 대조군을 병원에서 모집하는 경우가 많기 때문이다
- C. 사례 수가 대조군 수보다 대체로 적기 때문이다
- D. 노출을 과거의 기억에 의존해 수집하기 때문이다

5. 과잉짝짓기(over-matching)가 문제인 이유는?

- A. 짝짓기 변수가 늘수록 대조군 모집이 어려워지기 때문이다
- B. 짝지은 변수의 효과를 따로 추정할 수 없게 되기 때문이다

- C. 노출과 밀접한 변수로 짝지으면 노출 차이가 지워지기 때문이다
- D. 짝짓기 분석은 계산이 복잡해 오류가 생기기 쉽기 때문이다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 왜 오즈비인가** — 환자-대조군 연구에서는 왜 상대위험도(RR)를 직접 구하지 못하고 오즈비(OR)를 쓸까요? 그리고 OR을 '위험이 N배'라고 읽어도 되는 경우는 언제인가요?
2. **2단계 · 대조군과 편향** — 환자-대조군 연구의 성패는 '대조군을 어디서 뽑느냐'에 달렸다고 합니다. 병원 다른 과 입원환자를 대조군으로 쓰면 무엇이 문제가 될까요?

제7장 · 중재연구

제7장 중재연구

> **5장과 6장을 먼저 읽으셔야 합니다.** 이 장은 관찰연구가 못 하는 일을 다룹니다.

이 장에서 답할 질문

- 코호트로 충분하지 않은 때는 언제인가. 왜 굳이 사람에게 개입하는가
- 배정대로 분석하면 효과가 희석되는데, 왜 그래도 그렇게 하는가
- "위험을 절반으로 줄였다"는 문장은 무엇을 감추는가

7.1 왜 중재연구를 하는가

5장과 6장은 **일어난 일을 관찰**했다. 중재연구는 **연구자가 노출을 배정**한다. 이 한 걸음이 바꾸는 것이 있다.

관찰연구에서는 아무리 보정해도 **모르는 교란요인**이 남는다. ch5 §5.7.2에서 본 것처럼 보정은 아는 변수에만 할 수 있다. 무작위배정은 **알려지지 않은 교란요인까지** 평균적으로 두 군에 고르게 흩어 놓는다. 관찰연구가 원리적으로 할 수 없는 일이다.

대가는 크다 — 윤리적 제약, 비용, 그리고 **연구실에서 통한 것이 현장에서도 통하는가**라는 물음.

이 장의 관통 사례 — 니코틴 흡입기 시험

Bollinger et al. (2000), 금연을 못 하는 흡연자에게 경구 니코틴 흡입기를 준 **이중맹검 무작위배정 시험**이다.

금연이 최선이지만 많은 흡연자가 실패한다. 그렇다면 **줄이는 것**이라도 도움이 되는가 — 관찰만으로는 답할 수 없는 질문이다. 흡입기를 스스로 선택한 사람은 애초에 동기가 다르기 때문이다.

7.5의 군집 설계에서는 **Elley et al. (2003)** 의 '녹색 처방(green prescription)' 군집 무작위배정 시험을 함께 본다 (Paper C).

7.2 설계의 핵심 요소

무작위배정과 눈가림은 다른 장치다

시간축에 놓으면 자리가 분명해진다.

	무엇을 막나	없으면
무작위배정	배정 시점에 두 군이 달라지는 것	교란. 관찰연구와 같아진다
눈가림	배정을 안 뒤 행동·측정이 달라지는 것	참여자의 행동 변화, 측정자의 판정 편향

하나가 다른 하나를 대신하지 못한다. 무작위배정을 완벽히 해도 눈가림이 없으면 결과 판정이 흔들린다. ch6 §6.2.3에서 본 조사자 편향과 같은 기제다.

관통 사례가 **이중맹검**인 것은 그래서다 — 참여자도 연구진도 누가 활성 흡입기를 받았는지 모른다.

위약(placebo)은 두 가지 일을 동시에 한다. 첫째, **눈가림을 가능하게 한다** — 겉모습·맛·복용법이 활성 치료와 똑같아야 참여자도 연구진도 누가 진짜를 받았는지 모른다. 둘째, **"무언가를 받았다"는 사실 자체가 내는 효과(위약효과)를 두 군에서 똑같이 발생시켜 상쇄한다.** 그래서 두 군의 차이에는 활성 성분의 순효과만 남는다. 관통 사례가 위약 흡입기를 쓴 것이 이 때문이다.

7.3 결과의 분석

7.3.4 ★ 배정대로 분석 (intention-to-treat)

중재군에 배정된 사람 중 일부가 흡입기를 그만둔다. 그들을 어떻게 할 것인가.

언뜻 더 효율적으로 보이는 두 방법이 있다.

1. 배정대로 하지 않은 사람을 **분석에서 제외**한다
2. 중재군에서 그만둔 사람을 **대조군에 합쳐** 분석한다

둘 다 편향을 만든다. 이유는 하나다.

> **그만둔 사람은 끝까지 한 사람을 대표하지 않는다.**

대개 더 힘들었거나, 부작용이 있었거나, 애초에 동기가 약했다. 그들을 빼거나 옮기는 순간 두 군은 **더 이상 교란요인이 균형 잡혀 있지 않다** — 무작위배정을 한 이유가 통째로 사라진다.

그래서 **처음 배정된 군에 그대로 두고 분석한다.** 이것이 배정대로 분석이다.

희석과 편향은 다른 문제다

배정대로 분석에는 대가가 있다. 순응하지 않은 사람을 원래 군에 두므로 **효과가 희석된다.**

핵심은 이렇게 요약된다.

> 검출되는 효과는 참값의 **과소추정**일 뿐이며, 순응하지 않은 사람 때문에 **편향되지는 않는다.**

과소추정이지 편향이 아니다. 이 구분이 이 절의 전부다. per-protocol 분석은 반대로 **편향된다** — 값은 커지지만 어느 방향으로 얼마나 틀렸는지 알 수 없다.

> 덤으로 얻는 것이 있다. 현실에서도 사람들은 중도에 그만둔다. 그래서 배정대로 분석의 결과는 **실제 진료에서 이 치료가 낼 효과**에 더 가깝다. 이 지점은 수업 중 투표에서 다시 다룬다.

7.4 효과의 크기를 어떻게 말하는가

같은 결과를 세 가지로 말할 수 있고, 셋이 전혀 다른 인상을 준다.

지표	계산	갑 (8.0% → 4.0%)	을 (0.4% → 0.2%)
상대위험감소	1 − RR	**50%**	**50%**
절대위험감소	대조군 − 중재군	4.0%p	0.2%p
치료필요수 (NNT)	1 / 절대위험감소	**25명**	**500명**

상대위험감소는 두 사업에서 똑같다. 그런데 한 건을 막는 데 필요한 사람은 25명과 500명, 스무 배 차이다.

> **상대 지표는 기저위험을 지운다.** 지워진 그 정보가 정책 결정의 핵심일 때가 많다. "위험을 절반으로 줄였다"는 문장은 참이면서 동시에 아무것도 알려주지 않을 수 있다. 이 계산은 수업 중 투표에서 직접 해 본다.

7.5 군집 무작위배정

왜 집단째 배정하는가 — 오염

개인에게만 닿는 중재(알약)라면 개인 단위로 배정하면 된다. 그러나 상담·정보·게시물처럼 **집단에 닿는 중재**는 다르다.

같은 의원에 다니는 환자끼리 이야기가 오간다. 의료진이 두 군을 다르게 대하기도 어렵다. 대조군이 중재의 영향을 받는 것을 **오염(contamination)** 이라 한다. 오염이 있으면 두 군의 차이가 줄어 효과를 놓친다.

그래서 **의원째, 학교째** 배정한다. Elley 의 녹색 처방 시험이 그렇게 했다.

★ 군집을 무시하면 무엇이 망가지나

여기서 **5장·6장과 패턴이 달라진다.** 주의해서 읽으십시오.

같은 의원에 다니는 환자들은 서로 닮아 있다 — 진료 방식, 지역, 사회경제적 배경. 군집 안의 값이 서로 **상관되어 있으므로**, 1,000명이 갖는 실제 정보량은 1,000명분보다 적다.

이것을 무시하고 개인을 독립 관측치로 놓으면 어떻게 되는가.

> 이때 **표준오차가 과소추정**된다. 같은 말을 달리 하면 **정밀도가 과대평가**되고, 신뢰구간이 **인위적으로 좁아지며**, p값이 **인위적으로 작아진다.**

	ch5 차별 탈락	ch6 짝 무시	ch7 군집 무시
망가지는 것	**추정치**	**추정치**	**정밀도**
어떻게	1 쪽으로 끌려감	1 쪽으로 끌려감	신뢰구간이 좁아짐

점추정치는 대체로 자리를 지킨다. 대신 없는 확신이 생긴다. "설계를 무시하면 1 쪽으로"라는 패턴을 외운 학생은 여기서 틀린다. 수업 중 투표가 그 자리다.

