Research Workflows with AI

From data cleaning to literature management, live demos of current tools

Yohan Koo
CEO, CMDSPACE; Adjunct Professor, CHA University; Visiting Professor, KIRD

◆ 강의 마스터 노트

You will leave today with three things and real cases

01

Boundaries

At each of four stages, what AI does and what you decide

02

One rule file

One file that stops AI from deciding silently

03

Checking habits

Numbers from the log, citations from the DOI

◆ 강의 마스터 노트

Each stage has its own boundary

Stage	Hand to AI	Researcher decides
Data cleaning	Full checks, cleaning code	**What to fix or drop**
Statistics	Computation, assumptions	**Variables, interpretation**
Writing	Structure, draft, editing	**Claims and accountability**
Literature	Search, summary, matching	**What to cite**
Every stage	Logging, converting, searching	**What counts as the master copy**

◆ 자산-02 v1v2v3 진화 · ◆ 대규모 연구 시작 가이드

How we ask AI has changed three times

v1	**Prompt**	Ask well. Gone when you close the window
v2	**Context**	Provide materials too. Every single time
v3	**Harness**	Keep rules and materials in a folder. AI comes in and reads them

◆ 자산-02 v1v2v3 진화 · ◆ 자산-04 하네스

A public heart-failure cohort, seeded with hospital-style flaws

299

Heart-failure patients, Faisalabad, Pakistan, 2015

13

Clinical variables, including death during follow-up

28

Seeded flaws: missing, mixed units, label mismatches, impossible values, duplicates

No hospital data was used. This is a copy of public data (CC BY 4.0).

UCI 심부전 데이터셋 · § Ahmad 2017 PLOS ONE · § Chicco 2020 BMC MIDM

The demo uses public data. The real case is my own research

307

Pages in the final report (PDF). The manuscript is one Markdown file

1,000

Synthetic personas, each answering 190 items. Collected three times

4

Output formats: HWPX, PDF, HTML, DOCX. All generated from one manuscript

STEPI planning study for the 2045 S&T Frontier Strategy (Jun to Sep 2026), PI Yohan Koo

Data Cleaning

AI now finds problems well. The open question is who decides

◆ 자산-12 AI-ready 데이터

Never edit the original, have AI make a copy

Original	Read only. Not a single character changes
Cleaned copy	A new file in the clean folder
Change log	One line per changed value, with the reason

◆ 원본은 고치지 않는다 · ◆ 시연 키트 규칙 파일

AI found every seeded problem

Seeded problem	Count	No rules, 3 runs	Rule folder
Duplicate entries	3	All 3 runs	Found
Inconsistent labels	11	All 3 runs	Found
Mixed units	5	All 3 runs	Found
Impossible values	3	All 3 runs	Found
Missing values	6	All 3 runs	Found

Claude Code + Opus 5.5, same data and same question, run three times

◆ 사전 실행 기록 · UCI 심부전 데이터셋

Then it decided without asking

No rules, 3 runs

Sodium 1360 changed to 136

8 missing EF dropped or filled

Each run decided differently

→

Rule folder

Sodium 1360 left missing

Asked for a decision on EF

Every change logged with a reason

Original was 131: wrong the same way 3 times

◆ 사전 실행 기록 · ◆ 규칙 폴더 분석 로그

Ten lines made the difference

```
# CLAUDE.md  (put it in the folder, AI reads it first)
- Never modify the original. Write cleaned data as new files in clean/
- Check quality first. Do not delete on your own; propose reasons and a plan first
- Do not use variables set after the outcome (e.g. follow-up time) as predictors
- Not p-values alone: report effect sizes and 95% confidence intervals
- Append every analysis to notes/analysis-log.md
- Do not impute values that are not in the data
- No causal language for observational studies
```

◆ 시연 키트 규칙 파일

Before you add patient data

ASK	Does the IRB approval cover AI processing?
ASK	Is the environment allowed by data policy?
ASK	Pseudonymized, export approved?
STOP	~~Until confirmed,~~ **~~no patient data goes in~~**
DO	Build the procedure on public or synthetic data
DO	AI writes the code; run it in approved settings

◆ 시연 키트 규칙 파일

Leave received files as they are and make a reading copy

01

Received original

HWPX, PDF, XLSX kept exactly as received. Not one byte changed

02

Reading extract

Converted to Markdown. Record source name, hash and the tool used

03

One-line description

At the top: what, when, who. The line AI reads first

04

One master per role

Originals in the archive, judgments in notes, code in the run folder

Statistics

Do not ask for numbers; get the computation

◆ 시연 키트 규칙 파일

Get the computation, not just the numbers

Run	AI runs the code itself and checks the output
Log	Question, processing, rationale, and results in the log
Limits	What this data cannot answer, stated separately

◆ 시연 키트 규칙 파일

Same question 3 times, 3 answers

Same model, data, and question, but different analytic decisions changed the results.

