SIMULATION BENCHMARK · TYPE 2 DIABETES

When patients go quiet, claims data get it wrong

Synthetic matched worlds for evaluating medical disengagement: a benchmark in which truth is known by construction

Jérôme Vétillard · VP R&D · Twingital Institute · Revised manuscript R1 (v019i), November 2026

In one sentence: before evaluating, check that the method recovers the truth

In a simulated world where the effect of disengagement is known, the usual claims-based evaluation designs overstate it, dilute it or miss it.

+5.8 pts

10-year mortality

caused by disengagement, under the posited model

−€1,981

payer cost per patient

silence saves money at first

+€3,087

bias on the cost

if only patients who reappear in hospital are studied

Disengagement stacks three invisibilities that no real-world study can lift



The patient vanishes from the data

During the break, they leave no trace at all. They are found only afterwards, often in the emergency department.



Their counterfactual does not exist

The same patient, had they stayed in care, cannot be observed. The comparison is always made with someone else.



The benefit cannot be seen

An averted decompensation is an event that did not happen: unobservable at the individual level.

Consequence

An observational estimate can be matched on twenty covariates, consistent, elegant… and wrong, without anyone being able to tell.

Reverse the logic: generate data around a known truth

1. Causal mechanism

Hand-written cardio-reno-metabolic graph, parameters taken from the literature

2. Observation

Three channels: SNDS-like claims, connected devices, patient app

3. Analyses

The designs a real study could compute, with no access to the truth

4. Grading

Gap between each estimator and the truth, stored separately

Why a mechanistic rather than a learned generator? A model trained on real records reproduces their observation regime, blind spots included. It cannot produce what is missing: the trajectories of patients who have stopped being observed.

Five worlds, one population, the same random draws

REF

Continuous follow-up

Nobody ever disengages

SQ

Status quo

Breaks, spontaneous returns, reconnection after hospital

PA / PB

Re-engagement

Silent patients contacted, two priority rules

PC

IoMT variant

Re-engaged patients equipped with sensors

PE

PREDICARE sequence

Targeting, mobile-unit assessment, monitoring during the break

60,000

people with diabetes

20 years

of monthly trajectory

10 seeds

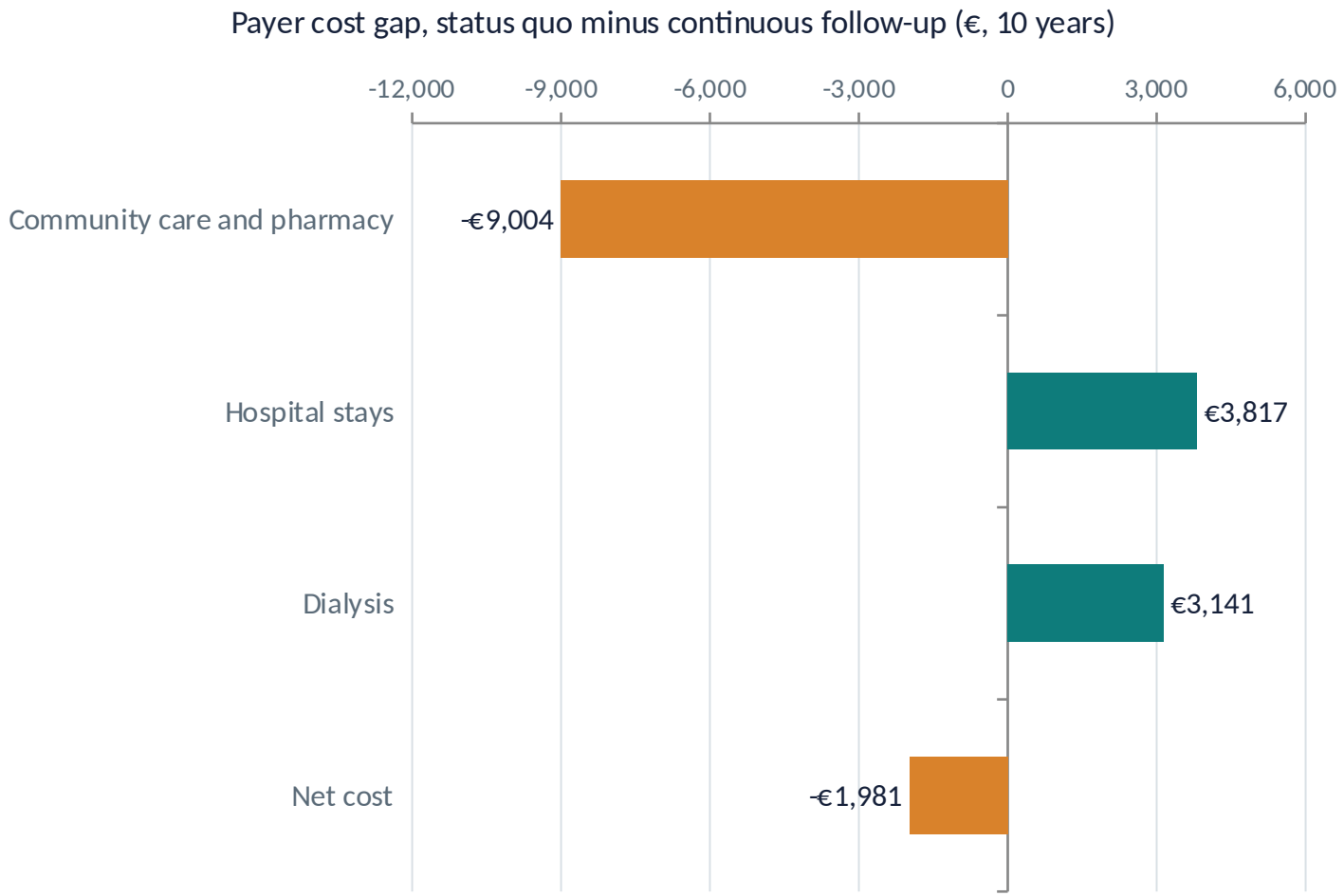
of replication (MCSE)