군집이 정보를 얼마나 깎는지는 **설계효과(design effect, DE)** 로 나타낸다.

```
DE = 1 + (m − 1) × ICC
```

`m`은 군집당 평균 인원, `ICC`(급내상관계수)는 같은 군집 안의 값들이 얼마나 닮았는가를 0~1로 나타낸 값이다. 예를 들어 군집당 20명(m=20), ICC=0.02이면 DE = 1 + 19×0.02 ≈ **1.38**. 개인 무작위배정과 같은 정밀도를 얻으려면 표본을 약 **38% 더** 모아야 한다는 뜻이다. 군집이 클수록, 군집 안이 닮을수록(ICC가 클수록) 손해가 커진다. 분석에서 군집을 무시하면 이 DE를 1로 친 셈이 되어 신뢰구간이 실제보다 좁아진다.

7.6 등록과 보고

시험은 시작 **전에** 주 결과변수와 분석 계획을 공개 등록부(예: ClinicalTrials.gov)에 못 박아 둔다. 그래야 결과를 본 뒤 유리한 지표만 골라 "주 결과"로 바꾸는 **사후 취사선택**을 막을 수 있다. 보고는 **CONSORT** 지침을 따라 배정·탈락·분석에 들어간 인원을 흐름도로 투명하게 밝힌다.

> 이 사전등록의 논리가 9장 체계적 문헌고찰의 **출판 편향** 대책과 곧장 이어진다. 유의한 결과만 세상에 나오고 나머지는 서랍에 묻히는 것을 막으려는 **같은 정신**이다.

정리 — 이 장에서 가져갈 다섯 줄

1. 무작위배정만이 **모르는 교란요인**까지 균형 맞춘다. 관찰연구가 못 하는 일이다
2. 무작위배정은 **배정 시점**, 눈가림은 **배정 이후**. 서로를 대신하지 못한다
3. **처음 배정된 군 그대로** 분석한다. 이탈자를 빼면 무작위배정이 무의미해진다
4. 배정대로 분석은 효과를 **희석**시키지만 **편향시키지는 않는다.** 둘은 다른 문제다
5. 군집을 무시하면 **추정치가 아니라 신뢰구간**이 망가진다. 앞 장들과 방향이 다르다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 무작위배정(randomisation)이 하는 일은?

- A. 두 군의 교란요인을 균형 맞춘다
- B. 측정자가 배정 결과를 모르게 한다
- C. 참여자의 중도 이탈을 줄여 준다
- D. 표본 크기를 줄여 비용을 아낀다

2. 무작위배정을 했는데도 눈가림(blinding)이 따로 필요한 이유는?

- A. 무작위배정만으로는 표본 크기가 부족해지기 때문이다
- B. 배정 이후의 측정과 행동이 왜곡될 수 있기 때문이다
- C. 무작위배정은 교란요인을 균형 맞추지 못하기 때문이다
- D. 눈가림이 있으면 중도 이탈자가 생기지 않기 때문이다

3. 배정대로 분석(intention-to-treat)이란?

- A. 중재를 실제로 받은 사람만 골라 분석하는 것
- B. 순응하지 않은 사람을 반대 군으로 옮겨 분석하는 것
- C. 중도 이탈자를 제외하고 남은 사람만 분석하는 것
- D. 처음 배정된 군에 그대로 두고 분석하는 것

4. 학교나 의원처럼 **집단 단위로** 무작위배정을 하는 주된 이유는?

- A. 집단 단위가 표본 크기를 줄여 주기 때문이다

- B. 같은 집단 안에서 중재가 대조군에 새기 때문이다
- C. 집단 단위로 배정하면 눈가림이 쉬워지기 때문이다
- D. 개인 배정은 윤리적으로 허용되지 않기 때문이다

5. 어떤 중재가 사건 발생을 대조군 4.0%에서 중재군 2.0%로 낮췄다. 치료필요수(NNT)는?

- A. 2
- B. 25
- C. 50
- D. 100

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 배정을 지키는 이유** — 중재를 끝까지 잘 따른 사람만 분석하면 효과를 더 정확히 잴 수 있을 것 같은데, 왜 '배정된 대로'(중간에 그만둔 사람도 포함해) 분석하라고 할까요?
2. **2단계 · 효과의 크기와 군집** — '위험이 50% 감소했다'는 결과를 봤습니다. 이것만으로 큰 효과라고 할 수 있을까요? 무엇을 더 물어야 할까요?

제8장 · 생존분석과 Cox 회귀

제8장 생존분석과 Cox 회귀

> **이 주에는 논문 워크숍이 없습니다.** 대신 계산과 그래프 읽기에 시간을 씁니다.

이 장에서 답할 질문

- 추적이 끝나기 전에 빠진 사람의 자료를 어떻게 쓰는가
- **"위험이 2배"** 라는 말이 코호트와 생존분석에서 왜 다른 뜻인가
- 두 생존곡선이 교차하면 무엇이 문제인가

8.1 왜 시간이 필요한가

5장에서 하던 방식의 한계

5장에서는 **인년(person-year)** 으로 발생률을 구했습니다. 그것도 시간을 쓰는 방법이지만, 답할 수 없는 질문이 있습니다.

> **"3년 뒤 생존 확률은 얼마입니까?"**

발생률은 **평균 속도**를 알려주지 시점별 확률을 주지 않습니다. 생존분석은 **시간에 따라 변하는 확률**을 그립니다.

★ 중도절단 — 이 장의 출발점

추적이 끝나기 전에 관찰이 끊기는 것을 **중도절단(censoring)** 이라 합니다.

절단되는 경우	절단이 아닌 경우
연구 종료 시점까지 사건 없음	사건이 일어남
이사·연락두절로 추적 실패	시점이 기록되지 않음(자료 결측)

절단된 사람도 **절단 시점까지는 사건이 없었다**는 정보를 줍니다. 생존분석은 그 정보를 버리지 않습니다.

절단에 딸린 가정

> **절단된 사람은 그 시점 이후**
> **남은 사람과 같은 위험을 가진다**

이 가정이 깨지면 결과가 틀립니다.

- 연구가 끝나서 절단 → 가정이 대체로 성립
- **부작용으로 치료를 중단하고 연락두절 → 가정이 깨짐**

후자에서 빠진 사람은 더 아픈 사람이고, 그들이 빠지면 남은 집단이 실제보다 건강해 보입니다. **생존율이 과대추정됩니다.**

> **5장의 차별 탈락과 같은 논리입니다.** 그때도 "누가 빠졌는가"가 답을 갈랐습니다. 이 지점은 수업 중 투표에서 다시 다룹니다.

이 장의 관통 사례

> **Whincup PH, Gilg JA, Emberson JR, Jarvis MJ, Feyerabend CF, et al. (2004).** *Passive smoking and risk of coronary heart disease and stroke: prospective study with cotinine measurement.* **BMJ 329, 200–205.**

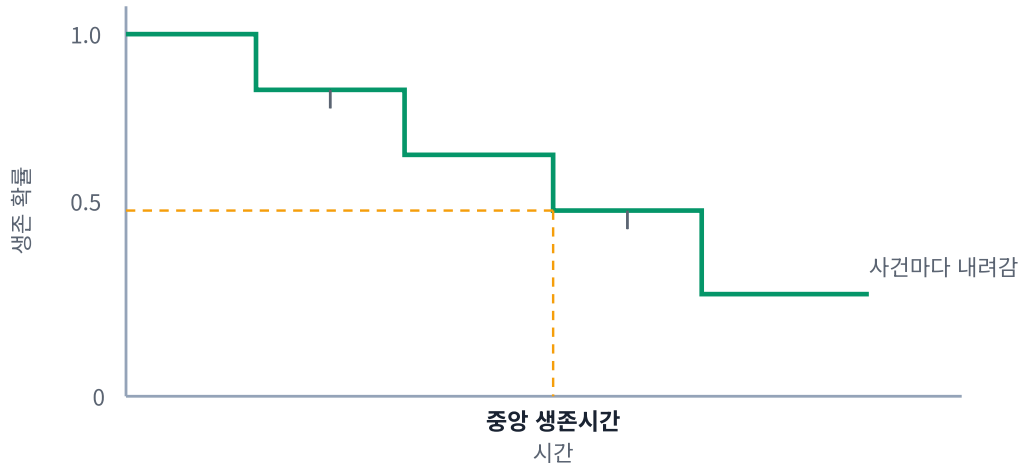
5장에서 본 영국 지역 심장 연구(BRHS) 입니다. 같은 코호트를 생존분석으로 다시 봅니다. 간접흡연 노출을 **코티닌(니코틴 대사물) 농도**로 객관적으로 재었다는 점이 강점입니다.