No rules	Analyzed	Missing data	Age OR	Cross-validated AUC
Run 1	291	Excluded	1.60	0.76
Run 2	299	Median imputation	1.73	0.755
Run 3	291	Excluded	1.71	0.75

Age odds ratio per 10 years. Runs 1 and 3 entered creatinine differently.

◆ 사전 실행 기록

When performance jumps, suspect it first

0.755

Cross-validated AUC without follow-up time (no rules, run 2)

0.873

Cross-validated AUC with follow-up time (same run)

Follow-up ends at death. Follow-up time already contains the outcome.

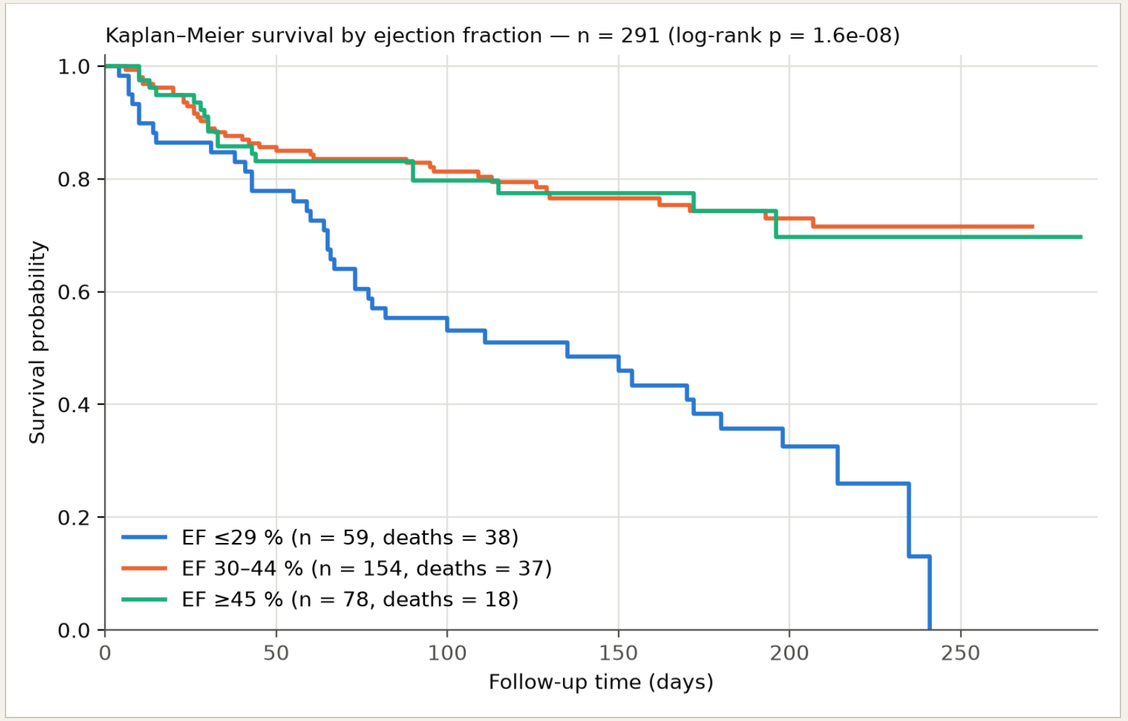
◆ 사전 실행 기록 · § Chicco 2020 BMC MIDM

The rule folder chose survival analysis

Follow-up varied from 4 to 285 days, so Cox instead of logistic

The ejection fraction association is mainly at 29% or below

It flagged curves after day 230 as unstable, with only 1 to 2 at risk



Survival curves by three ejection fraction groups, drawn by the rule folder (pre-run 2026-10-06, n=291)

