KCOR v7.5 concise full algorithm description

Goal

Our objective is to be able to analyze record-level retrospective data to answer questions related to net mortality risk/benefit for a vaccine intervention at any point in time.

In particular, how can we use the Czech Republic record level data to determine whether the COVID vaccines had a net mortality benefit as of a certain date.

Method

The estimand is the ratio of cumulative hazards of vaccinated vs. unvaccinated cohorts over the timeframe of interest — this is the quantity we want to estimate.

The estimator (KCOR) computes this ratio after correcting for the three sources of bias described below. The ratio is anchored to 1.0 at a reference quiet period post-enrollment, because even after bias correction the cohort baseline hazard levels need not be identical.

This gives us net risk/benefit as a function of t.

So the entire KCOR method can be summarized as follows:

> Using fixed cohorts defined by age and vaccination status at various enrollment dates, the ratio of cumulative hazards as a function of time is an objective way to assess risk/benefit of a vaccine, but before we do that, we need to neutralize the biases so that the cohorts are comparable. We do that bias correction by adjusting the mortality curves of the cohorts as a whole instead of using the more traditional approach requiring 1:1 matching of living people, which has been proven to be too inaccurate for this use (see Obel).

The key problem

In vaccine retrospective data studies, there are 3 sources of bias that must be neutralized before the cohorts can be compared:

1. Dynamic HVE — people who about to die avoid getting vaccinated, artificially lowering early vaccine-group mortality
2. Non-proportional hazards (NPH) — COVID mortality is more age-sensitive than background mortality, so the frailer unvaccinated cohort is hit disproportionately harder during waves
3. Static HVE — vaccinated individuals are systematically healthier than unvaccinated ones at baseline, a large and persistent bias that standard models cannot remove

How KCOR works (high level)

KCOR uses fixed cohorts defined typically after most people for a given dose have been vaccinated.

There is no migration or censoring because we are focused on characterizing the mortality of each cohort which we can't do if the cohort's mortality makeup keeps changing.

Next, we compute the h(t) for the cohorts.

We then apply the COVID NPH correction.

Finally, we neutralize the biases by adjusting the cumulative hazards. Then we take their ratio as a function of t to assess net harm/benefit as a function of t. We normalize to the value of the ratio at t=0.

The key insight

The key insight was that you can't use traditional 1:1 matching of living individuals to create comparable cohorts. Instead, you must observe the mortality over time of the cohort as a whole and neutralize the frailty variance of each cohort before comparing the curves.

The static HVE bias in vaccinated/unvaccinated retrospective observational studies is the elephant in the room. You cannot ignore it or neutralize it with 1:1 matching and Cox models. Most papers never notice that standard models cannot overcome the bias. The Obel authors were careful; they used negative controls and discovered their proportional hazards assumption was violated. There is no known way in the scientific literature to neutralize the bias they observed, so they concluded that RCTs are required to accurately compare cohorts and determine harm or benefit of vaccination.

Instead of 1:1 matching the observed characteristics of living INDIVIDUALS in each cohort to create comparable cohorts, KCOR takes the opposite approach: characterizing how each cohort dies over time.

So instead of 1:1 matching of living people we do cohort-level matching based on how the cohort dies over time. That was the key insight.

This is done by estimating the gamma frailty variance of each cohort using nonlinear regression to determine $\hat{\theta}_d$, the frailty variance at $t = 0$ (the first week after enrollment and the dynamic HVE stabilization skip). This is far easier and much more accurate than any 1:1 matching because it is done at the full cohort level using observed mortality rates, which are objective.

KCOR vs. traditional methods

Here is a quick comparison table comparing KCOR with traditional methods of analysis

Parameter	Traditional approach	KCOR
Estimand	Hazard Ratio (HR)	ratio of adjusted cumulative hazards

Parameter	Traditional approach	KCOR
Enrollment	Variable cohorts	Fixed cohorts defined based on age, vaccination status at enrollment
Cohort transitions	Allowed	Forbidden. Allowing transitions would change the mortality curve over time in unpredictable ways.
Censoring	Allowed.	Forbidden. If someone gets another shot, they are NOT removed from the cohort because that would change the mortality of the group that remains. This means any KCOR signal is conservative: if KCOR finds harm, the true harm is larger; if KCOR finds benefit, the true benefit was larger.
Comparability	1:1 matching of individuals based on observed living characteristics	Cohort-level matching based on how the cohort dies over time

Dynamic HVE bias

Dynamic HVE is the time-varying bias caused by people who are about to die avoiding vaccination — e.g., they are in a hospital or hospice. This is an exponentially decaying effect with a half-life of around 4 days. It causes death rates for approximately two weeks post-vaccination in the vaccinated group to be artificially low.

The graph below is from KCOR_ts_analysis.xlsx file in the KCOR github (see "dynamic HVE" tab)

Czech data. Time series h(t). All ages. Dose 2 in Feb 2021. The 2 week static HVE is obvious.
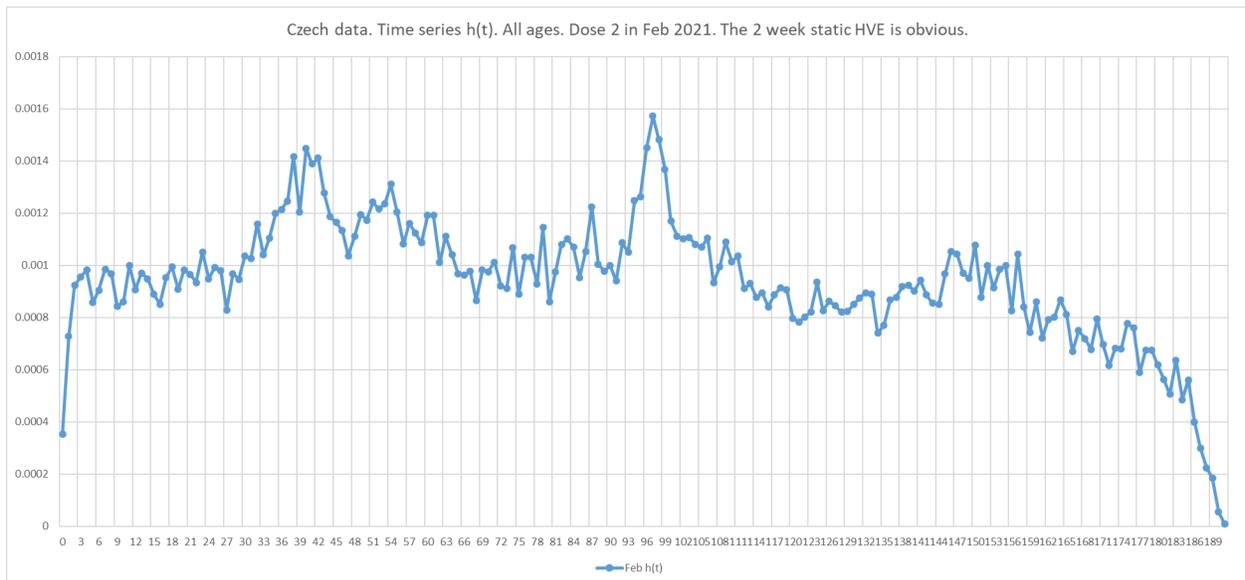
The adjustment here is simple: wait two weeks before cumulating hazards.

NPH bias

Non-proportional hazards (NPH) arise because COVID mortality is more age-sensitive than all-cause background mortality. The unvaccinated cohort contains a higher fraction of frail, older individuals (due to static HVE), so the unvaccinated cohort dies disproportionately faster during COVID waves than would be expected from its elevated baseline mortality alone. If uncorrected, this makes COVID waves appear to benefit vaccinated cohorts more than they actually did.

Deriving the NPH correction factor

All-cause mortality follows the Gompertz law, growing at $\gamma = 8.5\%$ per year with age.