5장은 이 코호트를 **인년당 발생률**, 곧 평균 속도 하나로 요약했습니다. 8장은 같은 자료를 **시간에 따라 변하는 생존 확률**로 봅니다. 발생률은 "평균적으로 얼마나 빨리 생기나"를 한 숫자로 주지만, 생존곡선은 "3년 뒤·5년 뒤 각각 얼마가 살아 있나"와 "두 군의 차이가 시간에 따라 어떻게 벌어지나"를 보여줍니다. 간접흡연처럼 효과가 서서히 쌓이는 노출에서는 이 시간 축이 특히 중요합니다.

8.2 Kaplan-Meier 곡선

계단으로 내려간다

사건이 일어날 때마다 생존 확률을 다시 계산해 **계단 모양**으로 그립니다.



절단된 사람은 곡선을 내리지 않습니다. 다만 그 시점부터 분모에서 빠집니다. 그래서 곡선에 작은 눈금(tick)으로 표시합니다.

⚠ 세로축을 먼저 읽으십시오

세로축	시작	방향	0.20이 뜻하는 것
생존 확률	1.0	내려감	20%가 살아 있다
사건 발생 확률	0.0	올라감	**20%가 사건을 겪었다**

> **바꿔 읽으면 결론이 정반대가 됩니다**

논문마다 다르게 그리므로 **축 이름을 먼저 확인하는 습관**이 필요합니다.

중앙 생존시간

생존 확률이 **0.5가 되는 시점**입니다. 절반이 사건을 겪는 데 걸린 시간입니다.

> **곡선이 0.5까지 내려가지 않으면 중앙 생존시간을 구할 수 없습니다.** 추적이 짧거나 예후가 좋은 경우이고, 그럴 때는 "중앙 생존시간 미도달"이라고 씁니다.

두 곡선을 비교하는 검정 — 로그순위검정

두 군의 생존곡선이 같은지 검정합니다. 원리는 5장의 카이제곱과 같습니다 — **각 시점에서 관찰된 사건 수와 기대되는 사건 수를 비교**해 전체를 합칩니다.

작은 예로 원리만 봅니다. A·B 두 군이 각각 10명으로 시작합니다.

사건 시점	그 직전 위험군 (A/B/합)	그 시점 사건	A군 관찰 (O)	A군 기대(E) = 사건×A위험/합	O−E
t_1	10 / 10 / 20	1 (A에서)	1	$1 \times 10/20 = 0.50$	+0.50
t_2	9 / 10 / 19	1 (B에서)	0	$1 \times 9/19 \approx 0.47$	−0.47

각 사건 시점마다 "이 시점에 A에서 몇 건이 기대되는가(E)"를 위험군 비율로 계산하고, 실제 관찰(O)과의 차이를 쌓습니다. 여기서는 $\Sigma O - \Sigma E \approx 0.50 - 0.47 \approx 0$ 이라 두 군의 차이가 거의 없습니다. 실제 로그순위검정은 이 $\Sigma O - \Sigma E$ 를 그 분산으로 나눠 카이제곱 통계량을 만듭니다. 핵심은 **매 사건 시점의 2×2 표를 만들어 관찰과 기대의 차이를 합친다**는 것입니다 — 3장·11장의 카이제곱과 같은 정신입니다.

8.3 Cox 회귀

무엇을 하는가

여러 변수를 **동시에 보정하면서** 생존을 비교합니다. 5장의 다중회귀, 6장의 로지스틱 회귀와 같은 자리에 있습니다.

장	결과 변수	모형	나오는 지표
5	연속형	선형회귀	회귀계수 β
6	이분형	로지스틱 회귀	**OR**
8	**시간-사건**	**Cox 회귀**	**HR**

셋 다 회귀계수의 지수(exponential)로 상대 효과를 얻습니다. 형태가 같습니다.

구체적으로 계수는 이렇게 읽습니다. 5·6장에서 8장으로 미뤄 둔 "계수를 어떻게 읽는가"가 여기서 한 꺼번에 정리됩니다.

- **선형회귀의 β** — 노출이 한 단위 늘 때 결과(연속형)의 **평균이 얼마나 변하는가**. 지수를 취하지 않고 그대로 읽습니다.
- **로지스틱 회귀의 exp(β)** — 다른 변수를 보정한 **오즈비(OR)**. 6장의 OR을 교란까지 걸러 낸 값입니다.
- **Cox 회귀의 exp(β)** — 다른 변수를 보정한 **위험비(HR)**.

세 모형 모두 **exp(계수) = 다른 변수를 고정했을 때의 상대 효과**라는 한 문장으로 통합니다. 그래서 5·6·8장의 다변량 분석은 이름만 다를 뿐 같은 골격입니다 — 이것이 계수 해석을 세 곳에 흩지 않고 여기 모은 이유입니다.

★ HR은 RR이 아닙니다

	무엇의 비인가
RR (상대위험도)	**기간 전체의 누적 위험**의 비
HR (위험비)	**어느 순간의 사건 발생 속도**의 비

HR = 0.60 은 **"추적 기간 내내, 매 순간의 사건 속도가 40% 낮다"** 는 뜻입니다.

> **"5년 뒤 사망자가 40% 적다"는 뜻이 아닙니다**

누적으로 몇 명이 덜 죽는지는 **기저 위험에 따라 달라집니다.** 7장의 상대위험감소·NNT 논의와 같은 문제입니다 — **상대 지표는 기저 위험을 지웁니다.**

> 이 함정은 수업 중 투표 문제로 다시 만납니다.

★ 비례위험 가정

Cox 회귀는 **HR이 추적 기간 내내 일정하다**고 가정합니다.

이 가정이 깨지는 가장 흔한 신호:

> **두 생존곡선이 교차한다**

교차한다는 것은 **초기와 후기의 효과가 반대**라는 뜻입니다. 그것을 하나의 HR로 평균하면 **1에 가까운 값이 나오고, "차이 없음"으로 잘못 결론**납니다.

실제로는 환자에게 매우 중요한 차이가 있습니다 — "초기 위험을 줄여야 하면 A, 2년을 넘길 수 있으면 B"라는 서로 다른 조언이 나와야 합니다.

> **KM 곡선을 먼저 그려 눈으로 확인하십시오.** Cox 회귀는 그다음입니다. 이 지점은 수업 중 투표에서 다시 다룹니다.

비례위험 가정은 두 가지로 확인합니다.

- **로그-로그 생존그림** — $\ln(-\ln S(t))$ 를 시간의 로그에 대해 그렸을 때 두 군의 선이 **나란히(평행)** 가면 가정이 성립합니다. 벌어지거나 **교차하면 위반**입니다.
- **Schoenfeld 잔차** — 잔차가 시간과 상관(추세)을 보이면 HR이 시간에 따라 변한다는 신호입니다.

이름과 "무엇을 보는가"만 기억하면 됩니다. 계산은 소프트웨어가 하고, 우리는 **먼저 KM 곡선을 눈으로 확인**한 뒤 이 도구로 뒷받침합니다.

8.4 생명표

현재 생명표(current life table) 는 어떤 기간의 **연령별 사망률**을 요약해 "이 사망률이 계속된다면" 이라는 가정 아래 기대수명을 계산합니다.

> **개인의 수명을 예측하는 것이 아닙니다**

인구 집단의 사망 수준을 하나의 숫자로 요약하는 도구입니다. 연령별 사망률을 쓰므로 **인구 구조가 달라도 공정하게 비교**할 수 있습니다 — 3장의 연령 표준화와 같은 목적입니다.

연령별 자료가 매년 단위로 없으면 **단축 생명표(abridged life table)** 를 씁니다. 보통 5년 간격입니다.

정리 — 이 장에서 가져갈 다섯 줄

1. **중도절단**된 사람도 절단 시점까지의 정보를 기여한다. 그게 생존분석의 이점이다
2. 다만 **"절단된 사람이 남은 사람과 같다"는 가정** 위에 서 있다. 깨지면 생존이 과대추정된다
3. **세로축을 먼저 읽는다.** 1.0에서 내려가면 생존 확률, 0.0에서 올라가면 사건 발생 확률
4. **HR은 RR이 아니다.** 순간 속도의 비이지 누적 위험의 비가 아니다
5. **곡선이 교차하면 HR 하나로 요약할 수 없다.** 비례위험 가정이 깨진 것이다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 생존분석이 누적 발생률보다 나은 점은?