◆ 규칙 폴더 분석 로그 · § Ahmad 2017 PLOS ONE

Get results in this order

01

Assumptions

Proportional hazards and linearity first

02

Effect size

Hazard ratios and 95% CIs

03

Sensitivity

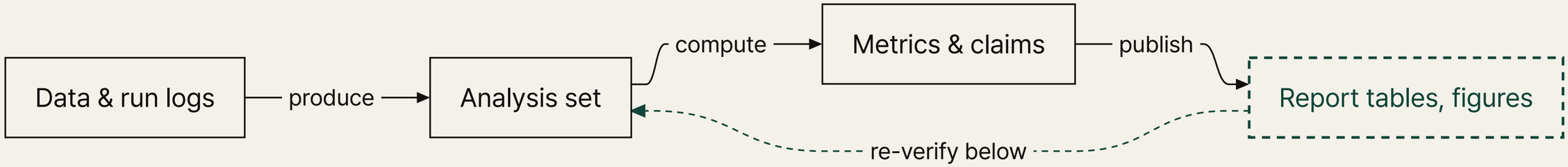
Does it hold if decisions change?

04

What we can't say

Causation, generalization, missing variables

Every number has a lineage



Every STEPI report table traces back to raw data. Dotted line: re-verification

◆ 대규모 연구 시작 가이드 · ◆ research-harness 스킬 · STEPI 탐색 웹

Writing

Never start from a blank page

IMRaD is not a format, it is questions in order

I	**Why**	What do we not know?
M	**How**	How did we find out?
R	**What**	What numbers are in the log?
D	**So what**	What can we claim, and what can't we?

STROBE

Every number in the abstract carries a log number

구조화 초록 초안 (2026-10-06)

숫자 출처는 `notes/analysis-log.md` 항목 번호. 인용은 `refs.md` 키만 사용.

Background: In this public heart failure cohort, age, serum creatinine and left ventricular ejection fraction (EF) have been reported as predictors of death [ahmad2017; chicco2020]. We re-analysed it with explicit quality and assumption checks.

Methods: We audited 302 records [L1.1], removing duplicates, converting unit-mixed creatinine and setting implausible values to missing without imputation [L1.1–L1.6]. Follow-up time, determined by the outcome, served only as the time scale [L2.1]. A pre-specified Cox model included all 11 baseline variables without stepwise selection, using complete cases [L3]. Proportional hazards and sensitivity to suspect platelet/CPK values were examined.

Results: Of 299 patients, 96 (32.1%) died over a median 115 days (4–285) [L3.1]. The model included 290 patients and 93 deaths (8.5 events per variable) [L3.2]. Adjusted hazard ratios were 1.50 (95% CI 1.24–1.82) per 10 years of age, 0.81 (0.73–0.89) per 5-point higher EF and 1.92 (1.43–2.56) per doubling of creatinine (all p<0.001), 1.63 (1.06–2.49) for hypertension and 1.53 (1.00–2.33) for anaemia [L3.3]. Hypertension, anaemia, EF and creatinine showed suspected non-proportional hazards [L3.6]; hypertension's association appeared mainly within 60 days (2.31 vs 1.07 later) [L3.7]. The EF association was concentrated at EF ≤29% (HR 2.86, 1.76–4.64, vs 30–44%) [L3.8]. Sodium's unadjusted association (0.71, 0.59–0.86) [L3.4] attenuated after adjustment (0.85, 0.67–1.07) [L3.3]. Age, EF and creatinine were consistent across sensitivity analyses [L3.9–L3.11]. Apparent C-index: 0.732 [L3.5].

Conclusions: Older age, lower EF and higher creatinine were consistently associated with death during follow-up; associations for hypertension, anaemia and sodium were less robust. These observational findings require external validation.

Structured abstract draft by the rule folder, from the output file (pre-run 2026-10-06, 248 words)

◆ 규칙 폴더 초록 초안 · ◆ 시연 키트 규칙 파일

How long to produce a 307-page report in four formats?

A A week

B A day

C 30 minutes

D **Seconds**

Think it over

ANSWER **D Seconds**

Conversion takes seconds. Research and writing took four months

◆ STEPI v1.6 발송 패키지 · ◆ 연구 3공간 원리 · ◆ STEPI v1.7 날짜 대조표

A 307-page report, one manuscript file

01

Master manuscript

One Markdown file. One author, many reviewers

02

Build snapshot

Freeze the manuscript in the run folder and record its hash

03

Four formats

A script generates HWPX, PDF, HTML and DOCX

04

Parity check

1,261 table rows and 49 figures verified identical across formats

Conversion takes seconds; research and writing took four months. What was saved was formatting time

◆ STEPI v1.6 발송 패키지 · ◆ 연구 3공간 원리 · ◆ STEPI v1.7 날짜 대조표

Humans are responsible for any submitted material that included the use of AI-assisted technologies.