14

alternative mechanisms

SQ − REF measures the effect of disengagement. Two worlds differ only where their hazards differ: identical for a patient until that patient's first break (verified by tests).

Under the posited model, disengagement costs lives, not money

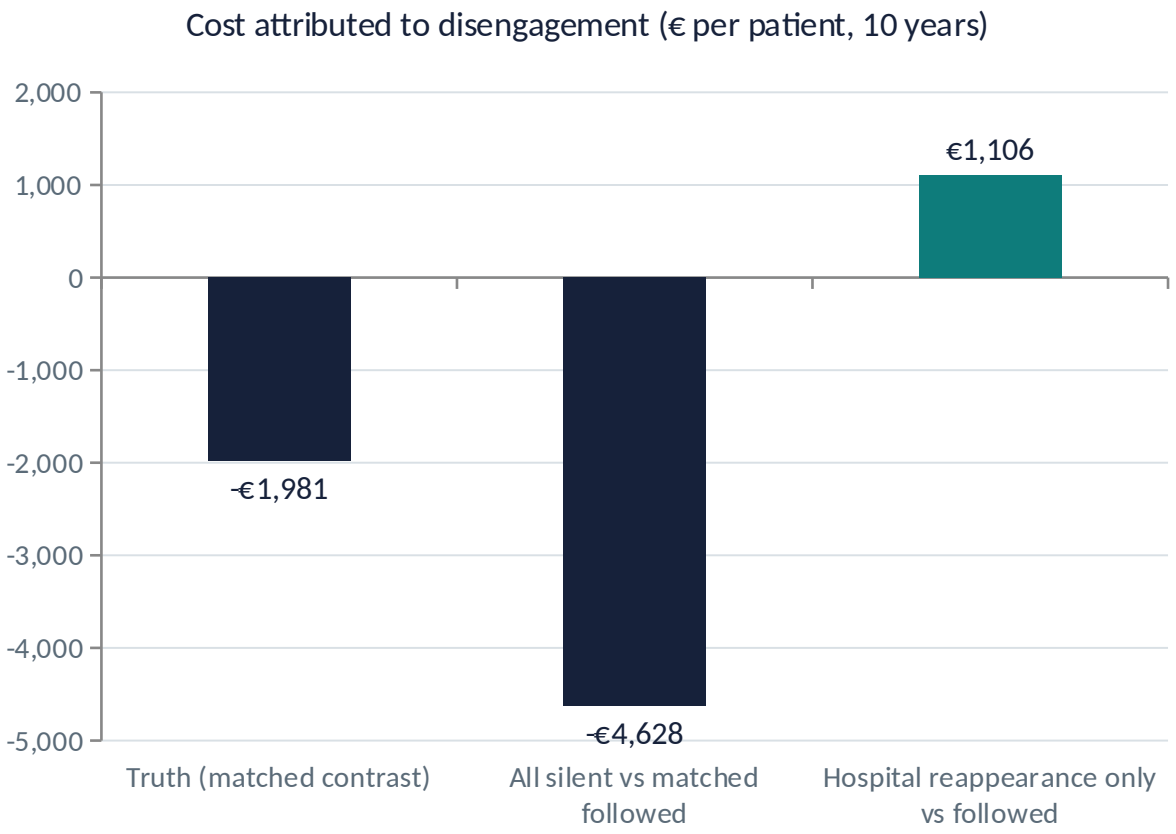


+5.8 points

of 10-year mortality (41.2% vs 35.3%), i.e. 0.42 life-years lost per patient

Silence looks cheap.
It removes useful care that is also paid care. The debt comes due later: in hospital, on dialysis, and in lives lost.

Three common evaluation designs, three ways of being wrong



