KCOR: Depletion-Neutralized Cohort Comparison via Gamma-Frailty Normalization Under Selection-Induced Hazard Curvature

Manuscript metadata (for journal submission)

- **Article type**: Methods / Statistical method
- **Running title**: KCOR via gamma-frailty normalization
- **Authors**: Steven T. Kirsch
- **Affiliations**: Independent Researcher, United States
- **Corresponding author**: stk@alum.mit.edu
- **Word count**: 5,770
- **Keywords**: selection bias; healthy vaccinee effect; non-proportional hazards; frailty; gamma frailty; negative controls; causal inference; observational studies; mortality curvature

---

Abstract

Retrospective cohort studies often estimate the mortality impact of medical interventions, but selection can create non-exchangeable cohorts that defeat standard comparisons. Selective uptake can induce systematic differences in mortality curvature—differences in the time-evolution of cohort hazards driven by frailty heterogeneity and depletion of susceptibles—violating the assumptions of Cox proportional hazards models, age-standardized mortality rates, and inverse-probability weighting. We introduce **KCOR (Kirsch Cumulative Outcomes Ratio)**, a method that normalizes selection-induced curvature before cohort comparison by estimating and inverting a gamma-frailty mixture model in **cumulative-hazard space**. KCOR fits cohort-specific selection parameters during epidemiologically quiet periods. Observed cumulative hazards are then transformed to depletion-neutralized baseline cumulative hazards and compared cumulatively via ratios. Because selection and treatment can be confounded in observational data, KCOR is presented as an integrated normalization-and-comparison framework; causal interpretation requires additional assumptions and is evaluated via prespecified control tests and simulations. We describe the KCOR framework, its mathematical foundation, and its validation using prespecified negative and positive control tests designed to stress curvature. KCOR requires only event timing—dates of intervention (enrollment) and dates of death, with birth date or year for age stratification. This makes it applicable to minimal record-level mortality datasets. Empirically, KCOR frequently estimates negligible frailty variance for some cohorts while estimating substantial frailty-driven depletion for others, reflecting asymmetric selection at cohort entry. After frailty neutralization, the depletion-neutralized cumulative hazards are expected to be approximately linear during epidemiologically quiet periods; departures from linearity serve as a built-in diagnostic of assumption violation or quiet-window contamination. KCOR defines a unified cumulative comparison whose output is the cumulative hazard ratio. KCOR therefore enables interpretable cumulative cohort comparisons in settings where treated and untreated hazards are non-proportional because selection induces different depletion dynamics, while also providing explicit failure-mode signals when curvature cannot be explained by the depletion model.

---

1. Introduction

1.1 Retrospective cohort comparisons under selection

Randomized controlled trials (RCTs) are the gold standard for causal inference, but are often infeasible, underpowered for rare outcomes, or unavailable for questions that arise after rollout. As a result, observational cohort comparisons are widely used to estimate intervention effects on outcomes such as all-cause mortality.

However, when intervention uptake is voluntary, prioritized, or otherwise selective, treated and untreated cohorts are frequently **non-exchangeable** at baseline and evolve differently over follow-up. This problem is not limited to any single intervention class; it arises whenever the same factors that influence treatment uptake also influence outcome risk.

1.2 Curvature (shape) is the hard part: non-proportional hazards from frailty depletion

Selection does not merely shift mortality **levels**; it can alter mortality **curvature**—the time-evolution of cohort hazards. Frailty heterogeneity and depletion of susceptibles naturally induce curvature even when individual-level hazards are simple functions of time. When selection concentrates high-frailty individuals into one cohort (or preferentially removes them from another), the resulting cohort-level hazard trajectories can be strongly non-proportional.

One convenient way to formalize "curvature" is in cumulative-hazard space: if the cumulative hazard $H(t)$ were perfectly linear in time, then its second derivative would be zero, whereas selection-induced depletion generally produces negative concavity (downward curvature) in observed cumulative hazards during otherwise stable periods.

This violates core assumptions of many standard tools:

- **Cox PH**: assumes hazards differ by a time-invariant multiplicative factor (proportional hazards).
- **IPTW / matching**: can balance measured covariates yet fail to balance unmeasured frailty and the resulting depletion dynamics.
- **Age-standardization**: adjusts levels across age strata but does not remove cohort-specific time-evolving hazard shape.