ICMJE, AI Use by Authors

AI cannot be an author. Responsibility for the content rests with people.

ICMJE AI 사용 규정

What to disclose at submission

Item	ICMJE recommendation
Authorship	AI tools cannot be authors
Where to disclose	Cover letter and manuscript
Responsibility	Check errors, omissions, bias
Citation	AI output is not a primary source
Nondisclosure	May be research misconduct

Also check the target journal's own policy.

ICMJE AI 사용 규정

Three models read it separately before submission

Axis	What was actually caught	Fix
Typo	Doubled sentence ending	Single ending
Who is counted	982 (an intersection) shown as total	Restore the condition
Statistics	ICC defined only as 0 to 1	Note negative estimates
Terminology	ROC-AUC called accuracy	Discrimination (AUC)
Arithmetic	26 vs 24 called 'half and half'	52% vs 48%

18 chunks x 3 models, 462 raw findings. Confirmation by a human re-reading the source

◆ STEPI 3모델 검증 기록 · ◆ research-verify 스킬

Not a well-written summary, one that shows how far to trust it

NUMBERS, GRADES

[n] per sentence, evidence grade A to I per claim

LINES

Solid lines are the source, dotted lines my reading

FIXED BOXES

Scope of evidence, where it fails, open questions

Literature

Citations come only from papers you have verified

§ Walters 2023 Sci Rep

Fabricated citations are fewer, but not zero

55 %

GPT-3.5 citations to nonexistent papers

18 %

GPT-4 citations to nonexistent papers

24 %

Real GPT-4 citations with bibliographic errors

A study checking 636 citations from literature reviews on 42 topics, using 2023 models.

§ Walters 2023 Sci Rep

Three steps for every citation

01

Acquire

Get the full text yourself with PubMed and Zotero

02

Index

Build one list of DOI-verified papers (refs.md)

03

Synthesize

AI cites only from that list

◆ 인용은 확인한 문헌 안에서

One DOI confirms it exists

```
curl -s api.crossref.org/works/10.1186/s12911-020-1023-5
200  Machine learning can predict survival of patients with heart ...
     BMC Medical Informatics and Decision Making, 2020
curl -s api.crossref.org/works/10.1186/s12911-021-0477-2
404  Resource not found
```

Actual output, 2026-10-06. The second DOI is a deliberate fake.

Crossref REST API · § Chicco 2020 BMC MIDM

Papers outside the list came only as candidates

Background 근거 문헌

인용한 문헌 (refs.md에 있고 DOI 확인됨)

- **ahmad2017** — Ahmad T, et al. *PLOS ONE* 2017. doi:10.1371/journal.pone.0181001. 이 코호트의 원 연구입니다. 나이, 크레아티닌, EF가 사망 예측 인자라는 근거로 썼습니다.
- **chicco2020** — Chicco D, Jurman G. *BMC Med Inform Decis Mak* 2020;20:16. doi:10.1186/s12911-020-1023-5. 같은 데이터를 다시 분석해 크레아티닌과 EF의 예측력을 보고한 논문입니다.

refs.md는 DOI만 확인한 목록입니다. 두 논문이 위 변수들을 실제로 유의한 인자로 보고했는지는 투고 전에 본문에서 한 번 더 확인해 주세요. 원 연구는 고혈압과 빈혈도 보고한 것으로 알고 있지만, 직접 확인하지는 못했습니다.

후보 문헌 (DOI 미확인 — 초록에는 인용하지 않았습니다)

심부전의 높은 사망률 같은 일반적인 배경 문장을 넣으려면 아래 문헌이 필요합니다. 그래서 지금 Background에는 그런 문장을 넣지 않았습니다.

1. Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev* 2017. 심부전 부담과 사망률에 관한 문헌입니다.
2. Heidenreich PA, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure. 2022 가이드라인입니다.
3. McDonagh TA, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. 2021 가이드라인입니다.
4. 저나트륨혈증과 심부전 예후 관련 문헌(예: OPTIME-CHF 분석, Klein L, et al. *Circulation* 2005). 나트륨 결과를 Discussion에서 다룰 때 쓸 수 있습니다.