COVID-specific mortality is steeper. The Levin meta-analysis of the IFR 27 studies shows COVID mortality grows at approximately 12.7% per year with age. The ratio of these slopes gives the amplification factor:

$$\text{amplification} = \frac{12.7\%}{8.5\%} = 1.49$$

This means a cohort that is R times frailer on background mortality (i.e., R times more likely to die from non-COVID causes) will be $R^{1.49}$ times more likely to die from COVID — an excess amplification of $R^{0.49}$ over and above the R-fold baseline hazard ratio.

Translating amplification into a correction multiplier

The NPH correction factor depends on the assumed static HVE — the mortality ratio between unvaccinated and vaccinated cohorts due to selection alone. Let R be the static HVE ratio (e.g., $R = 2$ means unvaccinated die at twice the rate of vaccinated in the absence of any vaccine effect or COVID). The NPH correction multiplier applied to unvaccinated excess COVID mortality is:

$$f_{\text{NPH}} = R^{(12.7/8.5 - 1)} = R^{0.49}$$

For two plausible HVE assumptions:

Static HVE (R)	Multiplier $R^{0.49}$	Interpretation
5× (upper bound from Mirror of Erised)	$5^{0.49} \approx 2.2$	Unvaccinated COVID mortality scaled down by 2.2× during waves
2× (conservative lower bound)	$2^{0.49} \approx 1.4$	Unvaccinated COVID mortality scaled down by 1.4× during waves

The analyses in this paper use the conservative $R = 2$ estimate, yielding a **1.4× NPH correction factor**.

How the correction is applied

The correction is applied **only during COVID epidemic waves** — calendar periods where COVID deaths constitute a material fraction of all-cause mortality. Outside of waves, COVID contributes negligibly to mortality and no adjustment is warranted; applying a correction during quiet periods would introduce bias rather than remove it.

During wave periods, the unvaccinated cohort's observed hazard contribution to the cumulative hazard is scaled down by the NPH factor before computing the KCOR ratio. This removes the differential amplification that would otherwise make wave mortality protection appear larger than it is in the vaccinated cohort.

Static HVE bias

The elephant in the room is neutralizing static HVE caused by fundamental differences between people who choose to be vaccinated (generally healthier, better access to healthcare, higher socioeconomic status, more likely to follow doctor and government guidance) and those who choose to be unvaccinated. There is often a 2X to 5X mortality difference between cohorts with the same age due to this effect. See the Mirror of Erised paper which notes:

> As presented above, the risk of death from non-COVID causes was up to five times lower among vaccinated individuals during periods with negligible COVID-19 mortality. This implies a risk ratio (entirely attributable to the HVE) close to 0.2, corresponding to an apparent vaccine effectiveness of approximately 80% against non-COVID mortality.

But there isn't just a difference in mortality. There is also a huge (and very unappreciated) difference in the gamma frailty variance of the cohorts.

This bias is readily observable if you are looking for it. It goes unnoticed if you are not.

For example, in Palinkas et al. (*Effectiveness of COVID-19 Vaccination in Preventing All-Cause Mortality among Adults during the Third Wave of the Epidemic in Hungary*), the Kaplan-Meier curves in Fig 1 and 2 for the unvaccinated (in red) curve up instead of down.
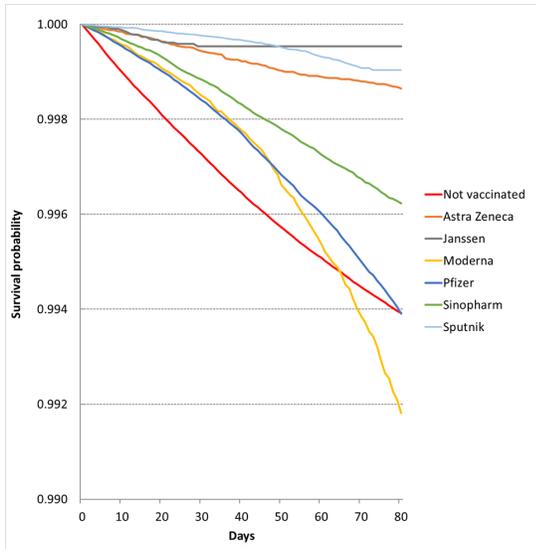
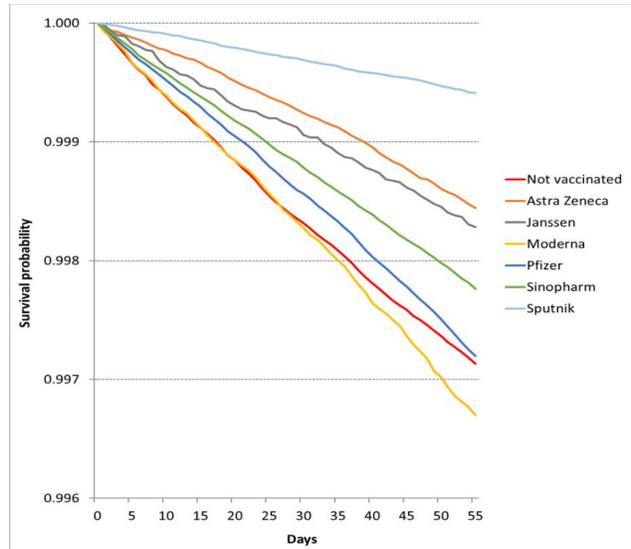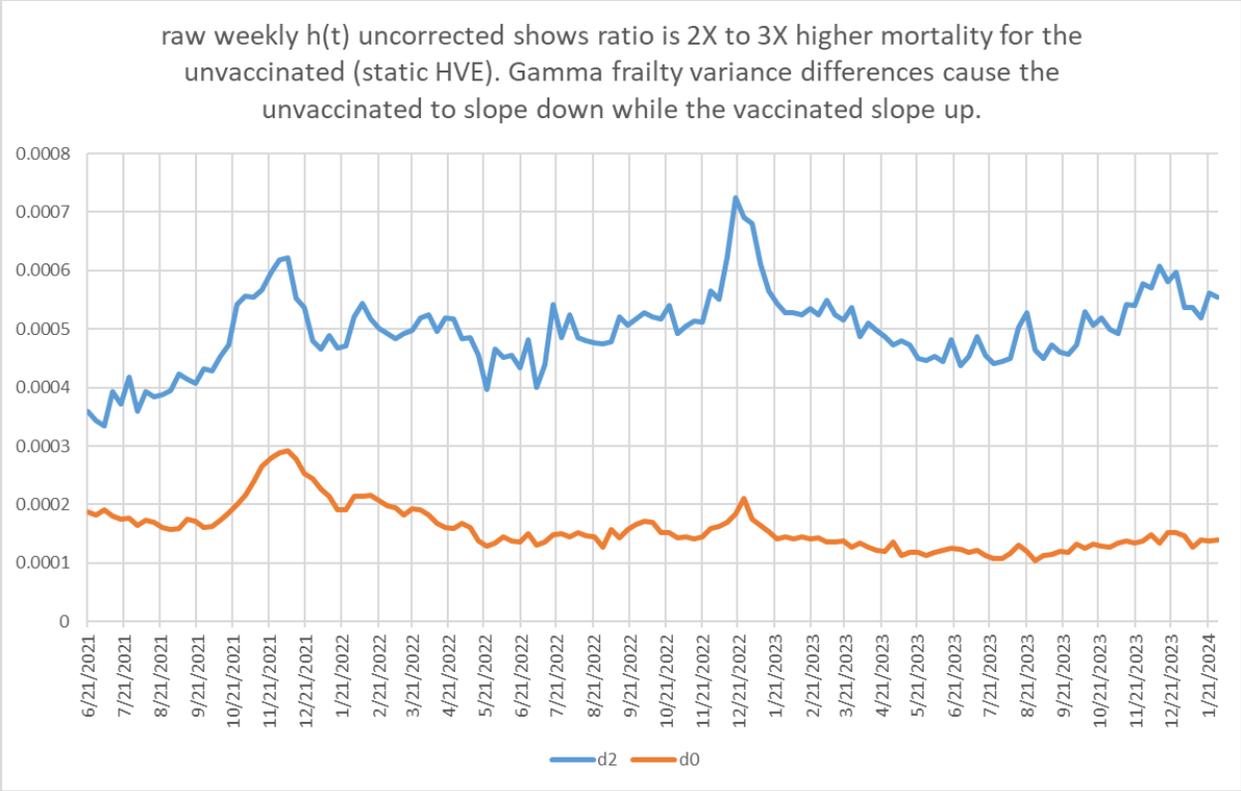**Figure 1.** Survival probabilities in vaccination cohorts in the epidemic period.