KCOR is designed for this failure mode: **cohorts whose hazards are not proportional because selection induces different depletion dynamics (curvature).** Approximate linearity of cumulative hazard after adjustment is therefore not assumed, but serves as an internal diagnostic indicating that selection-induced depletion has been successfully removed.

Although this manuscript is motivated in part by mortality analyses conducted during the COVID-19 vaccination period, the methodological problem addressed here is general. The COVID setting provides unusually clear examples of selection-induced non-proportional hazards—because uptake was voluntary, rapidly time-varying, and correlated with baseline health—making residual confounding easy to diagnose using control outcomes such as non-COVID mortality. However, KCOR is not specific to COVID, vaccination, or infectious disease. The estimator applies to any retrospective cohort comparison in which selection induces differential depletion dynamics that violate proportional hazards assumptions.

In this paper we distinguish two mechanisms often lumped as the 'healthy vaccinee effect' (HVE):

- **Static HVE:** baseline differences in latent frailty distributions at cohort entry (e.g., vaccinated cohorts are healthier on average). In the KCOR framework, this manifests as differing depletion curvature (different θ_d) and is the primary target of frailty normalization.

- **Dynamic HVE:** short-horizon, time-local selection processes around enrollment that create transient hazard suppression immediately after enrollment (e.g., deferral of vaccination during acute illness, administrative timing, or short-term behavioral/health-seeking changes). Dynamic HVE is operationally addressed by prespecifying a skip/stabilization window (§2.7) and can be evaluated empirically by comparing early-period signatures across related cohorts in multi-dose settings.

1.3 Related work: frailty, depletion of susceptibles, and selection-induced non-proportional hazards

KCOR builds on a long literature on unobserved heterogeneity ('frailty') and depletion of susceptibles, in which population-level hazards can decelerate over time even when individual hazards are simple. The gamma frailty model is widely used because its Laplace transform yields a closed-form relationship between baseline and observed survival/cumulative hazard, enabling tractable inference and interpretation[1].

A separate literature emphasizes that observational estimates of vaccine effectiveness can remain confounded despite extensive matching and adjustment, often revealed by negative control outcomes and time-varying non-COVID mortality differences[2,3]. KCOR is complementary: rather than using negative controls only to detect confounding, it targets a specific confounding geometry—selection-induced depletion curvature—and then requires controls and simulations to validate that the intended curvature component has been removed.

We do not claim that KCOR subsumes all approaches to confounding adjustment; rather, it provides a dedicated normalization and diagnostic toolkit for settings where non-proportional hazards arise primarily from selection-induced depletion dynamics.

1.4 Evidence from the literature: residual confounding despite meticulous matching

Two large, rigorously designed observational analyses illustrate the core empirical motivation: even extremely careful matching and adjustment can leave large residual differences in non-COVID mortality, indicating confounding and selection that standard pipelines do not eliminate.

1.4.1 Denmark (negative controls highlight confounding)

Obel et al. used Danish registry data to build 1:1 matched cohorts and applied negative control outcomes to assess confounding. Their plain-language summary includes the following:

> Meaning: The negative control methods indicate that observational studies of SARS-CoV-2 vaccine effectiveness may be prone to substantial confounding which may impact the observed associations. This bias may both lead to underestimation of vaccine effectiveness (increased risk of SARS-CoV2 infection among vaccinated individuals) and overestimation of the vaccine effectiveness (decreased risk of death after of SARS-CoV2 infection among vaccinated individuals). Our results highlight the need for randomized vaccine efficacy studies after the emergence of new SARS-CoV-2 variants and the rollout of multiple booster vaccines.[2]

This is a direct statement that observational designs—even with careful matching and covariate adjustment—can remain substantially confounded when selection and health-seeking behavior differ between cohorts.

1.4.2 Qatar (time-varying HVE despite meticulous matching)

Chemaitelly et al. analyzed matched national cohorts and explicitly measured the **time-varying healthy vaccinee effect (HVE)** using non-COVID mortality as a control outcome. They report a pronounced early-period reduction in non-COVID mortality among vaccinated individuals despite meticulous matching, followed by reversal later in follow-up, consistent with dynamic selection and depletion processes.[3]

Together, these studies motivate a methods gap: we need estimators that explicitly address **time-evolving selection-induced curvature**, not only baseline covariate imbalance. Table tbl. 1 summarizes these findings.

