

What Do We Need the Data For?

A data user's perspective on the inpatient sector

Esra Eren Bayindir

Hamburg Center for Health Economics, Universität Hamburg

IQWiG-Datensymposium · Köln · 10 September 2026

What we can answer today · What constrains us · What better patient data would deliver



The hospital data are public. Patient data are the constraint.

Minimum volumes, service groups, health care market restructurings — many reform questions turn on what happens to patients.

The hospital component

open and complete

Ownership, beds, department-level data on admissions, procedure, physician and nurse staffing, medical technology.

Published annually for every site, with persistent provider IDs (Institutionskennzeichen).

The patient component

fragmented and constrained

Reform evaluation requires longitudinal patient-level trajectories following individuals across

- Admissions
- Providers
- outpatient–inpatient boundary, and into
- Post-discharge survival

What researchers actually have

Fragmented samples from sickness funds.

- Patient-level for one sickness fund
- Case-level with no follow-up past discharge
- Full clinical depth at a single site.

- The hospital component is publicly available
- The patient component is what many reform questions actually turn on, and it is negotiated one fund at a time
- The binding constraint is the availability of patient data good enough to answer the questions

Example 1: Does hospital consolidation harm quality?

298 verified ownership changes among hospitals with ≥ 25 beds, 2009–2021

- US evidence is largely pessimistic: higher negotiated prices, no quality gain. But there, acquirers can raise prices.
- Germany has fixed DRG prices. An acquirer cannot extract a margin through price, so the question is genuinely open.

What answering it required

Patient component — sickness fund indicators

Risk-adjusted excess adverse event rates, hospital level, 2011–2019, with a consistent risk-adjustment method across all years

Hospital component — quality reports

Owner, beds, departments, department-level staffing, cath labs, stroke units, and location

Treatment definition

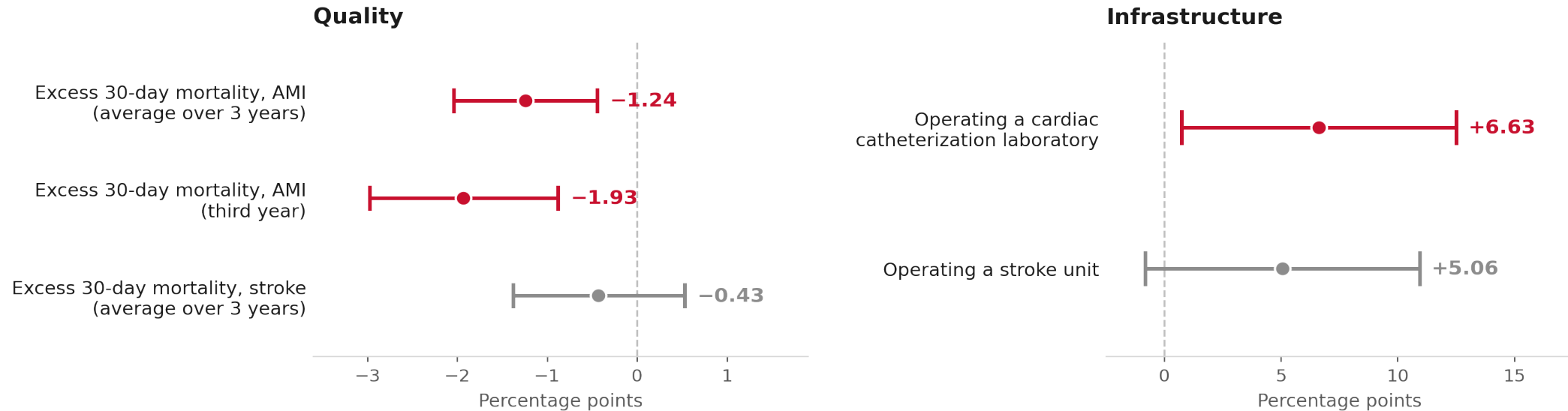
All 298 ownership changes validated against corporate registers and press, to separate real transactions from relabelling

Market definition

Hospital locations geocoded to build local markets and test spillovers onto competing hospitals

What we found

Acquired vs. ownership-stable hospitals, three years pre and post: 125 acquired and 821 control hospitals for AMI.