Selecting on reappearance

Sees only 46% of true breaks and keeps those who did badly: the cost is overstated by €3,087, to the point of reversing its sign.

Intention-to-follow survival

Dilutes the mortality effect: 3.3 points instead of 5.8, because followed patients disengage in turn and disengaged patients come back.

As-treated exposure

Misses delayed mortality: you decompensate during the silence, are reconnected, and die later, counted as "followed".

Only the matched counterfactual measures what was lost.

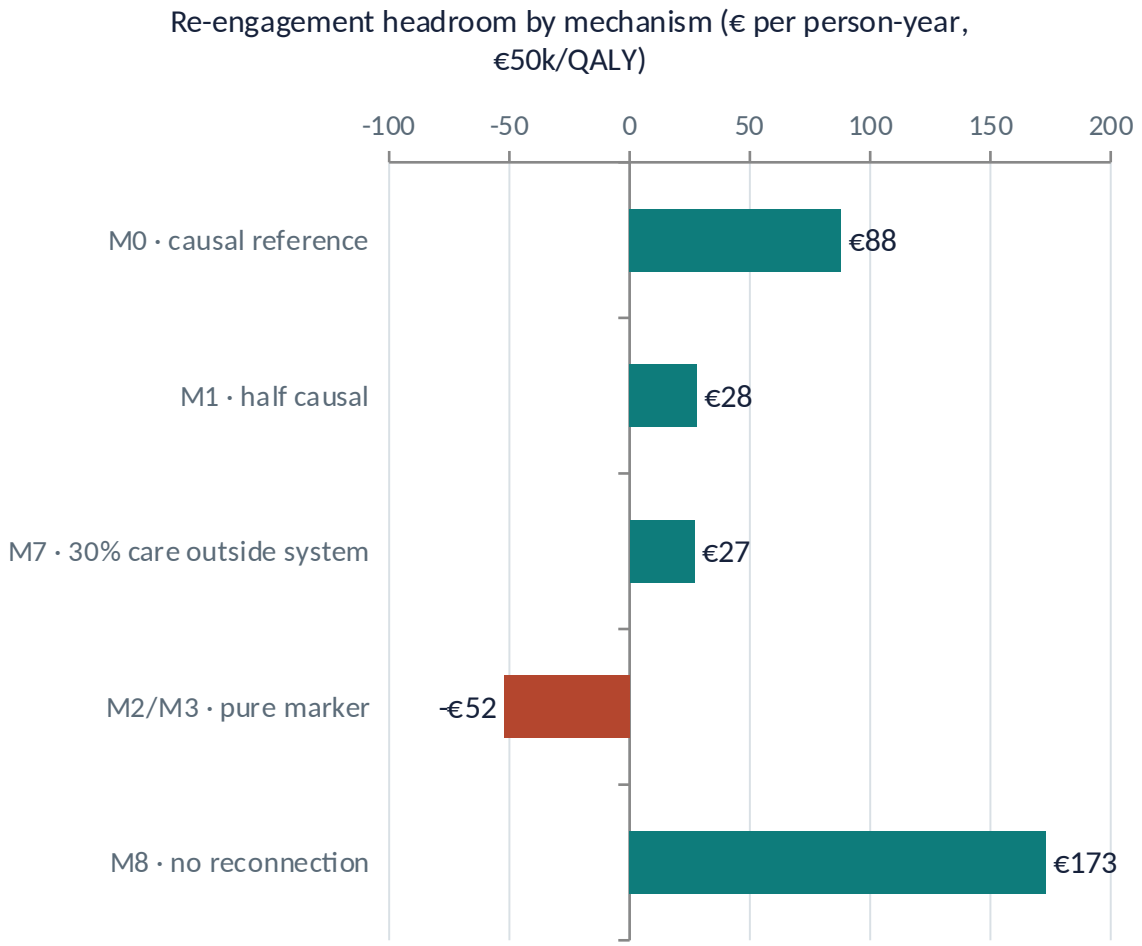
Some biases belong to the method, others to the assumed world