Table 1: Summary of two large matched observational studies showing residual confounding / HVE despite meticulous matching.

Study	Design	Matching/adjustment	Key control finding	Implication for methods
Obel et al. (Denmark)[2]	Nationwide registry cohorts (60–90y)	1:1 match on age/sex + covariate adjustment; negative control outcomes	Vaccinated had higher rates of multiple negative control outcomes, but substantially lower mortality after unrelated diagnoses	Strong evidence of confounding in observational VE estimates; "negative control methods indicate… substantial confounding"
Chemaitelly et al. (Qatar)[3]	Matched national cohorts (primary series and booster)	Exact 1:1 matching on demographics + coexisting conditions + prior infection; Cox models	Strong early reduction in non-COVID mortality (HVE), with time-varying reversal later	Even meticulous matching leaves time-varying residual differences consistent with selection/frailty depletion

1.5 Contribution of this work

This paper introduces **KCOR**, a method that transforms observed cohort hazards to remove selection-induced depletion dynamics prior to comparison, enabling interpretable cumulative cohort comparisons under selection-induced non-proportional hazards.

Table 2: Comparison of Cox proportional hazards, Cox with frailty, and KCOR across key methodological dimensions.

Feature	Cox PH	Cox + frailty	KCOR
Primary estimand	Hazard ratio	Hazard ratio	Cumulative hazard ratio
Conditions on survival	Yes	Yes	No
Assumes PH	Yes	Yes (conditional)	No
Frailty role	None	Nuisance	Object of inference
Uses partial likelihood	Yes	Yes	No
Handles selection-induced curvature	No	Partial	Yes (targeted)
Output interpretable under non-PH	No	No	Yes (cumulative)

Method family	Primary estimand (typical output)	Handles selection-induced depletion curvature?	What it requires	Primary failure mode
Kaplan–Meier / Cox PH	Instantaneous HR; survival differences under PH	No	Exchangeability; proportional hazards; adequate covariates	Non-PH from latent selection yields misleading HRs
Cox with frailty term	HR with random-effect heterogeneity	Partial	Correct frailty form; PH-centric interpretation	Depletion geometry can remain; HR interpretation unstable
Matching / IPTW / MSM	Model-based contrasts (ATE/ATT)	Indirect (via measured proxies only)	Correct models; rich covariates; positivity	Latent frailty and depletion persist unaddressed
Negative control methods	Bias detection (diagnostic)	No	Valid negative controls	Detects bias but does not remove it
KCOR (this work)	**KCOR(t): ratio of depletion-neutralized cumulative hazards**	**Yes (targeted)**	**DOB/DOD/DOI; valid quiet window; identifiable curvature; prespecified diagnostics and controls**	**If depletion model or quiet window fails, diagnostics flag nonlinearity or instability and comparison is not interpretable**

Table 1. Positioning KCOR among retrospective methods (corrected).
Most retrospective approaches either compare cohorts under proportional-hazards assumptions, balance measured confounding, or diagnose bias without removing it. KCOR occupies a distinct role: it **neutralizes selection-induced depletion dynamics** via gamma-frailty inversion and then **extracts the cohort contrast using a cumulative hazard ratio (KCOR), which is the estimand reported**. Normalization alone does not yield an interpretable signal; the KCOR ratio is the estimand that answers whether one cohort experienced higher or lower cumulative event risk than another under the stated assumptions. KCOR's estimand is cumulative by construction; instantaneous hazard ratios are not recovered even after normalization. {#tbl:positioning}

KCOR is not merely a frailty-normalization method. While gamma-frailty inversion is a necessary step, the central contribution of KCOR is the end-to-end comparison system that follows normalization. KCOR transforms observed cumulative hazards into a depletion-neutralized space and then defines the correct comparison operator in that space—a cumulative hazard ratio—together with diagnostics that determine when such comparisons are interpretable. Normalization alone does not yield a signal; the signal emerges only through the KCOR comparison itself. In this sense, KCOR should be understood as a complete retrospective comparison framework rather than a preprocessing adjustment that can be substituted into standard estimators. The integrated nature of KCOR—normalization, comparison, and diagnostics as a single system—is illustrated schematically in Figure fig. 2.

