Managing Analytic Multiplicity in Epidemiology

Reproducible Multiverse and Vibration of Effects Analyses in R

Alyssa Columbus

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health ·
Department of Statistics, LMU Munich · Fulbright U.S. Student Researcher

useR! 2026 · Warsaw, Poland · Wednesday, 8 July 2026

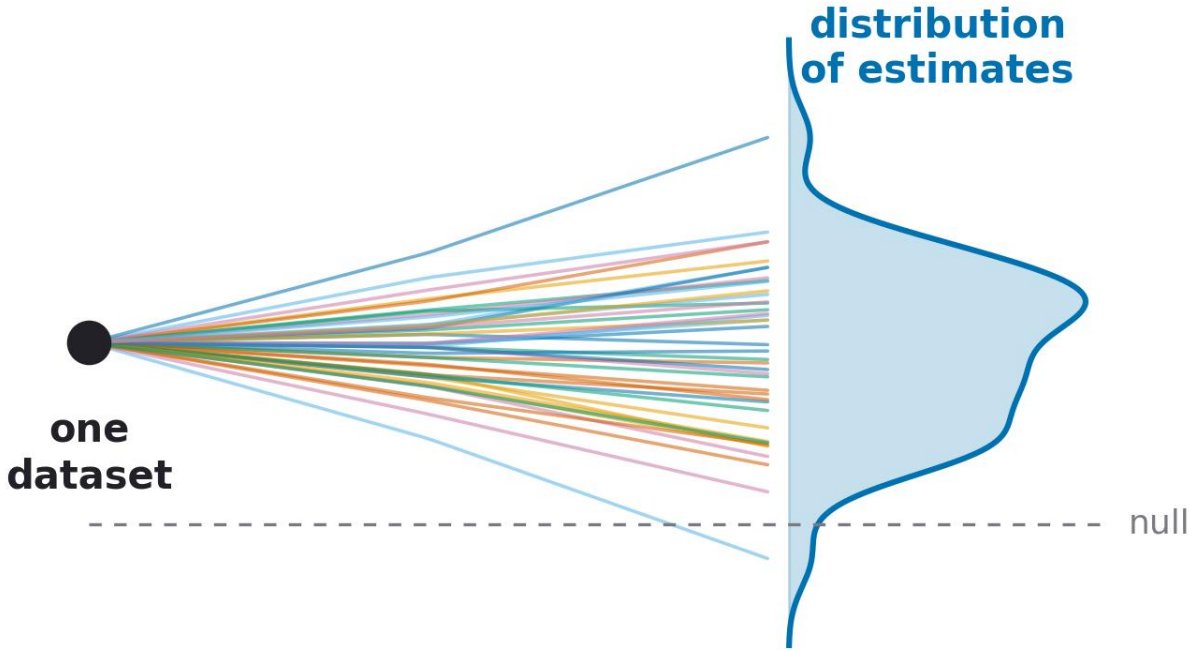
Slides to Follow Along `github.com/acolum/forkflow → slides.pdf`

One Dataset, Many Defensible Answers

A single analysis commits to **one path** through those choices.

Model form, covariate sets, transformations, and missing-data rules are all defensible choices, and each reasonable path can land on a **different estimate**.

The question this talk answers: how do we see the whole distribution of estimates, in R, and report it transparently?



Illustrative: paths from one dataset to many estimates.

Where the Analytic Choices Live

Model Specification

Functional form, link, interactions, random effects

Covariate Selection

Which confounders and adjusters enter the model

Transformations

Scaling, log or spline terms, categorization cut-points

Missing Data

Complete cases, single or multiple imputation

Exposure and Outcome Coding

Thresholds, composites, event definitions

Sample Construction

Inclusion criteria, outlier and eligibility rules

Six binary choices already give $2^6 = 64$ analyses. **Realistic studies reach thousands.**