- A. 표본 크기를 절반으로 줄여 준다
- B. 추적이 중단된 사람의 정보도 쓴다
- C. 교란요인을 자동으로 보정해 준다
- D. 사망 자료에만 적용할 수 있다

2. 중도절단(censoring) 에 해당하는 경우는?

- A. 사건이 일어났으나 시점이 기록되지 않음
- B. 관심 사건과 다른 이유로 사망함
- C. 연구 시작 전에 이미 사건을 겪음
- D. 사건 없이 추적이 종료되거나 중단됨

3. Kaplan-Meier 곡선에서 **중앙 생존시간**을 읽는 지점은?

- A. 곡선이 평탄해지기 시작하는 시간
- B. 곡선이 1.0을 유지하는 마지막 시간
- C. 생존 확률이 0.5가 되는 시간
- D. 곡선이 0.1을 지나는 시간

4. Cox 회귀의 **비례위험 가정**이 뜻하는 것은?

- A. 모든 관찰값이 절단되지 않아야 한다는 것
- B. 생존 시간이 정규분포를 따라야 한다는 것
- C. 위험요인의 효과가 시간이 지나면 사라진다는 것
- D. 위험요인의 상대 효과가 시간에 걸쳐 일정하다는 것

5. 세로축이 **'사건 발생 확률'** 인 곡선이 0.0에서 시작해 올라간다. 5년 시점의 값이 **0.20** 이라면?

- A. 5년까지 20%가 사건을 경험했다
- B. 5년까지 20%가 사건을 피했다
- C. 5년 시점의 생존 확률이 20%다
- D. 5년 시점의 발생률이 연 20%다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 생존곡선과 중도절단** — 생존분석에서 추적이 끝나기 전에 빠진 사람(중도절단)을 어떻게 처리할까요? 그 사람들을 '남은 사람과 위험이 같다'고 가정하면 언제 문제가 될까요?

2. **2단계 · 위험비를 읽는 법** — Cox 회귀에서 위험비(HR)가 2로 나왔습니다. '두 배 많이 죽는다'고 읽어도 될까요? 이 하나의 숫자가 성립하려면 어떤 가정이 필요할까요?

제9장 · 체계적 문헌고찰과 메타분석

제9장 체계적 문헌고찰과 메타분석

> 5·6·7장에서 배운 **개별 연구를 읽는 법**이 여기서 쓰입니다. 이 장은 그 위에 얹힙니다.

이 장에서 답할 질문

- 연구 하나가 아니라 **연구 전체를 근거로 삼으려면** 무엇이 필요한가
- 결과가 서로 다를 때, 합쳐도 되는가
- 우리가 찾지 못한 연구는 어디에 있는가

> **이 장을 관통하는 물음** — 메타분석이 통합하는 것은 '수행된 모든 연구'가 아니라 **'출판된 연구'**다. 둘이 다르면(출판 편향) 통합값은 현실의 평균이 아니라 **세상에 나온 것의 평균**일 뿐이다. 여기서는 개별 연구가 아니라 '증거의 집합'이 인구 현실을 담았는지를 묻는다.

9.1 왜 '체계적'이어야 하는가

14년

혈전용해제(심근경색 환자의 혈전을 녹이는 약)가 **1987년**에야 표준 치료로 권고되었습니다. 그런데 Antman 등(1992)이 70건의 무작위배정 시험을 **누적 메타분석**해 보니,

> **1973년에 이미 통계적으로 유의한 이득($p < 0.01$)이 확인될 수 있었습니다**

근거는 있었는데 **아무도 합쳐 보지 않았습니다.** 그 사이 교과서와 종설은 전문가의 인상에 따라 서로 다른 권고를 냈습니다.

> **체계적 문헌고찰은 새 자료를 만드는 일이 아니라, 이미 있는 자료를 낭비하지 않는 일입니다.**

무엇이 '체계적'을 만드나

	일반 문헌고찰	체계적 문헌고찰
논문 선정	저자가 아는 것	**사전에 정한 규칙**
검색	명시 없음	검색어·DB·기간을 기록
재현	불가능	**다른 사람이 반복 가능**
결론	저자의 판단	규칙의 결과

> **핵심은 재현 가능성입니다**

일반 고찰에서는 **저자가 고른 논문이 곧 결론**입니다. 고르는 순간 답이 정해집니다.

검색부터 최종 포함까지의 걸러짐을 한눈에 보이는 것이 **PRISMA 흐름도**입니다. 아래는 **예시 숫자**로 그 골격을 보인 것입니다(실제 값은 워크숍 논문의 표에서 확인하십시오).

숫자는 예시 · 실제 값은 논문 표 참조

핵심은 **각 단계에서 몇 편이, 왜 빠졌는지를 숫자로 남긴다**는 것입니다. 이 흐름도가 있어야 다른 연구자가 같은 검색을 반복해 같은 48편에 도달할 수 있습니다 — 이것이 '체계적'의 뜻입니다.

이 장의 관통 사례

> **Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, et al. (2005).** *Water, sanitation and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis.* **Lancet Infect Dis 5, 42–52.**

개발도상국의 설사병을 줄이기 위한 **여러 종류의 중재**(수질 개선, 위생시설, 손씻기 등)를 모아 비교한 연구입니다. **중재 종류가 다양하다는 점**이 이 장의 핵심 논점인 이질성과 곧바로 연결됩니다.

9.2 메타분석의 방법

통합은 평균이 아니다

통합 추정치는 각 연구 결과를 단순 평균한 값이 **아닙니다.** 연구마다 **정밀도가 다르므로** 가중치를 다르게 줍니다 — 분산의 역수로 가중합니다.

> 표본이 크고 신뢰구간이 좁은 연구가 더 큰 목소리를 냅니다. 당연한 이야기지만, **그래서 작은 연구가 여럿 모여도 큰 연구 하나를 못 이길 수 있습니다.**

★ 이질성 — 합쳐도 되는가

이질성(heterogeneity) 은 연구들의 결과가 **서로 어긋나는 정도**입니다.

숲그림에서 각 연구의 신뢰구간이 서로 겹치지 않는다면, 그 연구들은 **같은 것을 재고 있지 않을 가능성**이 큽니다.

	뜻
대상 집단이 다르다	도시 vs 농촌, 아동 vs 성인
중재의 강도가 다르다	손씻기 교육 1회 vs 6개월 프로그램
결과 정의가 다르다	자가보고 설사 vs 진단된 설사

이질성이 크면 통합값 하나로 요약하는 것 자체가 의미를 잃습니다. RR 0.4 인 연구와 2.6 인 연구를 평균해 1.12 를 얻어도, **그 1.12 를 경험하는 집단은 어디에도 없습니다.**

고정효과와 무작위효과 — 가정이 다르다

	가정	언제
고정효과	모든 연구가 **하나의 같은 참값**을 추정한다	이질성이 없을 때
무작위효과	**참값 자체가 연구마다 다를 수 있다**	이질성이 있을 때

계산 방법의 차이가 아니라 **세계관의 차이**입니다.

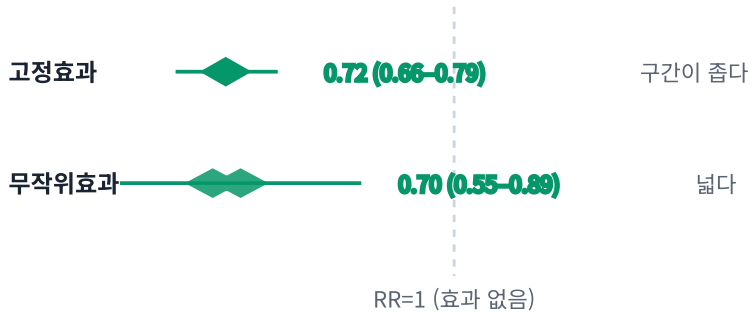
Paper A 의 실제 규칙은 이랬습니다.

> *"Random-effects models were used … if the test of heterogeneity was significant (**defined conservatively as p < 0.20**). In the absence of heterogeneity, fixed-effects models were used."*

기준이 0.05 가 아니라 0.20 입니다. 이질성 검정은 검정력이 낮아 실제 이질성을 놓치기 쉬우므로, 일부러 느슨하게 잡아 무작위효과 쪽으로 기울인 것입니다.

> **다만 무작위효과는 이질성을 인정하고 구간을 넓히는 것이지 해결하는 것이 아닙니다.** 이질성이 크면 통합보다 먼저 할 일은 **"왜 다른가"를 묻는 것**입니다. 이 지점은 수업 중 투표에서 다시 다룹니다.