Robust under all 14 mechanisms

The upward bias of reappearance cohorts (139 of 140 mechanism–seed pairs) and the overstated saving of matched comparisons: properties of the design.

Depends on the mechanism

The direction of the mortality bias reverses when disengagement is a marker of severity rather than a cause. A real study cannot tell whether its contrast is too small or too large.



Change the question: not "how much will it save" but "how much may it cost"

The usual question

"How much money will this programme save?"

It collapses if disengagement already saves the payer money.

The question we ask

"What fixed cost remains compatible with a reading threshold per QALY?"

This is the headroom: a conditional ceiling, not a budget.

€88

re-engagement headroom, 20 years, €50k/QALY, per person with diabetes per year

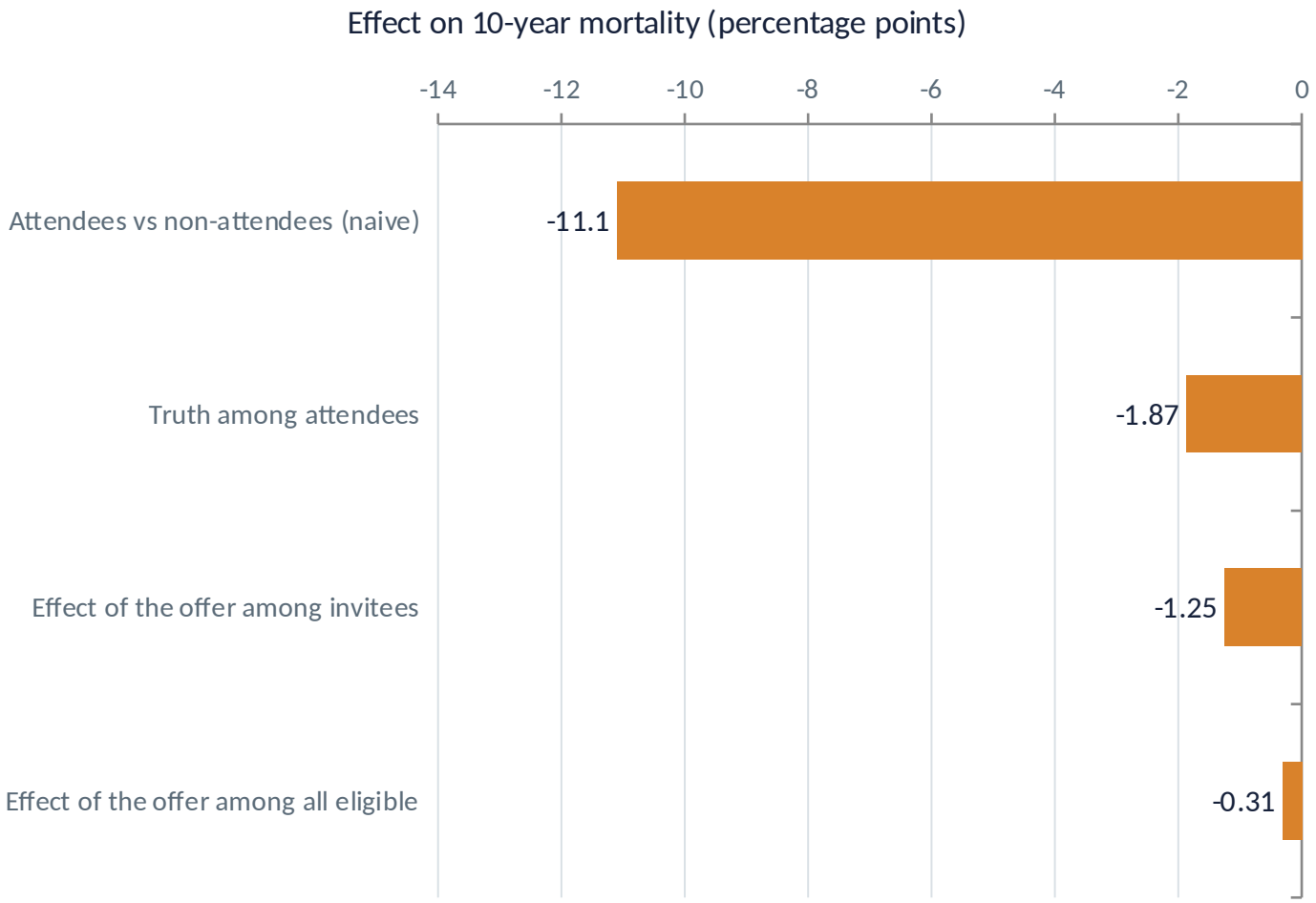
€24

the same headroom over 10 years: health gains arrive late, costs from day one

€13 to €70

headroom of the PREDICARE sequence, depending on unmeasured assumptions

Comparing those who come with those who don't multiplies the effect by six



Why? Those who don't come are sicker: 63% with high latent burden versus 47%.

What must be estimated: the effect of the offer among all eligible people. Under the simulated budget it is diluted to −0.31 points: only 28% of eligible people are invited, 18% attend.

What any demonstrator must do: randomise the offer, not the volunteer patient

Randomise the offer

By territory or practice, before targeting starts, with the same eligibility rule in each arm and its own capacity.

Keep everyone in the denominator

All eligible people, invited or not; report the shares invited and attending alongside.