Figure 2. Survival probabilities in vaccination cohorts in the nonepidemic period.

Secondly, for the Czech data, the h(t) for unvaccinated (d0) slopes down over time while the h(t) for the vaccinated (d2) slopes up. This is from the "primary shots h(t)" tab of the KCOR_v7.5_with_1.27 correction.xlsx in the KCOR github. This is for all ages (value=-2).

There are 4 things to note in this graph: 1. The vaccinated h(t) slopes up (minimal theta) 2. The unvaccinated h(t) slopes down (significant theta) 3. There is a step increase in the h(t) of the vaccinated cohort (fixed cohort defined on 6/14/2021) that happened during COVID/booster shots, so this is not selection bias since it is the same cohort. This is the mortality signal KCOR picks up on for the booster enrollment date. It's easy to see in the raw data here. 4. Because these are all ages cohorts and not a single narrow age range, the unvaccinated here have lower h(t) because the age mix is younger. Had we picked a narrow age range, the unvaxxed would have 2x or higher h(t).

raw weekly h(t) uncorrected shows ratio is 2X to 3X higher mortality for the unvaccinated (static HVE). Gamma frailty variance differences cause the unvaccinated to slope down while the vaccinated slope up.

To neutralize both the level and slope changes caused by static HVE bias, we use the gamma frailty neutralization built into KCOR.

KCOR method full detail

KCOR (Kirsch Cumulative Outcomes Ratio) is a depletion-neutralized cohort comparison framework that removes selection-induced frailty bias before computing cumulative outcome ratios. It operates on individual registry records — dates of birth, intervention, and death — aggregated into weekly cohort death counts for analysis, without requiring individual-level covariates such as comorbidities or socioeconomic status. The method proceeds in eight steps.

---

1. Fixed cohorts

Individuals are assigned to cohorts at enrollment based on intervention status (e.g., dose received). **No post-enrollment switching or censoring is permitted.** Individuals remain in their enrollment cohort for the full follow-up. This prevents immortal-time artifacts and keeps frailty composition stable within each cohort.

---

2. Discrete-time hazard and cumulative hazard

For each cohort d and weekly interval t, the discrete-time hazard is:

$$h_{\mathrm{obs},d}(t) = -\ln\left(1 - \frac{d_d(t)}{N_d(t)}\right)$$

where $d_d(t)$ is deaths in interval t and $N_d(t)$ is the risk set.

Cumulative hazards are accumulated from the **NPH-corrected effective hazard** $h_d^{\mathrm{nph}}(t)$ (defined in steps 4–5) — the observed hazard after applying the stabilization skip (step 4) and NPH wave correction (step 5):

$$H_{\mathrm{obs},d}(t) = \sum_{s \leq t} h_d^{\mathrm{nph}}(s)$$

---

3. The frailty depletion problem

Under gamma frailty, individuals with latent frailty $z_{i,d}$ have individual hazard $h_{i,d}(t) = z_{i,d}\,\tilde{h}_{0,d}(t)$. As high-frailty individuals die first, the surviving risk set becomes progressively healthier, causing the cohort-level hazard to decelerate relative to the true baseline. The gamma-frailty model links the observed and baseline cumulative hazards via:

$$H_{\mathrm{obs},d}(t) = \frac{1}{\theta_d}\ln\left(1 + \theta_d\,\tilde{H}_{0,d}(t)\right)$$

where $\theta_d \geq 0$ is the **frailty variance of cohort d at enrollment $(t = 0)$** — the variance of the initial gamma distribution before any depletion has occurred — and $\tilde{H}_{0,d}(t)$ is the depletion-neutralized baseline cumulative hazard. θ_d is a fixed scalar: it characterizes the cohort's frailty heterogeneity at the moment cohorts are defined. The time-varying narrowing of the surviving population's frailty distribution as high-frailty individuals die is captured entirely by $H_{\mathrm{obs},d}(t)$ in the inversion formula — not by a time-varying θ_d. When θ_d differs across cohorts — as it does when selection concentrates or depletes frail individuals at enrollment (static HVE) — naive cumulative comparisons are biased and hazards are non-proportional.

---

4. Stabilization skip

The first `SKIP_WEEKS` intervals after enrollment are excluded from accumulation to remove transient post-enrollment artifacts (dynamic HVE):

$$h_d^{\text{eff}}(t) = \begin{cases} 0, & t < \text{SKIP\_WEEKS} \\ h_{\text{obs},d}(t), & t \geq \text{SKIP\_WEEKS} \end{cases}$$

The time axis for frailty fitting is rebased so $t_{\text{rebased}} = 0$ at the first accumulating week:

$$t_{\text{rebased}} = t_{\text{raw}} - \text{SKIP\_WEEKS}$$

All frailty fitting uses t_{rebased} so that $H_{\text{gom},d}(0) = 0$ exactly at the first real data point.

---

5. NPH hazard correction

During COVID epidemic waves, the unvaccinated cohort's hazard is amplified relative to the vaccinated cohort beyond what their baseline frailty difference alone predicts. This occurs because COVID mortality is more age-sensitive than all-cause background mortality.

the Gompertz slope for COVID deaths is 11.5%/yr versus 8.5%/yr for all-cause mortality, an amplification of $11.5/8.5 = 1.35$. The frail-enriched unvaccinated cohort therefore dies disproportionately faster during waves, which would otherwise make wave-period mortality protection in vaccinated cohorts appear larger than it truly is.

The NPH correction factor $f_{\text{NPH}} = R^{0.35}$ depends on the assumed static HVE ratio R (the mortality ratio between unvaccinated and vaccinated cohorts attributable to selection alone):

Static HVE (R)	$f_{\text{NPH}} = R^{0.35}$
5× (upper bound)	≈ 1.75
2× (conservative)	≈ 1.27

The analyses in this paper use $R = 2$, giving $f_{\text{NPH}} = 1.27$.

The correction scales down the unvaccinated cohort's effective hazard **only during prespecified COVID wave calendar weeks** $\mathcal{W}$ — intervals identified a priori from Czech Republic COVID death curves where epidemic mortality constitutes a substantial fraction of all-cause deaths. The wave windows are fixed before examining KCOR output:

$$h_d^{\text{nph}}(t) = \begin{cases} h_d^{\text{eff}}(t) \, / \, f_{\text{NPH}}, & t \in \mathcal{W} \text{ and } d = d_{\text{unvax}} \\ h_d^{\text{eff}}(t), & \text{otherwise} \end{cases}$$

Outside wave periods, $h_d^{\text{nph}}(t) = h_d^{\text{eff}}(t)$ for all cohorts. This corrected hazard $h_d^{\text{nph}}(t)$ is what is accumulated into $H_{\text{obs},d}(t)$ in step 2 and used for all subsequent fitting and inversion. Because theta

estimation (step 6) uses only quiet-window timepoints — by construction outside $\mathcal{W}$ — the NPH correction does not affect the frailty estimates.

Why NPH correction is applied before the frailty inversion, not after. The gamma frailty model assumes both cohorts' mortality follows the same underlying Gompertz shape, differing only in frailty variance θ_d. NPH violates this assumption: COVID mortality has a steeper age-sensitivity than background mortality, so the unvaccinated cohort's hazard during waves does not conform to the shape the frailty model expects. Correcting $h(t)$ first restores this shape assumption before the model runs. Applying the correction afterward — directly to $\tilde{H}_0$ — would not be equivalent: the gamma inversion contains an exponential in H_{obs}, so inflated wave hazards fed into the inversion are amplified nonlinearly and cannot be undone by a simple post-hoc division. The two operations do not commute through an exponential.