같은 연구들에 두 모형을 돌리면 **통합값은 비슷해도 구간의 폭이 달라집니다.** 아래는 **예시**입니다.



무작위효과는 "참값이 연구마다 다를 수 있다"는 **연구 간 분산**을 구간에 더해 넣기 때문에 넓어집니다. 이질성이 있을 때 고정효과의 좁은 구간은 **없는 확신**을 줍니다. 그래서 이질성이 의심되면 무작위효과로 보고하되, **넓어진 구간이 곧 "우리는 하나의 답을 모른다"는 정직한 표현**임을 기억해야 합니다.

9.3 우리가 찾지 못한 연구

출판 편향

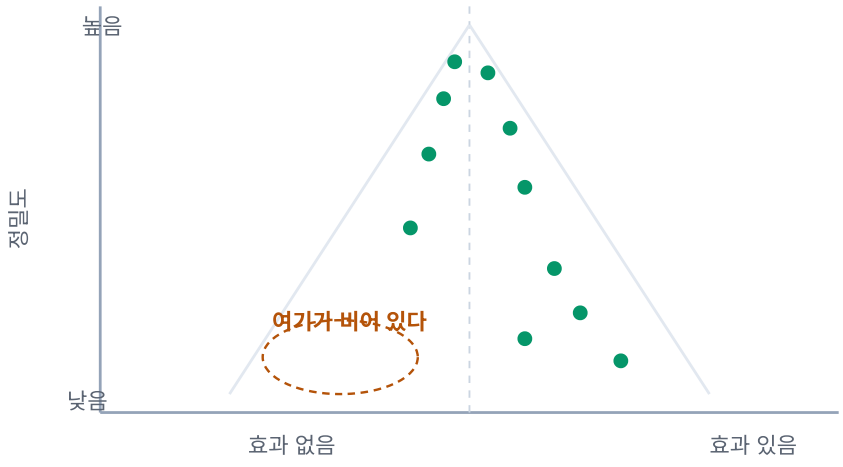
> **유의한 결과가 더 잘 출판됩니다**

- 작은 연구는 **우연히 큰 효과가 나와야** 유의해지고, 그래야 실립니다
- 효과가 없게 나온 작은 연구는 **서랍에 남습니다**
- 큰 연구는 결과와 무관하게 대체로 출판됩니다

그래서 메타분석에 들어온 작은 연구들은 **효과가 있는 쪽으로 선별**되어 있고, 통합 추정치는 참값보다 **크게** 나옵니다.

깔때기 그림

각 연구를 **효과 크기(가로) × 정밀도(세로)** 로 찍습니다.



왼쪽 아래가 비면 효과 없는 작은 연구가 묻혔다는 신호입니다.

Paper A 는 **Begg 검정**을 썼고, 판정 기준을 역시 **p < 0.20** 으로 느슨하게 잡았습니다. 이유는 같습니다 — 이 검정도 검정력이 낮습니다.

> **없는 자료를 통계로 만들어낼 수는 없습니다.** 무작위효과 모형도 빠진 연구를 되살리지 못합니다. 이 함정은 수업 중 투표 문제로 다시 만납니다.

9.4 편수보다 선정 기준

포함된 연구가 많을수록 좋은 메타분석이라는 생각은 틀립니다.

> **질 낮은 연구를 많이 넣으면**
> **통합값이 그쪽으로 끌려갑니다**

48편 중 39편이 편향돼 있다면 **그 39편은 근거가 아니라 잡음**입니다.

그래서 확인할 것은 편수가 아니라 이것입니다.

- 포함배제 기준이 **사전에** 정해졌는가
- 개별 연구의 **비뚤림 위험(risk of bias)** 을 평가했는가
- 설계를 제한했는가 (무작위배정 시험만? 관찰연구 포함?)

무작위배정 시험이 아니라 **관찰연구를 메타분석할 때는 한 가지를 더 조심**해야 합니다. 관찰연구는 저마다 **다른 교란요인이 다른 방식으로 남아 있습니다.** 어떤 연구는 흡연을 보정했고, 어떤 연구는 못 했습니다. 이런 연구들을 아무리 많이 모아 통합해도 **그 남은 교란은 사라지지 않습니다.** 오히려 편향된 연구가 여럿이면 통합값이 그 편향의 방향으로 더 단단히 굳습니다. 그래서 관찰연구 메타분석에서는 통합 이전에 **각 연구가 교란을 어떻게 다뤘는가(비뚤림 위험)** 를 반드시 함께 평가합니다.

정리 — 이 장에서 가져갈 다섯 줄

1. 체계적 고찰의 핵심은 **재현 가능성**이다. 일반 고찰은 저자의 선택이 곧 결론이다
2. 통합은 평균이 아니라 **정밀도로 가중한 합**이다

3. **이질성이 크면 통합값 하나가 의미를 잃는다.** 합치기 전에 왜 다른지를 물어야 한다
4. 고정효과와 무작위효과는 **가정이 다르다.** 무작위효과는 이질성을 인정할 뿐 해결하지 않는다
5. **묻힌 연구는 통계로 되살릴 수 없다.** 깔때기 비대칭은 통합값이 부풀려졌다는 신호다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 일반 문헌고찰과 **체계적** 문헌고찰을 가르는 것은?

- A. 반드시 메타분석을 함께 수행한다
- B. 사전에 정한 규칙으로 빠짐없이 찾는다
- C. 무작위배정 시험만을 대상으로 삼는다
- D. 저널 편집자의 사전 승인을 받는다

2. 메타분석에서 **이질성(heterogeneity)** 이 크다는 것은?

- A. 포함된 연구의 표본 크기가 서로 다르다
- B. 연구들의 효과 추정치가 서로 어긋난다
- C. 출판되지 않은 연구가 많이 존재한다
- D. 통합 추정치의 신뢰구간이 좁아진다

3. 고정효과 모형과 무작위효과 모형의 **가정**이 다른 지점은?

- A. 고정효과는 참값이 하나라고 본다
- B. 고정효과는 큰 연구에 가중치를 준다
- C. 무작위효과는 결측 자료를 보정한다
- D. 무작위효과는 출판 편향을 교정한다

4. 출판 편향(publication bias) 이 메타분석에 미치는 영향은?

- A. 통합 추정치의 신뢰구간이 넓어진다
- B. 이질성 검정의 검정력이 높아진다
- C. 효과가 실제보다 크게 나오기 쉽다

- D. 표본이 큰 연구가 과소 대표된다

5. 어떤 메타분석의 이질성 검정 결과가 **p = 0.03** 이었다. Paper A(Fewtrell 2005)가 정한 기준을 따른다면 어느 모형을 쓰는가?

- A. 고정효과 모형을 쓴다
- B. 두 모형의 평균을 낸다
- C. 통합을 포기하고 서술만 한다
- D. 무작위효과 모형을 쓴다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 무엇이 '체계적'인가** — 그냥 문헌고찰과 '체계적' 문헌고찰은 무엇이 다를까요? 그리고 논문을 많이 모을수록 더 좋은 메타분석일까요?
2. **2단계 · 통합과 그 함정** — 숲그림(forest plot)에서 연구들의 신뢰구간이 서로 겹치지도 않는데 맨 아래 마름모(통합값) 하나로 결론을 내려도 될까요?

제10장 · 예방 전략과 선별검사

제10장 예방 전략과 선별검사

> 이 장은 **연구방법이 정책으로 바뀌는 지점**입니다. 계산 하나가 사업 하나를 뒤집습니다.

이 장에서 답할 질문

- 위험이 높은 사람에게 집중하는 것이 왜 충분하지 않은가
- 좋은 검사인데 왜 양성의 대부분이 병이 없는가
- 조기 발견이 생존을 늘렸다는 말을 어떻게 검증하는가

> **이 장을 관통하는 물음** — 선별검사가 "효과 있다"는 말은 인구 집단의 **사망을 실제로 줄였는가**로만 확인된다. 생존기간이 길어 보이는 것은 진단 시점을 앞당긴 절차의 착시일 수 있다. 절차가 만든 좋아 보이는 숫자와, 인구의 실제 이득을 갈라 보는 것이 이 장의 핵심이다.

10.1 예방의 단계

단계	언제	예
1차 예방	발병 전	금연, 백신, 나트륨 규제
2차 예방	발병 후 · 증상 전	**선별검사**
3차 예방	발병 후	재활, 합병증 관리

선별검사는 2차 예방입니다. 병을 막는 것이 아니라 **이미 시작된 병을 일찍 찾는** 것입니다. 이 구분이 10.3에서 중요해집니다 — 일찍 찾는 것이 반드시 좋은 것은 아니기 때문입니다.

10.2 예방의 역설

어디에 자원을 쓸 것인가

어느 지역의 혈압 분포입니다.