The mechanism

The mortality improvement coincided with the expansion of catheterization capacity.
 No service lines closed — capital investment, not cost-cutting.

What we did not plan for

Examining the Chefarzt was not in the original design. It became central only once the results were in and we needed to explain why large chains and smaller acquirers behaved differently.

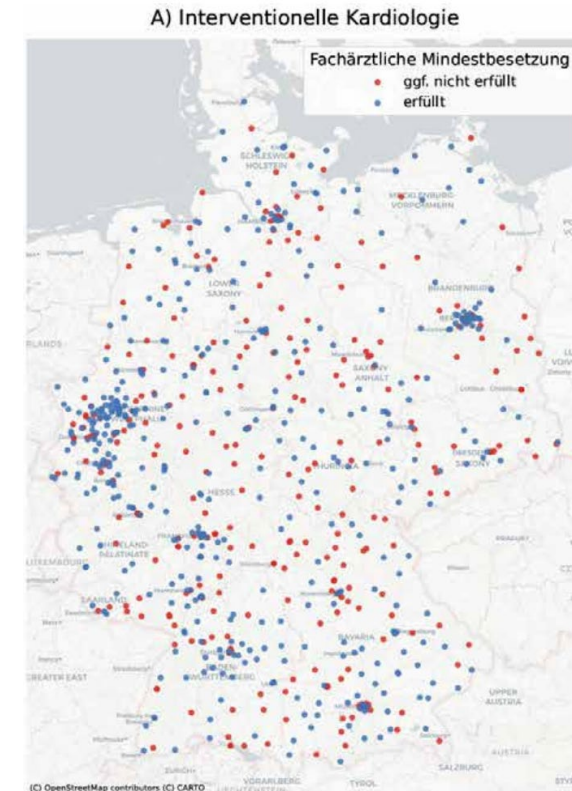
Exploratory research cannot pre-specify every mediating variable before the data are opened.

Example 2: Who can staff the new Leistungsgruppen?

Structured Hospital Quality reports 2023, physician FTE at department level against the minimum staffing thresholds, mapped for every hospital site.

Leistungsgruppe	Sites	Threshold met
Allgemeine Innere Medizin	1,070	98.8%
Allgemeine Chirurgie	1,014	90.3%
Komplexe Gastroenterologie	407	74.4%
Endoprothetik Hüfte	866	69.3%
Interventionelle Kardiologie	724	60.1%
Geriatric	623	58.8%

- Base groups are robust — over 90% of sites clear the physician threshold.
- Specialized groups are the bottleneck, and the dividing line is urban vs. rural, not East vs. West.
- Answered with the public hospital component alone.



The hospital component can map the pressure. It cannot tell us what the pressure does to patients.

Example 3: What only granular clinical data can reveal

21,269 cardiac operations by 46 surgeons at one German heart center, 2010–2023 — how surgeons learn.

- Procedure-specific volume in the preceding quarter is associated with lower 30-day risk-adjusted mortality for CABG: -0.49 pp (95% CI -0.88 to -0.10).
- Analogy-based learning shows up too — prior-quarter volume of technically related valve procedures also moves CABG mortality.
- Cumulative experience lowers the re-thoracotomy rate in mitral valve reconstruction.
- Directly relevant to minimum-volume rules: what matters is recent, procedure-specific practice, not lifetime counts alone.

What made it possible

Surgeon identity, so learning can be traced person by person. EuroSCORE II — comorbidities, smoking status, urgency — a clinical risk model claims data cannot construct.

What limits it

One hospital, so generalizability is limited. To the best of my knowledge, there is no national dataset in Germany carrying this clinical depth attached to provider identity.

Researchers must choose between national scale and clinical depth. We cannot have both.

Langenbeck, Bayindir, Schreyögg, Becker & Gummert, "Surgeon Experience and Outcomes: Specific, Analogy-Based and Cumulative Learning in Cardiac Surgery".