The correction is only applied for the first COVID wave after the groups are defined (Delta/Omicron). The groups are more similar in COVID frailty after these two depletion waves because the most frail succumbed making it less necessary to use the correction after that point.

---

6. Estimating $\hat{\theta}_d$: Delta-Iteration method

The key insight of v7 is that θ_d — the **enrollment-time frailty variance** — must be estimated from data anchored at $t_{\text{rebased}} = 0$, not from a late quiet window that reflects an already-depleted population.

Human all-cause mortality follows the Gompertz law $\tilde{h}_{0,d}(t) = k_d\, e^{\gamma t}$ with $\gamma = 0.085$ per year (≈ 0.00163 per week), a fixed biological constant. The correct marginal hazard for a fixed cohort under gamma frailty with Gompertz baseline is:

$$h_{\text{obs},d}(t) = \frac{k_d\, e^{\gamma\, t_{\text{rebased}}}}{1 + \theta_d\, H_0^{\text{eff}}(t_{\text{rebased}})}$$

where during quiet periods $H_0^{\text{eff}}(t) = H_{\text{gom},d}(t)$, the Gompertz cumulative hazard:

$$H_{\text{gom},d}(t_{\text{rebased}}) = \frac{k_d}{\gamma}\left(e^{\gamma\, t_{\text{rebased}}} - 1\right)$$

The wave-offset problem

During an epidemic wave, the population faces elevated mortality, accumulating excess cumulative hazard beyond H_{gom}. After the wave ends this excess persists permanently — the wave has fast-forwarded the population's frailty depletion clock:

$$H_0^{\text{eff}}(t) = H_{\text{gom}}(t) + \delta \quad \text{for all } t \geq t_{\text{wave end}}$$

Fitting quiet windows before and after a wave using the same $H_{\text{gom}}(t)$ — ignoring δ — causes the optimizer to inflate $\hat{\theta}_d$ to compensate, producing **+14% upward bias** on post-wave quiet windows. The delta-iteration method corrects this.

Step 1: Joint fit of $\hat{k}_d$ and $\hat{\theta}_d^{(0)}$ from the nearest quiet window

Both k_d and θ_d are fitted jointly from the **nearest quiet window after enrollment** — the first available quiet interval — using two-parameter nonlinear least squares with no delta correction:

$$(\hat{k}_d,\, \hat{\theta}_d^{(0)}) = \arg \min_{k \geq 0,\, \theta \geq 0} \sum_{t \in W_0} \left[h_d^{\text{nph}}(t) - \frac{k\, e^{\gamma t}}{1 + \theta\, H_{\text{gom},d}(t)} \right]^2$$

where W_0 is the nearest quiet window after enrollment. This gives both the baseline hazard $\hat{k}_d$ (held fixed for all subsequent steps) and a seed $\hat{\theta}_d^{(0)}$ to initialize the iteration. The "nearest quiet window" is simply the closest quiet interval to enrollment — it need not precede any wave.

Step 2: Iterate to convergence

Repeat the following until $\hat{\theta}_d$ converges (typically 2–3 iterations):

2a. Reconstruct H_0^{eff} via frailty inversion. Using the gamma-frailty relationship, compute the individual-level cumulative hazard from the observed cohort hazard:

$$h_0^{\text{eff}}(t) = h_d^{\text{nph}}(t) \cdot \left(1 + \hat{\theta}_d \cdot H_0^{\text{eff}}(t)\right), \qquad H_0^{\text{eff}}(t+1) = H_0^{\text{eff}}(t) + h_0^{\text{eff}}(t)$$

starting from $H_0^{\text{eff}}(0) = 0$. This reconstruction passes through **all weeks** — quiet and wave alike — with no special cases.

2b. Check delta sign (first iteration only). After the first reconstruction pass, compute each wave's provisional $\delta_i = H_0^{\text{eff}}(t_i) - H_{\text{gom}}(t_i)$ at wave end time t_i. If any $\delta_i < 0$:

- Log `DELTA_INAPPLICABLE` and skip to **Fallback** below
- Do not clamp and continue — a clamped negative delta produces a systematically wrong H_0^{eff} for all post-wave time points, corrupting $\hat{\theta}_d$

2c. Compute each wave's δ_i (incremental). For wave i with end time t_i, the incremental excess is:

$$\delta_i = \left[H_0^{\text{eff}}(t_i) - H_{\text{gom}}(t_i) \right] - \sum_{j < i} \delta_j$$

This is a **direct subtraction** — **not a fit.** The accumulated offset at time t is:

$$\Delta(t) = \sum_{i:\, t_i \leq t} \delta_i$$

2d. Refit $\hat{\theta}_d$ on all quiet windows with accumulated deltas. With δ_i values known, fit a single parameter to all quiet windows jointly:

$$\hat{\theta}_d = \arg\min_{\theta \geq 0} \sum_{t \in \bigcup_j W_j} \left[h_d^{\text{nph}}(t) - \frac{\hat{k}_d\, e^{\gamma t}}{1 + \theta \cdot \left(H_{\text{gom},d}(t) + \Delta(t) \right)} \right]^2$$

where W_j are all quiet windows and $\Delta(t)$ is the accumulated delta offset up to t.

Insufficient signal detection $->$ set theta to 0   When `theta0_raw` hits the upper bound (or equivalently when delta comes out negative as a consequence of theta overshoot), set `theta_applied = 0` and log as `INSUFFICIENT_SIGNAL`. The reasoning:

- These are young cohorts where frailty depletion over 3 years is negligible
- theta=0 means no HVE correction, which is correct — there's nothing to correct for
- The KCOR ratio is then computed on raw cumulative hazard, which is appropriate when frailty variance is unidentifiable ### Properties

Property	Delta-Iteration path	Fallback path
Applicable when	$\delta > 0$ (dose=0, unvaccinated cohorts)	$\delta \leq 0$ (vaccinated/selected cohorts)
Step 1 free parameters	2 (k_d and θ_d, nearest quiet window)	Same
Iteration free parameters	1 (θ_d only; k_d fixed after Step 1)	N/A
Wave offsets computed	N (δ_i per wave, analytically)	0
Iterations to convergence	2–3	1 (single fit)
Bias with perfect data	$< 0.01\%$	$\approx 0\%$ (nearest quiet window only)
Log message	(none — normal path)	`DELTA_INAPPLICABLE`

Identifiability. The estimator requires sufficient curvature in $h(t)$ over the observation window to distinguish θ_d from k_d. The total curvature is governed by annual_mortality $\times\, n_{\text{years}} \times \theta_d \times \gamma$. For cohorts where $\hat{k}_d$ is very small relative to γ, the cohort is flagged `GOMPERTZ_UNIDENTIFIABLE` and

$\hat{\theta}_d$ is set to zero (no correction applied). If the optimizer returns $\hat{\theta}_d > \theta_{\max}$ for other reasons, it is flagged `DEGENERATE` and similarly zeroed. Both conditions are logged explicitly.

---

7. Normalization (gamma inversion)

Given $\hat{\theta}_d$, the depletion-neutralized baseline cumulative hazard is obtained by exact inversion of the gamma-frailty identity, applied to the full post-enrollment trajectory of $H_{\mathrm{obs},d}(t)$:

$$\tilde{H}_{0,d}(t) = \frac{\exp\left(\hat{\theta}_d \, H_{\mathrm{obs},d}(t)\right) - 1}{\hat{\theta}_d}$$

implemented as `np.expm1(theta * H_obs) / theta` for numerical stability. When $\hat{\theta}_d = 0$ this reduces to $\tilde{H}_{0,d}(t) = H_{\mathrm{obs},d}(t)$ (identity; no correction). Note that $H_{\mathrm{obs},d}(t)$ here is the full observed cumulative hazard including wave periods — not restricted to quiet windows — so all real depletion, including epidemic-wave-accelerated depletion of frail individuals, is correctly incorporated into the correction.