혈압	인구	10년 위험	예상 환자
160 이상	5,000명	20%	**1,000명**
140–159	25,000명	8%	**2,000명**
140 미만	70,000명	4%	**2,800명**

고위험군(160 이상)은 **개인별 위험이 5배** 높습니다. 여기에 집중하는 것이 당연해 보입니다.

그런데 전체 환자 5,800명 중 **고위험군에서 나오는 것은 1,000명뿐**입니다.

> **83%는 위험이 낮거나 중간인 사람들에게서 나옵니다**

Rose 의 예방 역설

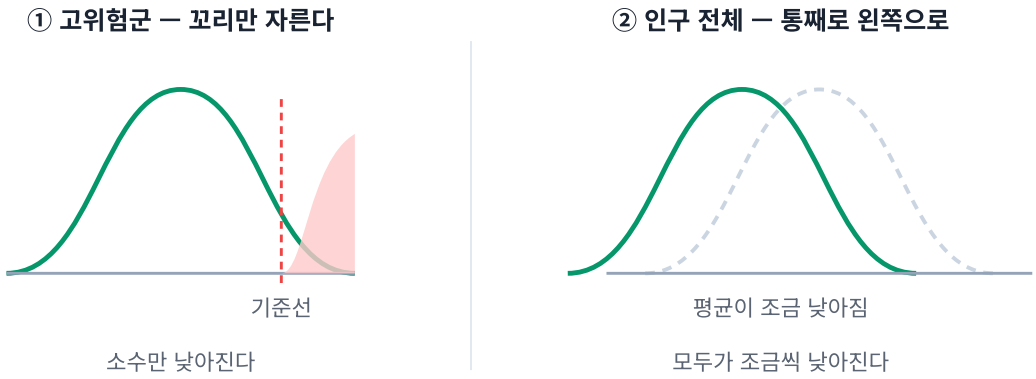
> **Rose G (1981).** *Strategy of prevention: lessons from cardiovascular disease.* **BMJ 282, 1847–1851.**

> *"위험이 조금 높은 다수가, 위험이 매우 높은 소수보다 더 많은 환자를 만들어낸다."*

	고위험군 전략	인구 전체 전략
대상	소수	전체
개인의 이득	**크다**	작다
인구 전체 효과	**작다**	**크다**
예	고혈압 약물치료	가공식품 나트륨 규제

둘 중 하나를 고르는 문제가 아닙니다. 고위험군 관리는 그 사람들에게 필요하고, 인구 전체 접근은 환자 수를 줄이는 데 필요합니다. **역할이 다릅니다.**

두 전략을 위험 분포 위에 그리면 차이가 한눈에 보입니다.



①은 기준선을 넘는 소수를 끌어내립니다. 개인에게는 큰 이득이지만, 환자의 대부분이 사는 곡선의 가운데는 그대로입니다. ②는 곡선 전체를 왼쪽으로 조금 밀어 **모두의 위험을 조금씩** 낮춥니다. 개인이 얻는 이득은 작지만, 가운데 다수가 함께 내려가므로 **인구 전체의 환자 수는 더 많이 줄어듭니다.** 이것이 예방 역설이 정책으로 번역되는 방식입니다.

10.3 선별검사의 성능

네 칸 표에서 시작한다

	실제 환자	실제 건강
검사 양성	참양성	위양성
검사 음성	위음성	참음성

지표	계산	분모가 무엇인가
민감도	참양성 / (참양성+위음성)	**환자**
특이도	참음성 / (참음성+위양성)	**건강한 사람**
양성예측도(PPV)	참양성 / (참양성+위양성)	**양성으로 나온 사람**
음성예측도(NPV)	참음성 / (참음성+위음성)	음성으로 나온 사람

> **분모가 다르면 다른 질문입니다**

- 민감도: **"환자를 놓치지 않는가"** — 검사를 만드는 사람의 질문
- PPV: **"내가 양성인데, 나는 환자인가"** — **검사를 받은 사람의 질문**

★ 같은 검사, 전혀 다른 의미

민감도 99%, 특이도 99%. 아주 좋은 검사입니다. **10만 명을 검사합니다.**

유병률 0.1%인 병

```
환자 100명 · 건강 99,900명

참양성 = 100 × 0.99     =     99명
위양성 = 99,900 × 0.01  =    999명
──────────────────────────────────
양성 합계                =  1,098명

PPV = 99 / 1,098 = 9%
```

양성이라고 통보받은 사람의 91%가 병이 없습니다.

같은 검사, 유병률만 바꾸면

유병률	양성자 수	그중 환자	PPV
0.1%	1,098명	99명	**9%**
1%	1,980명	990명	**50%**
30%	30,400명	29,700명	**98%**

> **검사 성능은 그대로인데 결과의 의미가 완전히 달라집니다**

> **드문 병을 전 인구에 검사하면 대부분의 양성이 거짓입니다.** 그래서 선별검사는 **유병률이 어느 정**

> **도 되는 집단**에 해야 합니다. 이 계산은 수업 중 투표에서 직접 해 봅니다.

우도비(likelihood ratio). 양성우도비 LR+ = 민감도 / (1 − 특이도) 로, 검사가 양성일 때 질병의 오즈가 몇 배가 되는지를 말합니다. LR+가 클수록 양성 결과가 강한 근거입니다. 민감도·특이도처럼 **검사 자체의 성질이라 유병률과 무관**하며, 사전확률(유병률)에 곱해 개인의 사후확률을 얻는 데 씁니다.

ROC 곡선. 연속형 검사에서 **양성/음성을 가르는 기준(cut-off)** 을 옮기면 민감도와 특이도가 맞바뀝니다 — 기준을 낮추면 더 많이 양성으로 잡아 민감도는 오르지만 특이도는 떨어집니다. ROC 곡선은 모든 기준값에서 (1 − 특이도)를 가로, 민감도를 세로로 찍은 것입니다. 곡선 아래 면적(AUC)이 검사의 전반적 변별력을 하나의 값으로 요약하는데, **AUC의 해석은 11장**에서 다룹니다.

10.4 선별검사를 평가할 때의 함정

선행기간 편향 (lead-time bias)

두 사람이 **같은 날 발병해 같은 날 사망**한다고 합시다.

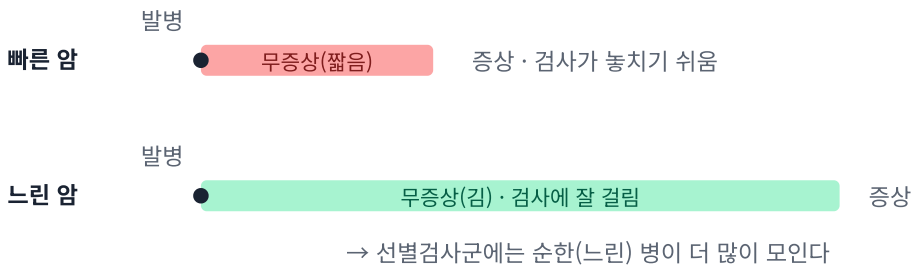
진단 후 생존기간은 3년 → 5년으로 늘었습니다. 그러나 아무도 더 오래 살지 않았습니다.

> 구분이 중요합니다 — 조기 발견으로 **실제로** 더 오래 사는 것은 **효과**이고, 진단이 빨라져 **길어 보이는** 것은 **편향**입니다. 생존율만으로는 둘을 구분할 수 없습니다.

> **선별검사의 효과는 생존기간이 아니라 사망률로 재야 합니다**

기간차 편향 (length-time bias)

진행이 느린 병은 **무증상 기간이 깁니다.** 그래서 **검사에 걸릴 확률이 높습니다.**



결과적으로 **선별검사군에는 순한 병이 더 많이 모입니다.** 검사가 효과 있어서가 아니라, **잡히는 병의 종류가 다르기** 때문에 예후가 좋아 보입니다.

자발적 참여 편향

검사를 받으러 오는 사람은 대체로 **건강에 관심이 많고, 금연·운동을 하고, 형편이 낫습니다.** 검사와 무관하게 예후가 좋습니다.

10.5 그래서 선별검사를 할 것인가

세 편향을 모두 피하려면 결국 **무작위배정 시험**이 필요합니다 — 검사군과 비검사군을 무작위로 나누고 **인구 10만명당 사망률**을 비교하는 것입니다.

그런 근거가 있는 선별검사는 생각보다 적습니다.

> **선별검사에는 비용이 있습니다.** 거짓양성으로 인한 불필요한 정밀검사와 불안, 그리고 평생 증상을 일으키지 않았을 병을 찾아내 치료하는 **과진단·과치료**입니다. 효과가 입증되지 않은 선별검사는 **해를 끼칠 수 있습니다.**

선별검사를 도입해도 되는지는 1968년 Wilson & Jungner 가 정리한 열 가지 기준으로 점검합니다. 검사의 성능만이 아니라 **병·치료·비용·운영**을 함께 봅니다.