The Same Data, Different Answers

8,192

Cox Models Per Exposure

2^{13} adjustment sets

417

Exposures Screened

for all-cause mortality

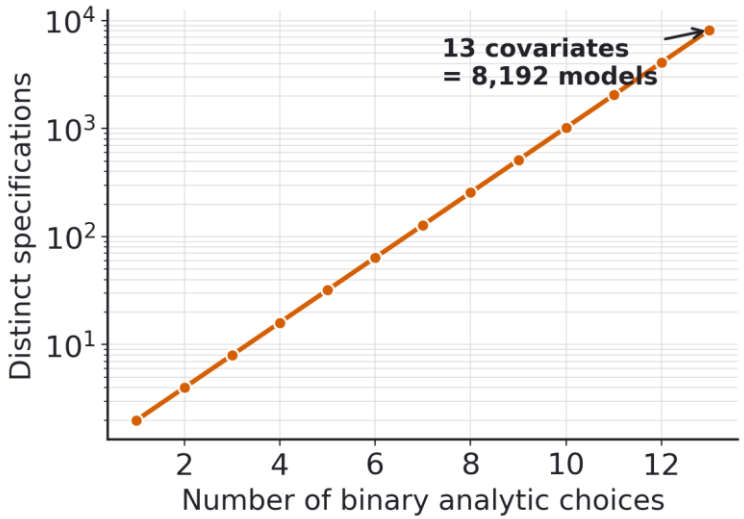
13

Adjustment Covariates

toggled on or off

Vibration of effects, applied to **NHANES** (National Health and Nutrition Examination Survey) mortality data, showed that swapping adjustment sets moves both the effect size and its p-value across a wide range.

The same pattern appears beyond epidemiology: **29 teams** analysing one dataset produced widely varying estimates (Silberzahn et al. 2018); **73 teams** on another question split on even the direction of the effect (Breznau et al. 2022).



Patel, Burford & Ioannidis 2015 · Klau, Hoffmann, Patel, Ioannidis & Boulesteix 2021

Two Complementary Lenses

Multiverse Analysis and Vibration of Effects

Multiverse Analysis

Steegen, Tuerlinckx, Gelman & Vanpaemel 2016

- Enumerate every defensible data-processing and modeling path
- Fit the full grid, then read the distribution of estimates
- Report the full specification curve of estimates
- Answers: how robust is the finding to reasonable choices?

Vibration of Effects

Patel, Burford & Ioannidis 2015

- Vary the adjustment set for a target association
- Plot effect size against $-\log_{10}(p)$ as a cloud of models
- Summarize spread with the relative hazard or odds ratio
- Flags Janus effects: the sign flips across specifications

One Reproducible Pipeline

A Multiverse Workflow, Built From Tidyverse Parts

1 Enumerate

`expand_grid()`
build the spec table

2 Fit

`future_map()`
run models in parallel

3 Tidy

`broom::tidy()`
one row per model

4 Summarize

`ggplot2 + specr`
curves and clouds

5 Report

Quarto
shareable, rerunnable

Every step is a data frame in, a data frame out. **The whole decision space becomes one reproducible object.**

Turn the Decision Space Into a Table

```
library(tidyverse)

# each analytic choice is a set of options
specs <- expand_grid(
  adjust_age = c(TRUE, FALSE),
  adjust_sex = c(TRUE, FALSE),
  adjust_bmi = c(TRUE, FALSE),
  transform  = c("identity", "log"),
  missing    = c("complete", "impute"),
  outliers   = c("keep", "winsorize"))

nrow(specs)    # 64 defensible specifications
```

One Row = One Analysis

expand_grid() takes the Cartesian product of every choice.

Each row is a complete, defensible specification.

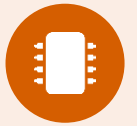
Adding a choice is one more argument, and the grid grows on its own.

Fit the Whole Grid, in Parallel

```
library(broom); library(furrr)
plan(multisession)          # use every core

fit_spec <- function(spec) {
  dat <- prepare(cohort, spec)  # apply choices
  m   <- coxph(
    Surv(time, event) ~ exposure + adjusters(spec),
    data = dat)
  tidy(m, conf.int = TRUE) |> filter(term == "exposure")
}

results <- specs |>
  mutate(out = future_map(seq_len(n()), \(i)
                          fit_spec(specs[i, ]))) |>
  unnest(out)             # tidy: one row per spec
```