Methods에서 쓴 방법에도 문헌이 필요할 수 있습니다. EPV 기준은 Peduzzi et al. 1996, 비례위험 검정은 Grambsch & Therneau 1994가 후보입니다. 둘 다 "추가 확인 필요"입니다.

Part of the rule folder's literature response, from the output file (pre-run 2026-10-06)

◆ 규칙 폴더 문헌 단계 · § Chicco 2020 BMC MIDM · § Ahmad 2017 PLOS ONE

Zotero is the reference point. Judgments live in my notes

01

Read and mark

PDFs and annotations stay in Zotero, the one reference point for bibliography

02

Send to a note

Pick a recently read paper; annotations become a note in reading order

03

Write the judgment

Supports, contradicts, uncertain. Which page it came from

04

Cite by key

Write [@citekey]; the bibliography builds itself. Unknown keys warn

My CMDS Zotero plugin. Updates go plan, review, apply. Never overwrites silently

◆ 근거와 승인 계약 · CMDS Zotero (GitHub) · ◆ CMDS Zotero 0.5 명세 · ◆ 위키 ↔ Zotero 가이드

Ask seven times separately, then judge by the source

7

Independent sessions: ChatGPT 2, Gemini 3, Grok, Claude. None sees another's answer

5

5 completed, 2 failed (usage limit, error). Failures recorded as they are

756

First editions in the Luhmann Archive list (1958 to 2021) checked against

'70 books, 400 papers' turned out to be a number spread from an unsourced Wikipedia sentence.

◆ 루만 서지 멀티모델 조사 · Niklas-Luhmann-Archiv

Across the stages

Three things that keep running beneath all four stages: harness, context, judgment