#	기준
1	그 병이 **중요한 건강 문제**여야 한다
2	발견된 환자에게 쓸 **인정된 치료법**이 있어야 한다
3	진단·치료 **시설과 자원**이 갖춰져 있어야 한다
4	**잠복기 또는 초기 증상기**가 있어(일찍 잡을 여지가 있어야) 한다
5	적절한 **검사 방법**이 있어야 한다
6	그 검사가 대상 인구에게 **받아들여질** 만해야 한다
7	병의 **자연 경과**가 충분히 알려져 있어야 한다
8	누구를 환자로 보고 치료할지에 대한 **합의된 기준**이 있어야 한다
9	사례 발견의 **비용이 전체 의료비에 비추어 균형**을 이뤄야 한다
10	선별은 한 번으로 끝내지 않고 **지속되는 과정**이어야 한다

실제 검사 하나를 이 표에 대입해 보면 논쟁의 지점이 드러납니다. 예를 들어 무증상 성인의 갑상선암 초음파는 4·9번(과진단으로 이어질 잠복기, 비용 대비 이득)에서 걸립니다 — 이 판정을 워크숍에서 합니다.

정리 — 이 장에서 가져갈 다섯 줄

1. 선별검사는 **2차 예방**이다. 병을 막는 것이 아니라 일찍 찾는 것이다
2. **위험이 조금 높은 다수에서 더 많은 환자가 나온다** — Rose 의 예방 역설
3. **분모가 다르면 다른 질문이다.** 민감도의 분모는 환자, PPV의 분모는 양성자다
4. **드문 병에서는 좋은 검사도 양성의 대부분이 거짓이다.** PPV는 유병률에 좌우된다
5. 선별검사의 효과는 **생존기간이 아니라 사망률**로 잰다

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. 선별검사(screening)는 예방의 어느 단계에 해당하는가?

- A. 1차 예방 — 발병 자체를 막는다
- B. 3차 예방 — 합병증을 줄인다
- C. 4차 예방 — 과잉의료를 막는다
- D. 2차 예방 — 조기에 찾아 개입한다

2. 검사의 **민감도(sensitivity)** 는 무엇을 재는가?

- A. 양성인 사람 중 실제 환자의 비율
- B. 건강한 사람을 음성으로 가려내는 비율
- C. 환자를 양성으로 잡아내는 비율
- D. 검사 결과가 반복 측정에서 일치하는 정도

3. 선행기간 편향(lead-time bias) 이란?

- A. 진행이 느린 병이 검사에 잘 걸리는 것
- B. 검사를 받는 사람이 원래 더 건강한 것
- C. 검사 장비가 시간이 지나며 낡는 것
- D. 진단이 앞당겨져 생존이 길어 보이는 것

4. 환자 100명 중 90명이 양성, 건강한 900명 중 90명이 양성으로 나왔다. **양성예측도(PPV)** 는?

- A. 50%
- B. 90%
- C. 10%
- D. 45%

5. Rose(1981)가 말한 **예방의 역설(prevention paradox)** 은?

- A. 예방사업은 비용이 치료보다 더 많이 든다
- B. 위험이 조금 높은 다수에서 환자가 더 많이 나온다
- C. 예방이 성공하면 그 성과가 보이지 않는다
- D. 고위험군은 개입을 잘 따르지 않는 경향이 있다

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · 선별검사의 함정** — 선별검사로 암을 발견한 사람들이 증상이 나타난 뒤 발견된 사람보다 오래 산다는 자료가 있습니다. 그럼 선별검사가 수명을 늘린 걸까요?
2. **2단계 · 양성예측도와 예방 전략** — 민감도 99%, 특이도 99%인 아주 좋은 검사가 있습니다. 이 검사에서 양성이 나오면 병이 있다고 봐도 될까요? 무엇에 따라 달라질까요?

제11장 · 확률분포와 가설검정

제11장 확률분포와 가설검정

> **논문 워크숍이 없습니다.** 대신 지금까지 열 개 장에서 써온 도구들의 **바닥을 봅니다.**

이 장에서 답할 질문

- `p = 0.02` 를 정확히 뭐라고 읽어야 하는가
- 신뢰구간은 무엇에 대한 구간인가
- 스무 번 검정하면 무슨 일이 생기는가

> 이 장은 새로운 것을 배우는 자리가 아닙니다. **5장부터 10장까지 쓰던 숫자들이 실제로 무엇을 뜻했는지** 확인하는 자리입니다.

11.1 표본에서 모집단으로

우리는 늘 **표본**을 보고 **모집단**을 말해 왔습니다.

- 코호트 7,735명 → 영국 중년 남성 전체
- 사례 155명 → 오클랜드의 임신부 전체

그 도약에는 **불확실성**이 따릅니다. 통계는 그 불확실성을 **숫자로 표현**하는 도구입니다.

> **통계는 답을 주는 것이 아니라**
> **답이 얼마나 불확실한지를 알려 줍니다**

11.2 신뢰구간

무엇에 대한 구간인가

> **참값(모수)에 대한 구간입니다**

개별 관측값의 분포가 아닙니다.

	무엇의 범위	표본이 커지면
신뢰구간	**참값**	**좁아진다**
참고범위	개별 관측값	거의 그대로

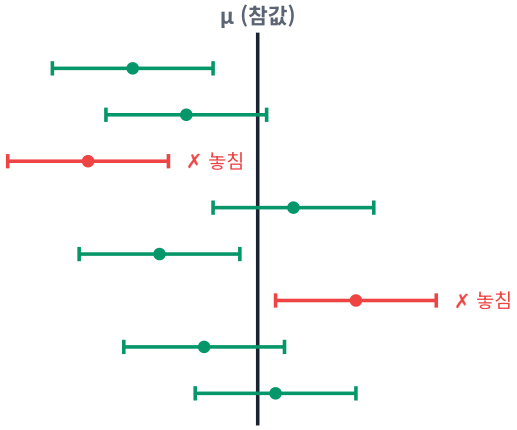
"환자의 95%가 이 범위 안에 있다"로 읽으면 **완전히 다른 것**을 말하게 됩니다.

정확한 뜻

표본을 100번 뽑아 각각 95% 신뢰구간을 만들면, **그중 약 95개가 참값을 포함**합니다.

우리는 그중 **하나**를 갖고 있고, 그것이 참값을 포함하는지는 알 수 없습니다.

같은 모집단에서 표본을 거듭 뽑아 매번 95% 신뢰구간을 그리면 이렇게 됩니다(세로선이 참값 μ).



100개를 그리면 약 95개가 참값을 가로지른다

100개를 그리면 **약 95개가 세로선을 가로지르고, 약 5개가 빗나갑니다.** "95%"는 이 절차의 성질입니다. 우리는 그중 **하나의 구간**만 손에 쥐고 있고, 그 하나가 참값을 품었는지 아닌지는 알 수 없습니다. 그래서 "이 구간에 참값이 있을 확률이 95%"라는 표현은 엄밀히는 틀립니다 — 확률은 구간을 만드는 절차에 붙지, 이미 그어진 한 구간에 붙지 않습니다.

폭이 말하는 것

신뢰구간의 폭 = 정밀도입니다. 유의성은 그 구간이 1(또는 0)을 포함하는지의 문제일 뿐입니다.

	해석
RR 1.15 (1.02–1.30)	작은 효과를 **정밀하게** 추정
RR 2.80 (0.90–8.70)	**아무것도 배제하지 못함**

5장에서 다룬 그 문제입니다. **유의하지 않다 ≠ 효과가 없다.**

11.3 가설검정

두 가지 오류

	실제로 효과 없음	실제로 효과 있음
있다고 결론	**제1종 오류 (α)**	옳음
없다고 결론	옳음	**제2종 오류 (β)**

- **α** — 헛것을 볼 확률. 보통 0.05로 정합니다
- **β** — 있는 것을 놓칠 확률
- **검정력 = 1 − β** — 있는 효과를 잡아낼 확률. 보통 80%로 잡습니다

5장 §5.6의 표본 크기 계산이 바로 이 값들을 정하는 일이었습니다.

★ p값이 말하지 않는 것

```
p값이 말하는 것 :  P( 이만큼 극단적인 자료 | 효과가 없다 )
흔한 오해       :  P( 효과가 없다 | 이 자료 )
```

> **조건이 뒤집혀 있습니다**

10장의 PPV와 정확히 같은 구조입니다.