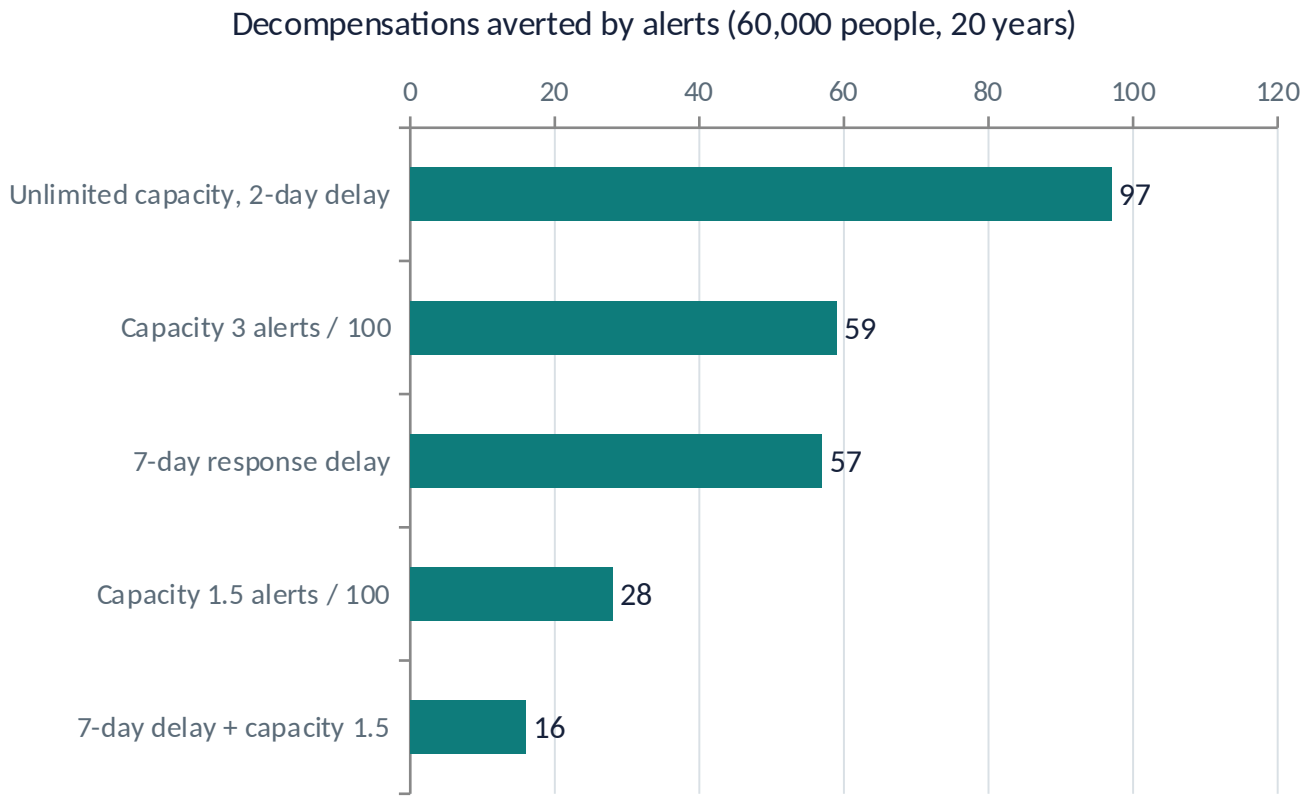
Aim for intermediate endpoints

Resumed follow-up, treatment continuity, decompensations. Mortality alone would need 400,000 eligible people per arm.

Prespecify the estimand

Under shared capacity, territory, place lottery or rule comparison do not estimate the same effect.

In the model, value runs through the care team, not the algorithm



QALY gain per 1,000: 9.4 in the first scenario, 9.3 in the last.

The detector, honestly

0.57 → 0.70

ROC-AUC without then with connected sensors (30-day heart-failure decompensation, internal experiment)

5%

positive predictive value: about 19 alerts per true episode, and poor calibration

What this work does not say, and does not claim to say

Nothing about real-world France

Parameters have the status of assumptions, calibration is pending; some simulated rates (dialysis, myocardial infarction) look high.

Circularity acknowledged

The benchmark predicts what it injected: only differences between designs and between worlds are interpreted, never absolute values.

No simulated learning

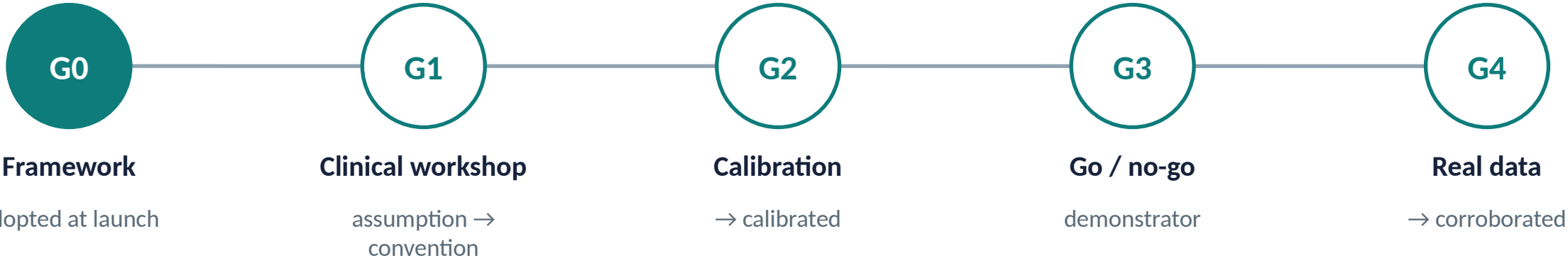
The IoMT registry gain is an imposed curve: training, temporal validation and transfer are not represented.

A selected registry

It would observe about one eligible person in six, and trajectories under the programme, not the natural history of disengagement.

The benchmark shows neither that PREDICARE works nor what it might cost. It shows how to evaluate it without getting it wrong.

The benchmark applies the governance it advocates: no claim outside its domain



An epistemic status changes only through a milestone, never through drafting.

21

components declaring their validity domain, machine-readable

23

sources in the provenance registry, SHA-256 fingerprints

25

architecture decision records

21

automated tests, including a deliberate negative test

Three conclusions, and one method

1

Claims-based studies get medical disengagement wrong

Selecting on reappearance overstates its cost under every mechanism tested; the direction of the mortality bias depends on the mechanism.

2

A vigilance programme is defended as cost per QALY, not as savings

Under an explicit headroom, conditional on the causal share of the excess risk of disengagement.

3

A demonstrator must randomise the offer

And keep every eligible person in the denominator: those who come are not those who stay away.

You cannot grade a method on data where the answer is unknown. You build worlds where it is known.

Manuscript and code: twingital-ventures.com · predicare-synth v0.19, Zenodo 10.5281/zenodo.23475000 (Apache 2.0) · Synthetic data 10.5281/zenodo.23485752 (CC-BY-4.0) · No real patient data