◆ research-harness 스킬

A research harness is four records

01

Project contract

Purpose, master locations, deliverable list, approval basis

02

Evidence ledger

From data to deliverables, the lineage of every number

03

Decision record

Who decided what, when, under which conditions

04

Handoff check

Checkpoint, open items, the one next action

Agents start by reading these four and finish by writing them

◆ research-harness 스킬 · ◆ 대규모 연구 시작 가이드

Context must survive for AI to pick up my research

01

Session link

Each note links back to the conversation that created it

02

Handoff note

On finishing, write the checkpoint and the next action

03

Packet size

Over 40KB in one call, hand over a summary instead

04

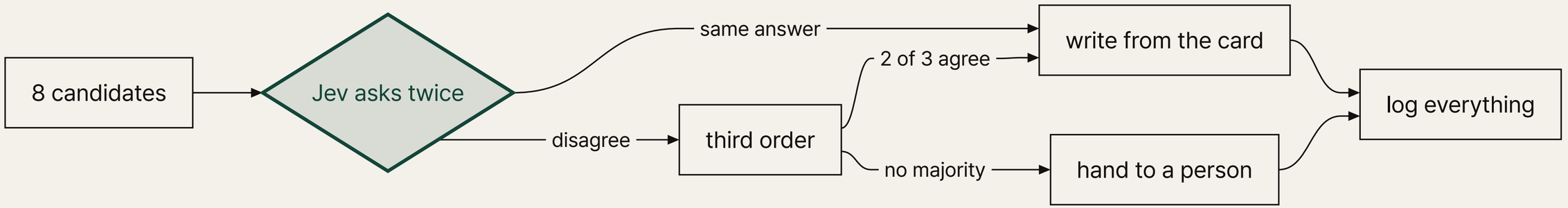
Search all sessions

'Did we discuss this before?' across every past conversation

138KB in one call errored; 88KB crashed the session twice. Measured

◆ 세션 링크 시스템 · ◆ 인계·납품 점검 템플릿

Choosing and writing are separated



ask.cmdspace.work. 58 visits (Oct 1-7): answered 26, held 29, to a person 3

◆ 자산-14 고르는 모델과 쓰는 모델 · ask.cmdspace.work · Jev (TypeSafe) · ◆ Jev 활용법 총정리

One rule file exposes silent decisions

01

Boundaries

Researchers decide what to fix or drop

02

One rule file

The folder's CLAUDE.md keeps the same procedure

03

Checking habits

Numbers from the log, citations from the DOI

◆ 시연 키트 규칙 파일 · ◆ research-harness 스킬

Start tomorrow in 30 minutes

☑ DO — Make one ongoing study a folder; keep originals apart

☑ DO — Adapt ten lines of the rule file to your study

☑ DO — Log date, question and reasoning for each analysis

◆ 시연 키트 규칙 파일

Ask your questions, we will work them out live

Post your research stage and where you are stuck in the chat

ask.cmdspace.work

Appendix

Materials for those who want to go deeper

◆ 자산-04 하네스

Tools for each stage

Web chat

ChatGPT, Claude, Gemini. Quick questions, one or two files

Folder-based agents

Claude Code, Codex. Work where rule files and logs build up

PubMed

Search. Automate with the E-utilities API, no key needed

Zotero

Stores full texts and references. Its export is where refs.md starts

Crossref

Confirms a DOI exists. Lookup with no key

Obsidian

An editor that keeps research notes and rule files in one folder

PubMed E-utilities · Zotero · Crossref REST API

Web chat vs. folder-based agents

When web chat fits

- One-off questions
- Summarizing one or two files
- Polishing sentences, translation
- Exploration that needs no record

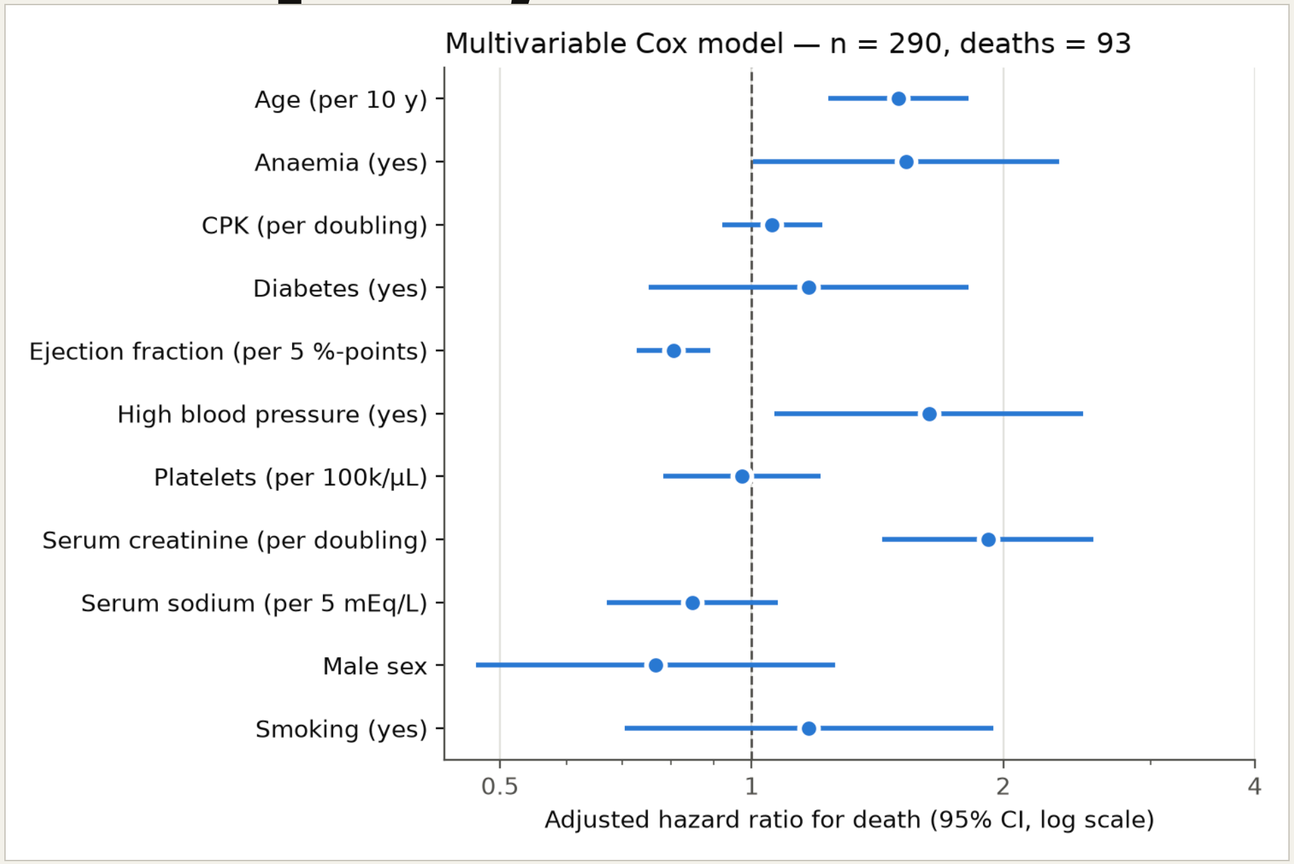
When a folder fits

- Analyzing the same data repeatedly
- Numbers going into a paper
- Aligning procedures with coauthors
- Research that runs for months