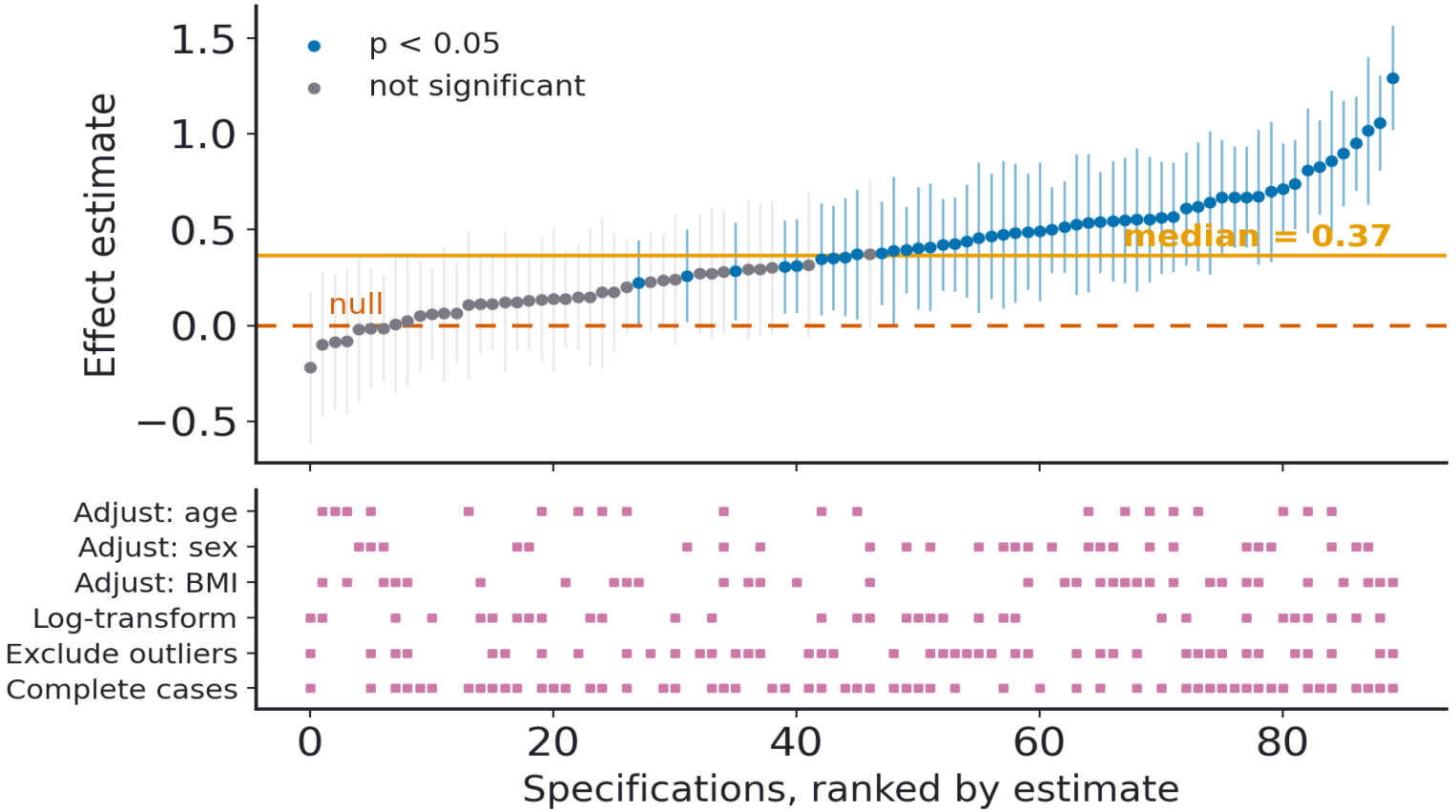
Keep It Fast

furrr spreads fits across cores with one plan() call.

broom::tidy() gives a uniform estimate table.

Cache prepared data and reuse the design matrix across specs that share it.

Reading the Whole Specification Curve



Illustrative specification curve.