---

8. KCOR ratio

Cohorts are compared via the ratio of depletion-neutralized cumulative hazards, anchored to 1.0 at a reference quiet window $[t_0, t_0 + W_{\mathrm{norm}}]$ post-enrollment:

$$\mathrm{KCOR}(t) = \frac{\tilde{H}_{0,A}(t)}{\tilde{H}_{0,B}(t)}, \qquad \mathrm{KCOR}_{\mathrm{norm}}(t) = \frac{\mathrm{KCOR}(t)}{\overline{\mathrm{KCOR}}_{[t_0, t_0 + W_{\mathrm{norm}}]}}$$

where $\overline{\mathrm{KCOR}}_{[t_0, t_0 + W_{\mathrm{norm}}]}$ is the mean KCOR over the `NORMALIZATION_WEEKS` ($W_{\mathrm{norm}} = 4$) reference window immediately following the anchor period. This anchoring accounts for any residual level difference between cohorts after frailty correction.

A flat trajectory ($\mathrm{KCOR}_{\mathrm{norm}}(t) \approx 1$) indicates no divergence between cohorts after depletion normalization. Deviations from unity represent cumulative outcome differences that persist after frailty correction.

Key parameters

Parameter	Default	Role
SKIP_WEEKS	2	Dynamic HVE stabilization skip

Parameter	Default	Role
`k_anchor_weeks`	4	Weeks post-skip used to fix $\hat{k}_d$
`gompertz_gamma`	0.085/yr	Gompertz aging constant (fixed biological constant)
`theta0_max`	100	Degenerate/unidentifiable fit threshold
`min_quiet_deaths`	30	Minimum deaths in quiet windows to attempt fit
`NORMALIZATION_WEEKS`	4	Reference window for KCOR anchoring
`theta_estimation_windows`	see below	Quiet calendar intervals (ISO year-week)

Theta estimation windows (quiet calendar intervals, ISO year-week):

- '2021-26', '2021-36' — post primary
- '2022-22', '2022-25' — post booster
- '2023-19', '2023-40' — post COVID
- '2024-11', '2024-14' — post 2024

What changed from v6 to v7

The gamma inversion (step 7) and KCOR ratio (step 8) are **unchanged**. The only change is step 6 — how $\hat{\theta}_d$ is estimated:

Property	v6	v7
Baseline model	Flat k_d (constant)	Gompertz $k_d e^{\gamma t}$
Free parameters (initial fit)	2 (k_d and θ_d)	2 (k_d and θ_d, nearest quiet window)
Free parameters (iteration)	N/A	1 (θ_d only; k_d fixed after initial fit)
$\hat{k}_d$ source	Jointly fitted with late data — confounded with $\hat{\theta}_d$	Jointly fitted from nearest quiet window after enrollment; held fixed
Estimation windows	Single late quiet window	All quiet windows simultaneously (with wave-offset correction)
$\hat{\theta}_d$ anchored at	Late depleted population — underestimates true enrollment-time θ_d	Enrollment ($t_{\text{rebased}} = 0$) — correct enrollment-time θ_d
Vaccinated $\hat{\theta}_d$	Near zero by construction	Non-zero where frailty signal exists
Monotonicity of $\hat{\theta}_d$ by age	Frequently violated	Holds empirically (YoB 1920–1970)

Estimating θ_0 simulator results

Using synthetic mortality data with two epidemic waves (104-week follow-up), we swept across 400 parameter combinations — 20 annual mortality rates (1%–20%/yr) × 20 theta values (1–20) — and recovered θ_0 using the delta-iteration method. The estimator converges in 2–3 iterations across all tested parameter combinations.

The error in estimating θ_0 is not an iteration problem; it is a **geometric identifiability problem**. The estimator requires the $h(t)$ curve to have enough curvature over the observation window for the optimizer to distinguish θ_d from k_d. This curvature is governed by:

$$\text{curvature} \propto \text{annual\_mortality} \times n_{\text{years}} \times \theta \times \gamma$$

When this product is small — low mortality, short follow-up, small theta — the $h(t)$ curve is nearly flat and k_d and θ_d become nearly collinear. This is a fundamental geometric property of the model, not a limitation of the algorithm, and cannot be overcome by any estimator without additional assumptions.

Identifiability sweep results

Follow-up	Worst-case error	Problem region
2 years (104 weeks)	4.93%	mortality < 3%/yr AND theta > 10
3 years (156 weeks)	3.49%	mortality < 2%/yr AND theta > 12
4 years (208 weeks)	2.14%	mortality < 2%/yr AND theta > 15

Outside the low-mortality/high-theta corner, errors are consistently < 0.5% regardless of follow-up length. For typical all-cause mortality cohorts ($\geq 4\%$/yr) with theta in the range 1–10 and 2+ years of follow-up, the estimator is well-identified with errors well under 1%. The Czech Republic registry data ($\approx 6\%$/yr, 2+ year follow-up) falls comfortably in the well-identified zone.