Format of one log entry

Field	Example from the rule folder
Question	Which quality issues, how to handle
Processing	01_clean.py, one row per change
Result	Sodium 1360 → missing
Rationale	No evidence to confirm the cause
Post-check	Original was 131; fixing was wrong

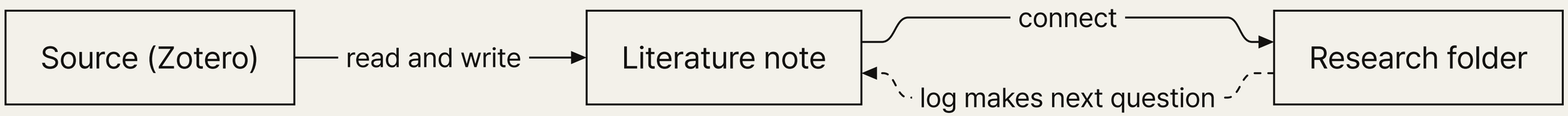
Hazard ratio plot, rule folder



Adjusted hazard ratios and 95% CIs from a multivariable Cox model (pre-run 2026-10-06, n=290, 93 deaths)

◆ 규칙 폴더 분석 로그

Read papers feed the next study



Sources in Zotero, judgment in notes, procedure in the folder. Dotted: back to the next question

◆ 자산-10 Second Brain · ◆ 자산-12 AI-ready 데이터

Folder structure of today's kit

data

Original

Read-only, plus codebook

clean

Cleaned copy

New files and change log

notes

Log

Analysis log and drafts

refs.md

Verified list

Only DOI-verified papers

Want to measure your AI practice?

AKM Index: diagnoses, on five axes, how well you supply knowledge and context to AI. akm.cmdspace.work

Judgment model measurements (Sep to Oct 2026)

Task	Result	Rule
Score replay	97% agreement at confidence 0.7+	Recalibrate the gate per task
87-way classification	Names only 51%, with definitions 76%	Never give names without definitions
Reversed option order	27% changed their answer	Ask in both orders
Search re-ranking	+16.7 pts over pool top-1	The hold signal is the value
Question desk	58: answered 26, held 29	If unsure, to a person

Zero human gold labels. Replay against existing labels, not an accuracy claim

◆ Jev 활용법 총정리 · Jev 공개 분류기 · Jev (TypeSafe)

Seven roles in a multi-model research run

Service	Mode	Role
ChatGPT	Max reasoning	Edition and criteria consistency
ChatGPT	Deep research	Primary bibliography trace
Gemini	Deep Think	Where the list could be wrong
Gemini	Deep research	Broad source coverage
Gemini	Flash, fast	Community and social sources
Grok	Heavy	Full research plus challenge
Claude	Max reasoning	Independent synthesis

Model names as shown on screen at run time. Agreement is a signal to verify, not evidence

Three spaces: think, exchange, run

Space	Master of	Holds
Obsidian vault	Judgment, writing	Questions, design, decisions, manuscript
External archive	Exchange evidence	Institution originals, delivered editions
Run folder (DEV)	Build, reproduce	Code, raw outputs, builds
Downloads	Convenience copy	PDF just before sending. Not a master

Split by edit rights, evidence role and lifespan, not by file extension

◆ 연구 3공간 원리 · ◆ 대규모 연구 시작 가이드

Frame a new study in ten minutes

01

One decision question

What, once known, changes whose decision

02

Count deliverables

From the contract: format, quantity, reviewer, deadline

03

Fix the three spaces

Actually check the vault, archive and run folder

04

Existing evidence first

Raw data, code, frozen copies, the last handoff

05

One next action

Owner, input, output, completion evidence, stop condition

Ten minutes is how to start preparing, not a bar for approving a design

◆ 대규모 연구 시작 가이드 · ◆ research-harness 스킬

Thank you

ask.cmdspace.work

Questions welcome after the seminar

ask.cmdspace.work

cmdspace.work

CMDSPACE

cmdspace.work

bio

Yohan Koo's bio and work

bio.cmdspace.work

Slides and the demo kit will be distributed via the Research Planning Center.

Last updated 2026-10-21 01:59 KST · R3