How to Read It

1. Where the mass sits

the typical estimate across defensible models

2. How wide it spreads

a wide band means the result is sensitive to choices

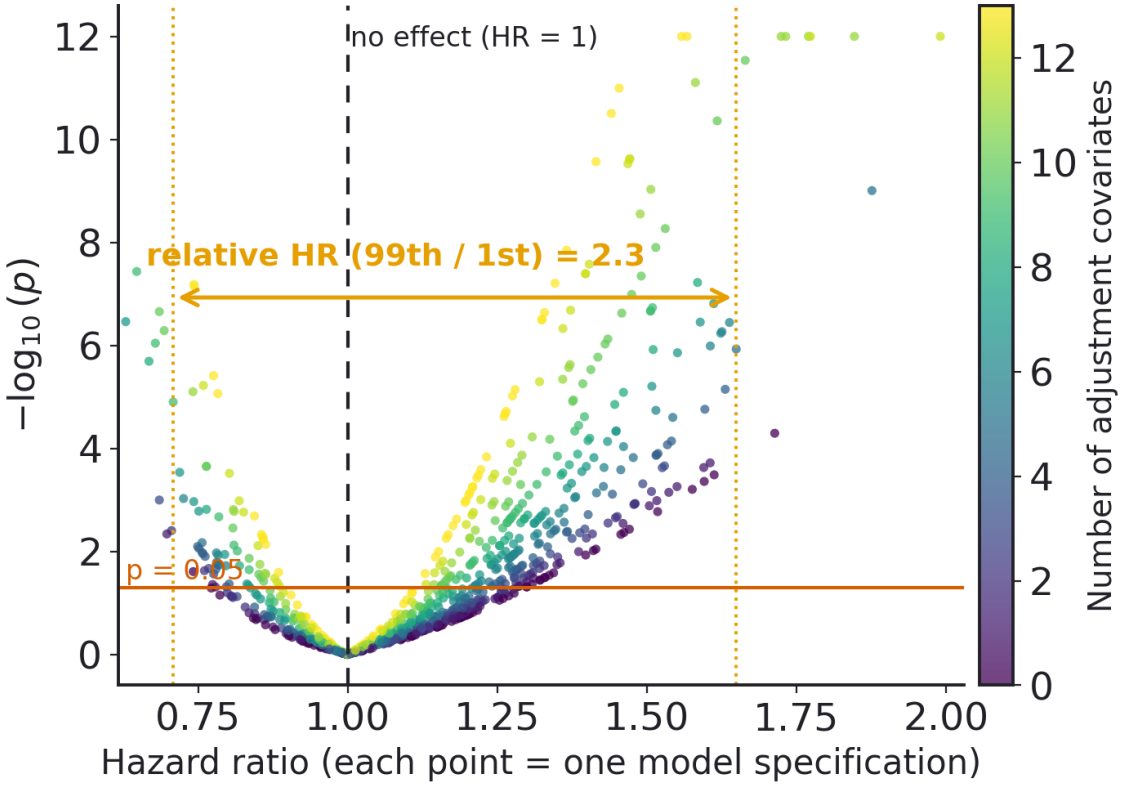
3. What crosses the null

the share of models whose interval includes the null

4. Which choices move it

the lower panel marks the active choices per model

Effect Size Against Significance, Across Models



Illustrative vibration-of-effects cloud.

`voe(results)` returns the relative ratio and Janus fraction

The Cloud

Each point is one adjustment set. A tight cloud means a stable association; a wide one means it depends on the model.

Relative Ratio

99th ÷ 1st percentile of the hazard ratio. Near 1 the effect barely moves; the further above 1, the more fragile the finding.

Janus Fraction

Share of estimates above the null. Near 0 or 1 the sign holds; near one half it flips across models.

Packages: `voe` `quantvoe`

Reproducible Scaffolding Around the Grid

targets

A dependency graph reruns only the specs that changed

renv

Pins package versions so the grid fits the same next year

Quarto

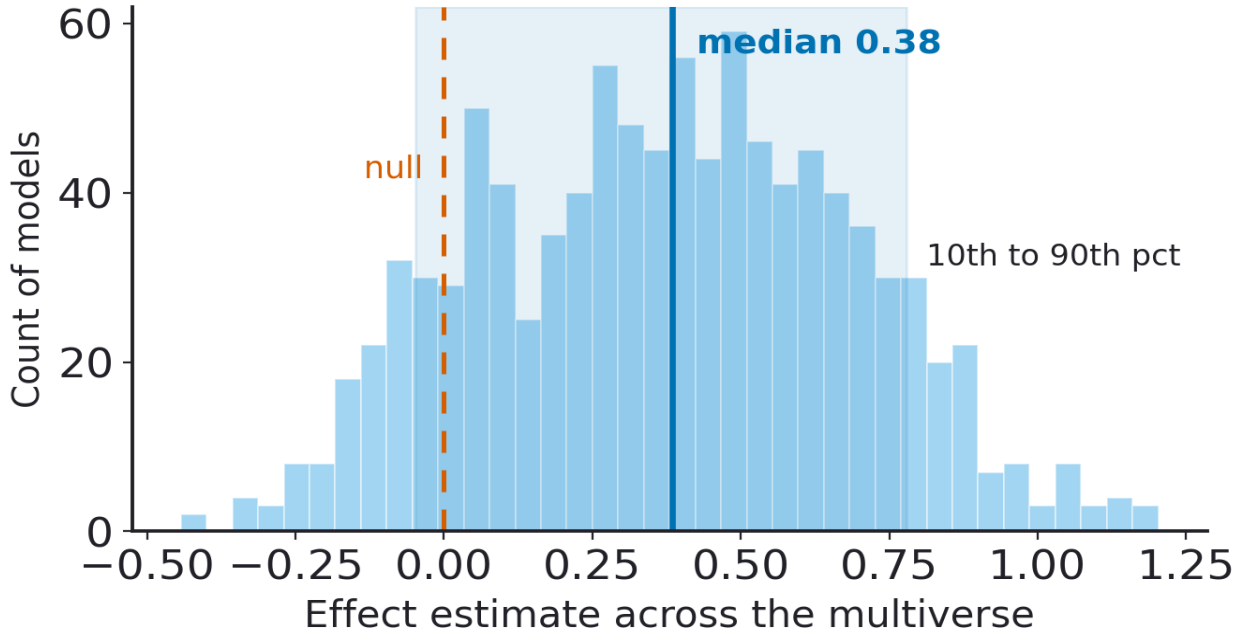
Weaves code, curves, and prose into one shareable report