Delta-Iteration Theta0 Estimator — 2-Wave Simulation
True θ=2.0   Estimated θ=2.00053929   Error=+0.026965%
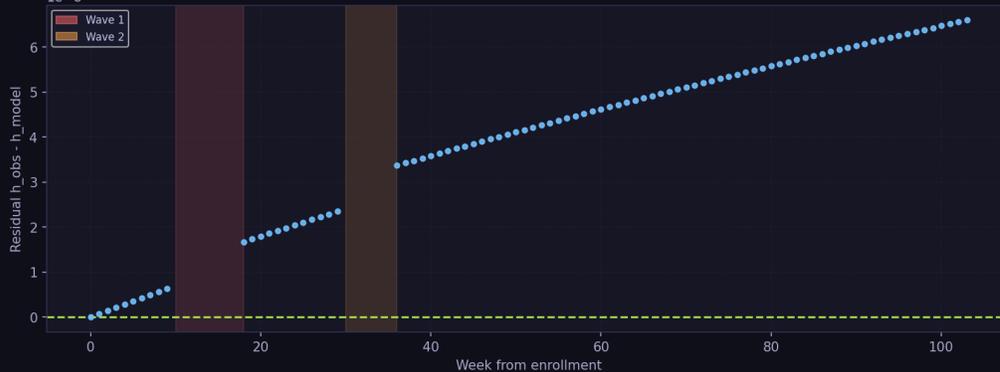
(a) Observed hazard and fitted model
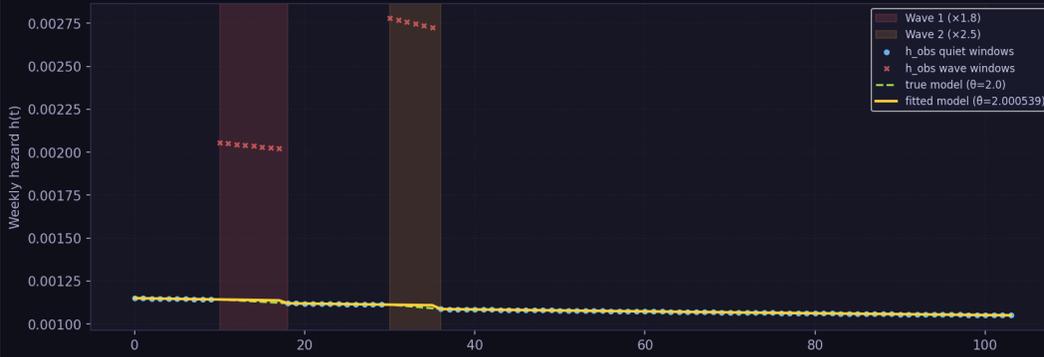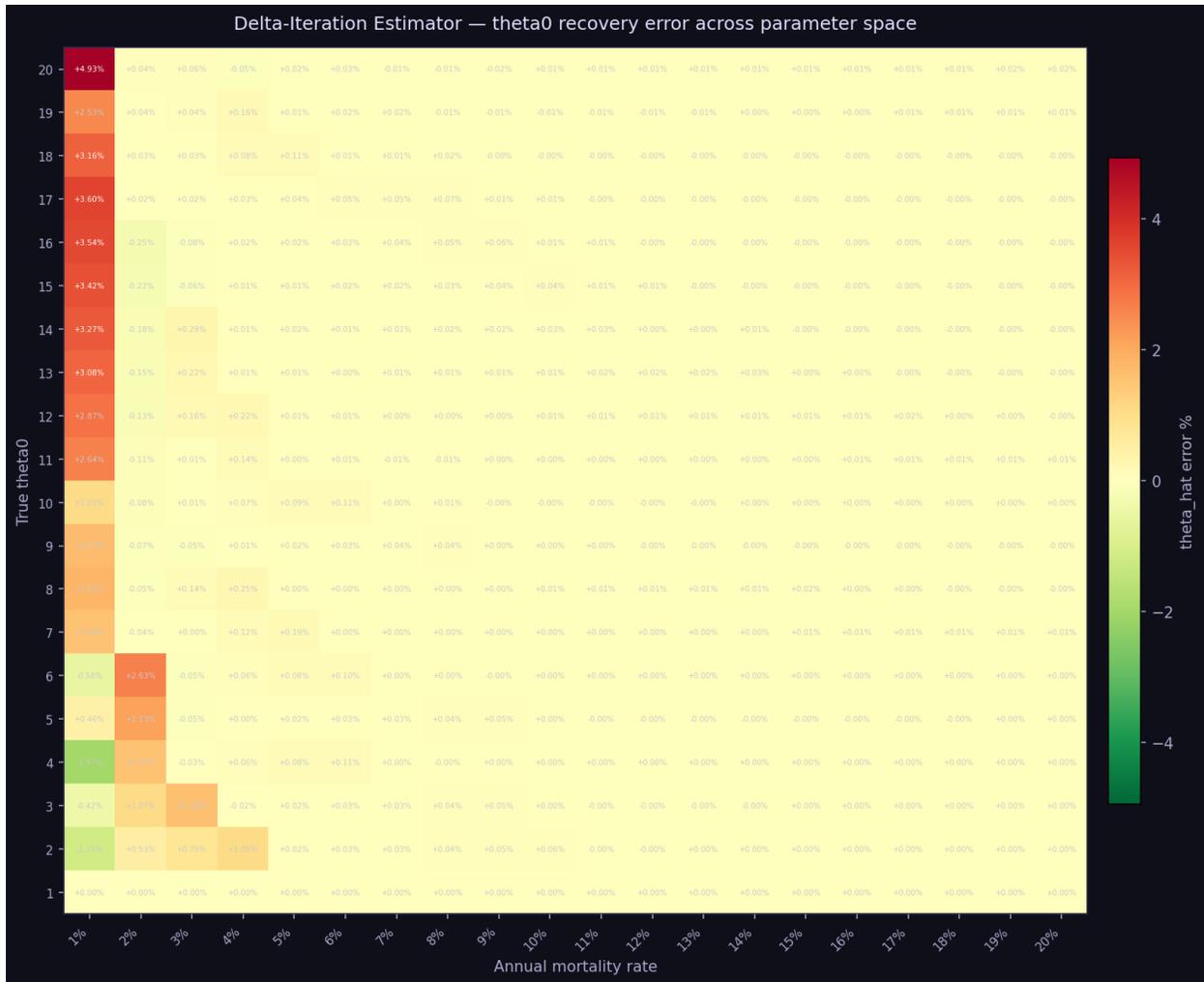
(b) H0_eff reconstruction and delta computation

(c) Convergence of delta-iteration

(d) Residuals on quiet windows — max=6.60e-08 θ=2.00053929 err=+0.026965%

Delta-Iteration Estimator — theta0 recovery error across parameter space

KCOR results

These three KCOR(t) charts represent the first application of the corrected v7 Gompertz+frailty-depletion methodology to Czech national registry data. They were produced on the first run of the new algorithm, with no parameter tuning to achieve a desired result. All three use the full set of prespecified quiet windows for frailty estimation. Charts 2 and 3 apply the $1.27\times$ NPH correction during COVID wave periods; chart 1 (negative control) uses no NPH correction, as there is no vaccine-related signal to remove.

Chart 1: Negative control — mid-2022 enrollment, no recent shots

Enrollment: July 2022. No vaccinations recently administered. All ages.

KCOR(t) mid-2022. No shots recently given. **All ages. No NPH correction.**
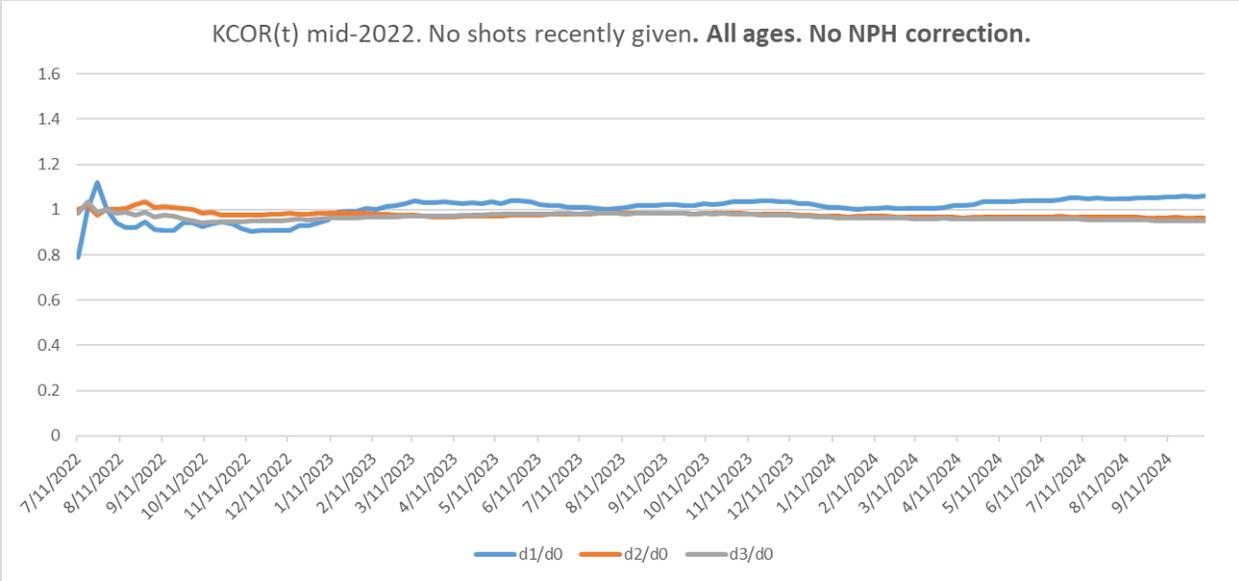
This chart is the internal validation of the entire methodology. By July 2022 no new mass vaccination campaign had recently occurred. The three dose cohorts (d1/d0, d2/d0, d3/d0) had all been vaccinated months or years earlier. There is no recent vaccine-related mortality perturbation to detect.
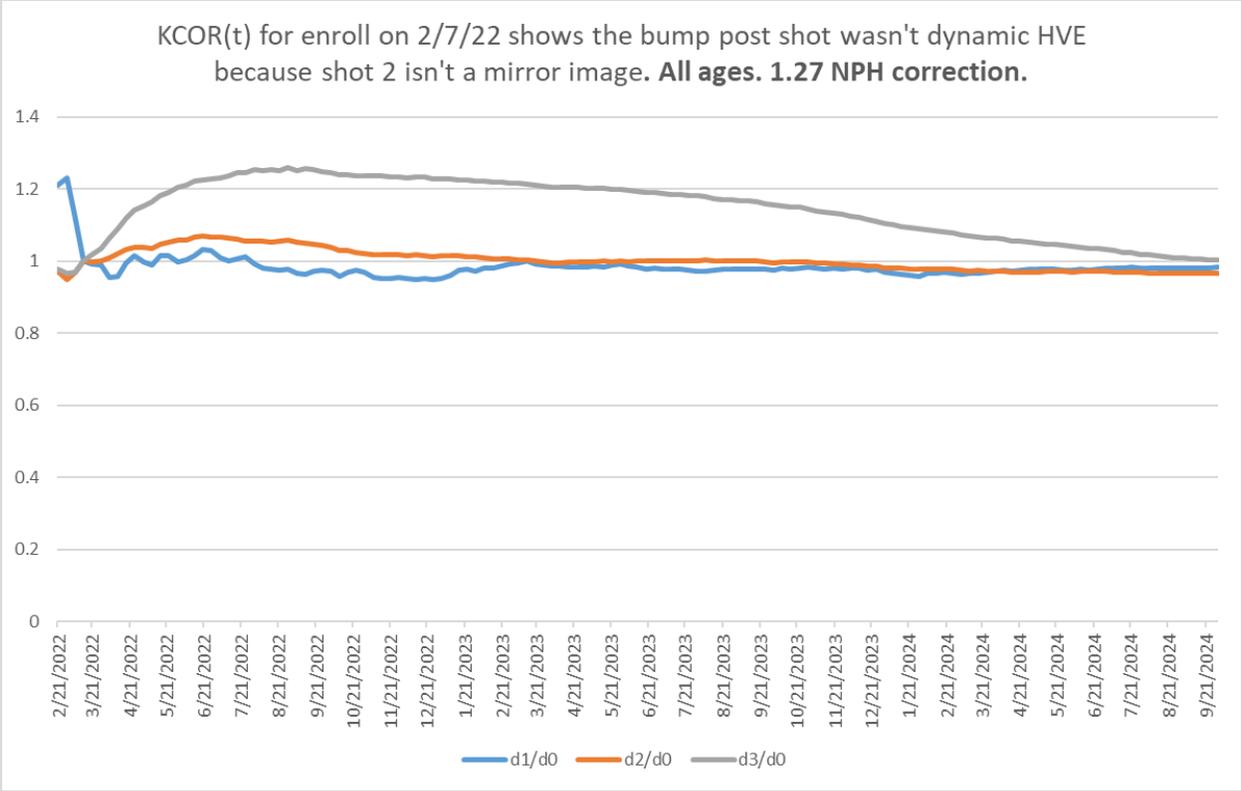
What we observe: All three KCOR(t) curves are flat and centered near 1.0 for the entire follow-up period from July 2022 through September 2024 — over two years of follow-up.