Limitation I: Why single sickness funds cannot answer system-level questions

Not a criticism of the funds — this collaboration is the reason the studies exist at all.

Demographic and regional selection

Sickness funds are not random population samples. AOK skews toward elderly, multimorbid and blue-collar populations with strong rural footprints; TK and BARMER skew younger, urban and higher-earning.

The denominator problem

Even the largest single fund typically covers only 20% to 35% of admissions in a given hospital.

A researcher looking at one fund cannot measure actual hospital case volume, crowding, capacity strain, or market concentration.

Spatial distortion and churn

Cross-regional patient flows and fund-switching break longitudinal tracking. When an insured individual switches funds, their trajectory is truncated.

A population-wide, centrally governed source answers all three.

Limitation II: How research access has worked so far

Provider identity is not always available

In the research extracts we have worked with, hospital identifiers could be dropped.

The consequence was that claims could not be linked deterministically to the public quality reports — staffing, ownership, medical technology — or to county-level infrastructure.

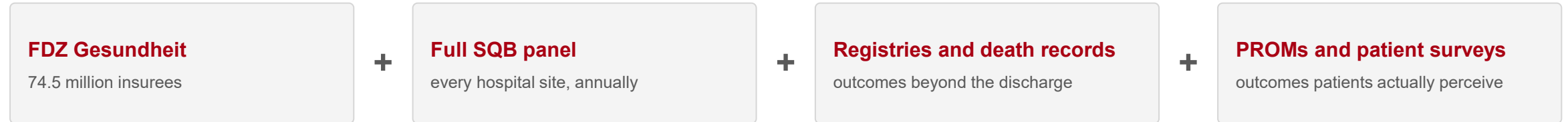
Access ran on a fixed menu

Every diagnosis, procedure code and covariate must be specified at the time of application.

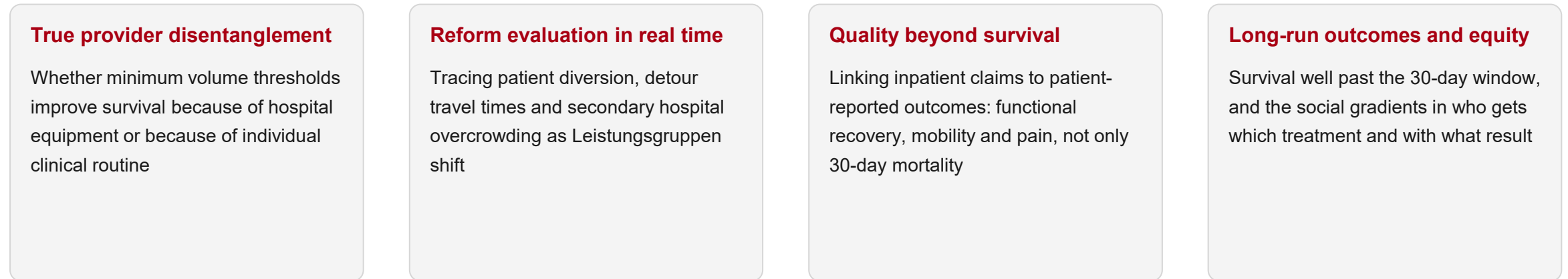
Causal inference is iterative. The variable that explains the result could be the one nobody thought to request.

What better data provision would open up

Other systems have solved parts of this. None of them holds all the inputs Germany already collects.



What this unlocks



With 74.5 million insurees, universal risk structures and complete hospital structural data, Germany would possess the largest and richest health care dataset in the world.

What it would take to unlock Germany's data potential

- 1 The linkage on public provider IDs**
Allow routine, deterministic merging of patient records with public hospital identifiers.
- 2 Population-scale, longitudinal trajectories**
Shift from case-level analysis to individual-level tracking across outpatient, inpatient and post-acute care.
- 3 Move from pre-approval gating to secure research environments**
Replace the “predict your exact variables” model with secure virtual research environments that allow agile data exploration.

Germany already publishes the most transparent hospital structural data in the world

If we connect it to patient-level data, we can make evidence-based policy on many more of these questions for the patients who rely on it