10장	11장
민감도 = P(양성 | 환자)	p값 = P(자료 | 귀무가설)
PPV = P(환자 | 양성)	우리가 알고 싶은 것 = P(가설 | 자료)
유병률이 낮으면 PPV가 9%	**사전 확률이 낮으면 유의해도 틀릴 수 있음**

민감도 99%인 검사도 드문 병에서는 양성의 91%가 거짓이었습니다. 같은 이유로, **$p < 0.05$ 라도 애초에 가능성이 낮은 가설이었다면 여전히 틀릴 확률이 높습니다.**

> 이 지점은 수업 중 투표에서 다시 다룹니다.

다중검정

유의수준 0.05로 **20개를 검정하면 효과가 하나도 없어도 평균 1개가 유의하게** 나옵니다.

```
20 × 0.05 = 1
```

찾아낸 뒤 **그것만 골라 보고하면 거짓 발견**입니다.

논문을 읽을 때 물어야 할 것:

- **몇 개를 검정했는가**
- **주 결과를 사전에 정했는가**
- 하위군 분석이 **탐색적 결과**로 표시되어 있는가

> **7장의 시험 사전등록이 이 문제의 구조적 해법입니다.** 주 결과를 미리 못 박아 두면 사후 취사선택이 불가능해집니다. 이 지점은 수업 중 투표에서 다시 다룹니다.

⚠ 사후 검정력은 쓰지 마십시오

결과가 유의하지 않았을 때 **관측된 효과 크기로 검정력을 다시 계산**하는 일이 흔합니다.

> **그 값은 p값을 다시 쓴 것일 뿐입니다**

p가 크면 사후 검정력은 **반드시** 낮게 나옵니다. 순환논법입니다. "검정력이 낮아 유의하지 않았다"는 **"유의하지 않아서 유의하지 않았다"** 와 같은 말입니다.

대신 신뢰구간의 폭을 보십시오. 무엇을 배제하지 못했는지가 답입니다.

> 이 함정은 수업 중 투표 문제로 다시 만납니다.

11.4 어떤 검정을 쓰는가

5장 §5.5.4의 표를 다시 봅니다. 이제 각 칸의 뜻을 압니다.

비교하려는 것	자료	모수적	비모수적
두 독립 집단의 평균	연속형	두 표본 t검정	Mann-Whitney
짝지은 두 값의 차	연속형	대응 t검정	Wilcoxon 부호순위
세 집단 이상	연속형	분산분석	Kruskal-Wallis
두 범주형 변수	범주형	카이제곱	Fisher 정확검정
두 연속형의 관계	연속형	Pearson 상관	**Spearman 순위상관**

모수적 검정의 가정이 충족되면 모수적을 씁니다 — 검정력이 더 높기 때문입니다. 가정이 깨지면 비모수적으로 갑니다. 검정력을 잃는 대신 가정에서 자유로워집니다.

> 자료가 심하게 치우쳐 있으면 **로그 변환** 후 모수적 검정을 쓰는 방법도 있습니다.

11.5 베이즈 방식

사전 확률을 명시한다

```
사후 오즈 = 사전 오즈 × 우도비
```

- **사전 오즈** — 검사 전에 이미 알고 있던 것 (유병률, 임상 정황)
- **우도비** — 이 검사가 확률을 얼마나 움직이는가
- **사후 오즈** — 검사 후의 확률

> **10장의 PPV 계산이 바로 이 구조였습니다**

유병률(사전 확률)이 0.1%일 때 민감도·특이도 99%인 검사로도 PPV가 9%였던 것은, **우도비가 아무리 커도 사전 확률이 너무 낮으면 사후 확률이 크게 오르지 못하기** 때문입니다.

왜 이 방식이 유용한가

빈도주의는 **"이 자료가 얼마나 이상한가"** 를 묻습니다. 베이즈는 **"그래서 지금 무엇을 믿어야 하는가"** 를 묻습니다.

임상과 정책은 후자를 필요로 합니다.

> **관통 사례**: Powell J, Geddes J, Deeks J, Goldacre M, Hawton K (2000). *Suicide in psychiatric hospital inpatients: risk factors and their predictive power.* **Br J Psychiatry 176, 266–272.**
>
> 위험요인의 **예측력(predictive power)** 을 다룬 논문입니다. 자살은 드문 사건이므로, 위험요인이 통계적으로 유의해도 **개인을 예측하는 데는 거의 쓸모가 없습니다.** 10장의 PPV 문제가 정신과 임상에서 그대로 반복됩니다.

예시 숫자로 이 논문의 논점을 계산해 봅니다(실제 값은 워크숍에서 논문 표로 확인하십시오).

입원 정신과 환자에서 자살은 드뭅니다. 기저율(사전 확률)을 **1%** 라 합시다. 어떤 위험요인이 통계적으로 유의하고 양성우도비 **LR+ = 2** 라고 합시다(유의하지만 약한 요인).

```
사전 오즈  = 0.01 / 0.99          ≈ 0.0101
사후 오즈  = 0.0101 × 2           ≈ 0.0202
사후 확률  = 0.0202 / (1+0.0202)  ≈ 0.020  = 2%
```

이 위험요인을 가진 사람의 자살 확률은 1%에서 **2%로** 오릅니다. 통계적으로 분명히 유의하지만, **여전히 98%는 자살하지 않습니다.** 이 요인 하나로 개인을 선별하려 하면 거의 전부가 거짓양성입니다. **"유의한 위험요인"과 "쓸모 있는 예측도구"는 다릅니다.** 드문 사건에서는 사전 확률이 너무 낮아, 어지간한 우도비로는 사후 확률이 실용적인 수준까지 오르지 못합니다 — 10장 PPV 문제의 재현입니다.

정리 — 이 장에서 가져갈 다섯 줄

1. **신뢰구간은 참값에 대한 것**이지 개별 관측값의 분포가 아니다
2. **p값은 P(자료 | 가설)이지 P(가설 | 자료)가 아니다.** 조건이 뒤집혀 있다
3. **20번 검정하면 효과가 없어도 1개가 유의하다.** 몇 개를 검정했는지 물어라
4. **사후 검정력은 p값의 재표현**이다. 대신 신뢰구간의 폭을 보라
5. 사전 확률이 낮으면 좋은 검사도, 유의한 결과도 **여전히 틀릴 수 있다**

확인 문제

> 읽었는지 스스로 점검하는 문제입니다. 답은 싣지 않았습니다. yonsei QST 에서 풀고, 틀리면 AI 튜터와 확인하십시오.

1. p값이 뜻하는 것은?

- A. 귀무가설이 참일 확률을 나타낸 값
- B. 대립가설이 참일 확률을 나타낸 값
- C. 같은 결과가 재현될 확률을 나타낸 값
- D. 귀무가설 하에 이 자료가 나올 확률

2. 평균의 **95% 신뢰구간**이 뜻하는 것은?

- A. 관측값의 95%가 이 범위 안에 들어 있다
- B. 표본의 95%가 이 범위에서 추출되었다
- C. 이 범위가 참값을 포함할 것이라 확신한다
- D. 다음 환자의 값이 95% 확률로 이 안에 든다

3. 제2종 오류(type II error) 는?

- A. 실제로 있는 효과를 없다고 결론짓는 것
- B. 실제로 없는 효과를 있다고 결론짓는 것
- C. 표본을 잘못 추출하여 생기는 오류
- D. 자료 입력 과정에서 생기는 오류

4. 유의수준 0.05로 **서로 독립인 20개의 검정**을 수행했다. 실제로는 아무 효과도 없다고 할 때, 유의하게 나오는 검정의 기대 개수는?

- A. 0개
- B. 1개
- C. 5개
- D. 20개

5. 베이즈 방식으로 검사 후 확률을 구할 때 필요한 두 가지는?

- A. 표준 오차와 p값
- B. 민감도와 특이도
- C. 상대 위험도와 기여 위험도
- D. 사전 오즈와 우도비

스스로 생각해 보기

> 정답을 찾는 질문이 아닙니다. 개념을 **자기 말로** 세워 보십시오. yonsei QST 의 AI 튜터가 같은 질문으로 대화하며, 여러분의 답이 어디까지 닿았는지 짚어 줍니다.

1. **1단계 · p값과 신뢰구간** — $p = 0.03$이 나왔습니다. 이것을 '귀무가설이 참일 확률이 3%'라고 읽어도 될까요? p값은 정확히 무엇의 확률인가요?
2. **2단계 · 유의성의 함정** — 표본이 아주 큰 연구에서 $p < 0.05$가 나왔습니다. 그럼 이 효과는 크고 중요한 걸까요? 그리고 한 연구에서 20개를 검정해 1개가 유의했다면 그것을 보고해도 될까요?