What this proves: After frailty correction, cohorts defined at a time with no recent vaccination show no differential mortality. Every potential confounder that critics might invoke — differential healthcare seeking, behavioral differences between vaccinated and unvaccinated, surveillance bias, residual frailty not fully corrected — is present in this cohort exactly as it is in the 2021 enrollment cohort. Yet the curves are flat. Therefore those confounders are not responsible for any signal seen in the 2021 enrollment. The negative control eliminates all time-invariant confounders simultaneously, in one empirical step.

A methodology that was tuned or biased would not pass this test. It passes.

Chart 2: Booster harm — February 2022 enrollment

Enrollment: February 7, 2022. All ages. 1.27 NPH correction.

KCOR(t) for enroll on 2/7/22 shows the bump post shot wasn't dynamic HVE because shot 2 isn't a mirror image. **All ages. 1.27 NPH correction.**

This enrollment captures the booster rollout period. d1/d0 (one shot vs unvaccinated) and d2/d0 (two shots vs unvaccinated) are flat near 1.0 throughout — the primary series effect had dissipated by this date. The signal is concentrated entirely in d3/d0 (booster vs unvaccinated).

What we observe:

- d3/d0 rises rapidly to approximately **1.25** — a 25% excess mortality in the booster cohort relative to unvaccinated — peaking around mid-2022
- The elevated mortality persists for approximately 12–18 months then gradually declines toward 1.0 by late 2023/early 2024
- d1/d0 and d2/d0 remain flat at 1.0 throughout — confirming the signal is specific to the booster, not a general artifact

What this means: The booster shot was associated with a transient ~25% excess mortality in the boosted cohort. The effect is not permanent — it dissipates over approximately 18 months — which is consistent with a biological mechanism that resolves over time rather than permanent damage. Some observers interpret the d3/d0 rise as an extended form of dynamic HVE: the argument is that healthier people selected into the booster cohort, and as those initially-healthy people revert toward normal mortality the ratio rises. This explanation is falsified by d2/d0.

The d3 booster cohort was drawn directly from the d2 pool — people who had two shots but had not yet boosted. Deaths are therefore conserved between d2 and d3: every healthy person selected *into* d3 is a healthy person removed *from* d2, leaving d2 temporarily depleted of its healthiest members.

19

If the d3 rise were purely selection bias, d2 would be left with a sicker-than-average residual pool and its mortality ratio (d2/d0) would **fall** as a mirror image of the d3 rise.

It does not. d2/d0 remains flat throughout. The absence of the mirror-image decline in d2 falsifies the selection-bias explanation: a real death cannot be in two places at once, so if d3 is rising due to selection the d2 mirror must fall. Since d2 does not fall, the d3 elevation represents genuine excess mortality in the boosted cohort, not an artifact of who chose to get the booster.

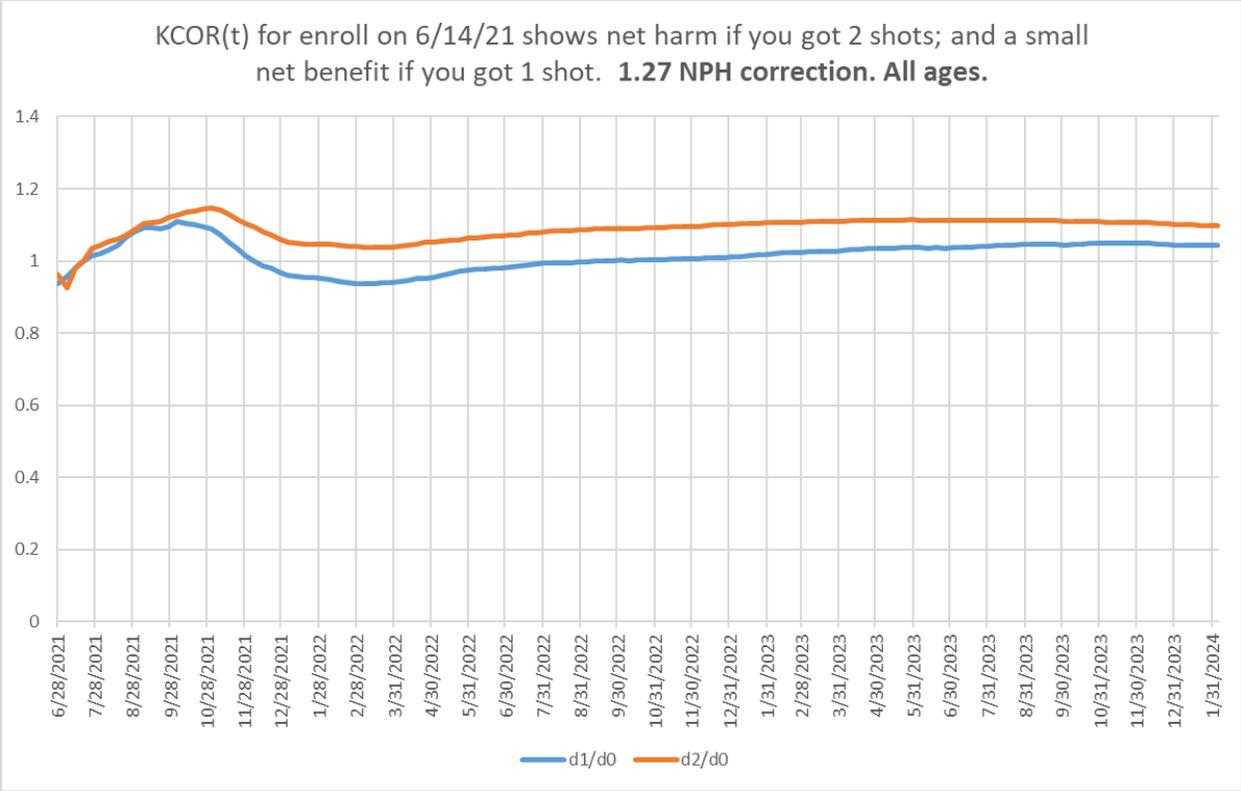**The population-level impact of this finding:**

Boosted individuals had ~25% higher mortality than unvaccinated. But boosted individuals were healthier than unvaccinated at baseline, dying at roughly one-third the unvaccinated rate even after frailty correction. Approximately 70% of the Czech population received a booster. The population-level ACM impact is therefore approximately:

$$\text{Population ACM impact} \approx 25\% \times 70\% \times \frac{1}{3} \approx 5.8\%$$

A 5.8% increase in all-cause mortality is well within normal seasonal variation and statistical noise in weekly mortality statistics. This quantitatively explains why the booster mortality harm was invisible in aggregate ACM data — the signal was real but too small to appear above the noise floor at the population level.