set.seed()

Fixes imputation and resampling so results are exact

Reproducibility is the deliverable here: anyone can rerun the full decision space and land on the same distribution.

Reporting Analytic Uncertainty in Full



Illustrative distribution of estimates.

1. Show the Distribution

Median plus a percentile interval across specifications

2. Name the Drivers

Which choices move the estimate, and by how much

3. Mark Defensibility

Separate principled specs from arbitrary ones

4. Share the Grid

Publish the spec table and code alongside the summary

Doing This Well, and Its Limits

Not Every Fork Is Equal

Weight or gate specifications by defensibility; a wide multiverse of weak models misleads

Pre-Specify the Grid

Decide the decision space before seeing outcomes, and register it

Budget the Compute

Prune redundant specs, cache shared steps, and sample the grid when it is huge

Read It as Robustness

A robustness check on the finding, to be read as a range across specifications

The Whole Multiverse, in a Few Lines of forkflow

```
library(forkflow); library(survival); library(broom)
library(furrr); plan(multisession)

d     <- colon |> filter(etype == 2) |> drop_na()    # deaths
specs <- adjuster_sets(c("age","sex","obstruct",
           "perfor","adhere","differ","surg")) # 128 sets

results <- fit_multiverse(specs, \(s)
  coxph(reformulate(c("node4", s$adjusters[[1]]),
        "Surv(time, status)"), d) |> tidy(conf.int = TRUE) |>
filter(term=="node4"),
  .parallel = TRUE)

voe(results)  spec_curve(results)  voe_volcano(results)
```

Install: `pak::pak("acolum/forkflow")`

One Package

forkflow: spec_grid, fit_multiverse, voe, spec_curve, report

Real Data

survival::colon, a public trial. 2^7 = 128 fits, in parallel

Outputs

A voe() summary, a specification curve, and a VoE volcano

This Talk Is an Excerpt

A Package, a Book, an Open Project

The Package

forkflow: enumerate, fit, summarise, and report a multiverse. MIT licensed, with tests and continuous integration, headed for rOpenSci and CRAN.

The Book

Analytic Multiplicity in Epidemiology, a Quarto book of tutorials. This talk draws on Chapters 1, 6, and 8. All content is CC BY 4.0.

The Open Project

Reproducible-research tooling that anyone can build on, to increase trust in the process of data analysis. Issues and pull requests are welcome at the repository.

Roadmap v0.1 Workflow Verbs › rOpenSci / JOSS Review › CRAN › Book Published

All Resources at a Glance

Docs `acolum.github.io/forkflow`

Book `acolum.github.io/forkflow/book`

Code `github.com/acolum/forkflow`

Slides `github.com/acolum/forkflow → slides.pdf`

Takeaways for Any Analyst

1. **Report the whole distribution**
   One defensible analysis is one path among many; show the full spread of estimates.

2. **Write the decision space down first**
   Pre-specify the choices before seeing outcomes, and register them.

3. **Weight specifications by defensibility**
   Let a causal diagram choose the adjustment sets; treat arbitrary forks as checks.

4. **Read the spread as robustness**
   A relative ratio near 1 and a Janus fraction near 0 or 1 mean a stable finding.

5. **State the limits plainly**
   A multiverse is only as trustworthy as the specifications you include.

The Take-Away

Make the analytic decision space visible and reproducible.

Multiverse and vibration of effects analyses turn hidden researcher degrees of freedom into an auditable object you can build, test, and share. That workflow is packaged in forkflow.

R Toolkit

forkflow · tidyverse · broom · furrr · specr · multiverse / mverse · voe / quantvoe · targets · Quarto

Resources for Future Reference

Docs acolum.github.io/forkflow

Book acolum.github.io/forkflow/book

Code github.com/acolum/forkflow

Slides github.com/acolum/forkflow → slides.pdf

alyssahcolumbus@gmail.com

alyssacolumbus.com