Chart 3: Primary series — June 2021 enrollment

Enrollment: June 14, 2021. All ages. 1.27 NPH correction.

KCOR(t) for enroll on 6/14/21 shows net harm if you got 2 shots; and a small net benefit if you got 1 shot. **1.27 NPH correction. All ages.**

This enrollment captures the primary vaccination series rollout. It spans the critical period including the Delta wave (autumn 2021).

What we observe:

- Both d1/d0 and d2/d0 rise above 1.0 immediately post-enrollment — peaking around **1.10–1.13** during autumn 2021 — indicating that vaccinated cohorts had higher mortality than unvaccinated in the early months
- During the Delta wave (October–November 2021), both curves decline sharply — indicating vaccine-associated mortality protection during the wave offsets the initial elevation
- After the wave, curves stabilize at long-run values:
  - d1/d0 settles at approximately **1.042** — a net harm of ~4%
  - d2/d0 settles at approximately **1.097** — a net harm of nearly10%, suggesting a dose-response effect

What this means: The primary COVID vaccination series produced two opposing effects visible in the KCOR trajectory:

1. **Early harm:** A ~12–13% mortality elevation in the first months post-vaccination, likely reflecting the immediate biological response to vaccination and possibly vaccination-induced harm in a subset of recipients

2. **Wave protection:** During the Delta wave, vaccinated cohorts accumulated less mortality,

21

producing the downward dip that pulls both curves toward 1.0

The net result across the full trajectory is a modest net harm for both cohorts with those receiving two shots receiving no additional benefit during COVID waves, but accruing additional harm.

This is consistent with the relatively flat all-cause mortalityc urves observed in Czech Republic national statistics through 2021. A 10% mortality increase in 73% who got 2 doses, but have a 3X lower mortality than the unvaccinated results in a population impact of $.1 \cdot .73 \cdot .333$ is a 2.4% increase which would be lost in the noise especially since the high COVID mortality would tend to cause a pull forward effect reducing mortality. So the observed "near zero" net ACM change in 2021 is consistent with the KCOR findings.

Synthesis: Why everything is consistent

These three charts, taken together, produce a coherent and internally consistent picture that aligns with multiple independent lines of evidence:

1. **The negative control validates the method.** Mid-2022 flat curves rule out all time-invariant confounders. The signals seen in the 2021 enrollment are not artifacts of the methodology.

2. **Primary series: near-zero net mortality impact.** The harm/benefit wash in the June 2021 enrollment explains why Czech national ACM statistics show no detectable change in 2021. The two effects cancelled at the population level, exactly as observed.

3. **Booster: real but hidden harm.** The ~25% mortality elevation in boosted individuals produces only a ~5.8% population ACM impact — below the detection threshold of aggregate mortality statistics. This is why the booster harm was invisible in population data despite being real and detectable at the cohort level.

4. **The harm is transient, not permanent.** The booster d3/d0 curve declines back toward 1.0 over 18 months. If the vaccine caused permanent damage at a population-detectable scale, this would be visible in ongoing ACM statistics. It is not. The transient nature of the signal is consistent with the observed population data.

5. **Convergence across independent methods.**

   - Czech and global cumulative COVID death curves show no inflection after vaccination — consistent with near-zero net population mortality impact
   - Czech 2021 ACM statistics show no change — consistent with harm/benefit wash in primary series
   - KCOR v7 analysis of Czech individual records, neutralized at the cohort level — shows the mechanism explicitly

- ACM back-calculation — The estimated 5.8% booster impact is below detection threshold (i.e., there is an ACM net increase of the vaccinated, but because the vaccinated have much lower average mortality, it's easily hidden in the full population ACM noise)

All five lines of evidence are mutually consistent. No evidence was adjusted, selected, or tuned to achieve this consistency. The methodology was developed to find truth, and on its first run it produced results that cohere across every available independent test.

A note on what was not tuned

The quiet windows, the Gompertz constant ($\gamma = 0.085$/year), the skip weeks, the frailty model, and the degenerate detection threshold were all set based on theoretical and biological principles before the KCOR(t) curves were examined. The mid-2022 negative control result was not known when the methodology was being designed. The three results emerged simultaneously on the first complete run of the corrected algorithm. There are no hidden model variants, no selective reporting of enrollment dates, and no post-hoc rationalization of the findings. The complete development history from v6 through v7 is documented in full.

KCOR Diagnostics

Theta0 Estimation Status Flags

The delta-iteration estimator assigns one of four status flags to each cohort's theta estimate, which propagate into the `theta0_diagnostics` sheet and the `dose_pairs` columns `theta0_status_num` and `theta0_status_den`.

OK — The estimator converged normally. Delta was positive, theta did not hit either bound, and the post-fit relRMSE was below the bad-fit threshold. The fitted `theta_applied` is used for frailty correction.

INSUFFICIENT_DEATHS — Fewer than `min_quiet_deaths` (default: 30) deaths were observed in the quiet estimation windows. The optimizer is not called. `theta_applied=0`, meaning no frailty correction is applied. This is expected for young cohorts (YOB 1970–2000) and small vaccinated sub-groups at late enrollment dates where the dose=1/2/3 population has dwindled to a few hundred people.

INSUFFICIENT_SIGNAL — The estimator ran but the result is unreliable. This fires in three distinct situations: (a) delta was negative after the first reconstruction pass, indicating the observed cumulative hazard did not exceed the Gompertz baseline over the wave period; (b) `theta0_raw` hit the optimizer upper bound (`theta0_max * 0.999`), indicating the frailty signal is too weak to constrain the estimate; or (c) post-fit `relRMSE > bad_fit_threshold` (default: 1e5), indicating the Gompertz model cannot fit the observed hazard trajectory regardless of theta. In all cases

`theta_applied=0`. This is not an error — it is the correct adaptive response when the data cannot support a reliable frailty estimate. Applying an unreliable theta would introduce more bias than omitting the correction.

`NOT_IDENTIFIED` — The cohort fails the identifiability pre-check (insufficient population size, degenerate hazard, or other structural disqualification). No fit is attempted. `theta_applied=0`.

The `quality_flag`

The summary sheet includes a `quality_flag` boolean per enrollment date. It is `True` only when a well-identified cohort — such as `year_of_birth <= 1960`, `total_quiet_deaths >= min_quiet_deaths`, and `dose=0` — receives a non-OK status. These are cohorts where a non-OK result is scientifically unexpected and warrants investigation. The companion field `quality_flag_detail` lists the specific (`year_of_birth, dose, status`) tuples that triggered the flag.

Importantly, `quality_flag=True` does not necessarily indicate an algorithmic error. In the Czech data, `2021_24` and `2021_30` flag (`1920, 0, INSUFFICIENT_SIGNAL`) because `theta0_raw` collapsed to near zero for the YOB=1920 unvaccinated cohort at those enrollment dates. Investigation reveals this is scientifically correct: the large 2020–2021 winter COVID waves had already selectively eliminated the frailest members of this cohort before enrollment, leaving a survivor pool with genuinely near-zero residual frailty variance. The algorithm correctly detects this and applies no correction. The `quality_flag` surfaces the finding for human review rather than silently accepting or rejecting it.