This manuscript is **methods-only**:

- We present the estimator, model assumptions, and uncertainty quantification.
- We validate the method using prespecified negative and positive controls designed to stress selection-induced curvature.
- We defer any applied real-world intervention conclusions to a separate, dedicated applied paper.

KCOR is proposed as a diagnostic and normalization estimator for selection-induced hazard curvature; causal interpretation requires additional assumptions beyond the scope of this methods paper.

1.6 Relation to causal inference frameworks

KCOR is not intended to replace established causal inference designs such as instrumental variables, regression discontinuity, difference-in-differences, or target trial emulation. Those frameworks address distinct identification problems and typically require either exogenous instruments, sharp intervention thresholds, rich covariate histories, or well-defined intervention regimes.

KCOR is designed for a complementary setting in which such requirements are not met—specifically, retrospective cohort data where only dates of birth, death, and intervention are available, and where selection-induced depletion produces strong non-proportional hazards that invalidate hazard-ratio-based estimators. In this setting, KCOR targets a different failure mode: curvature in cumulative hazards arising from latent heterogeneity and selection rather than from time-varying treatment effects.

By neutralizing depletion geometry and defining a cumulative comparison operator in the resulting space, KCOR enables interpretable cohort contrasts under minimal data constraints. When stronger causal designs are feasible, they should be preferred; when they are not, KCOR provides a principled way to assess whether observed cohort differences persist once selection-induced depletion is removed.

Box 1: KCOR in one page

KCOR (Kirsch Cumulative Outcomes Ratio) is a cumulative-hazard normalization and comparison framework for retrospective cohort studies affected by selection-induced non-proportional hazards. The method proceeds in six steps:

1. **Fixed cohorts**: Individuals are assigned to cohorts at enrollment based on intervention status; no switching or censoring is permitted in the primary estimand.

2. **Cumulative hazard estimation**: Discrete-time hazards are computed from event counts and risk sets, then accumulated into observed cumulative hazards $H_d^{\text{obs}}(t)$ after an optional stabilization skip period.

3. **Quiet-window frailty fit**: During epidemiologically quiet periods (free of external shocks), cohort-specific frailty variance parameters θ_d and baseline hazard levels k_d are estimated via nonlinear least squares by fitting the gamma-frailty identity $H_d^{\text{obs}}(t) = \frac{1}{\theta_d} \log(1 + \theta_d k_d t)$ to observed cumulative hazards.

4. **Gamma inversion**: The fitted frailty parameters are used to invert the gamma-frailty identity, transforming observed cumulative hazards into depletion-neutralized baseline cumulative hazards $\tilde{H}_{0,d}(t) = \frac{e^{\hat{\theta}_d H_d^{\text{obs}}(t)} - 1}{\hat{\theta}_d}$.

5. **KCOR ratio**: Cohorts are compared via the ratio of depletion-neutralized cumulative hazards: $\text{KCOR}(t) = \tilde{H}_{0,A}(t)/\tilde{H}_{0,B}(t)$.

6. **Diagnostics**: Post-normalization linearity in quiet periods, fit residuals, and parameter stability under window perturbations serve as internal checks that depletion normalization is valid and assumptions are met.

---

2. Methods

Table tbl. 4 defines the notation used throughout this section.

Table 4: Notation used throughout the Methods section. θ_d denotes the cohort-specific depletion (frailty variance) parameter governing curvature in the observed cumulative hazard.

Symbol	Meaning
d	cohort index (enrollment definition × age group × intervention count (discrete exposure index))
t	event time since enrollment (discrete bins, e.g., weeks)
$h_d^{\mathrm{obs}}(t)$	observed cohort hazard at time t
$H_d^{\mathrm{obs}}(t)$	observed cumulative hazard (after skip/stabilization)
$h_{0,d}(t)$	depletion-neutralized baseline hazard for cohort d
$H_{0,d}(t)$	depletion-neutralized baseline cumulative hazard
θ_d	frailty variance (selection strength) for cohort d
k_d	baseline hazard level for cohort d under default baseline shape

For COVID-19 vaccination analyses, intervention count corresponds to the number of vaccine doses received; more generally, this can index any discrete exposure level.

2.1 Conceptual framework: level vs curvature under selection

Differences in mortality between cohorts can arise from:

- **Level effects**: multiplicative shifts in hazard that are constant over time.
- **Curvature effects**: differences in the time-evolution of cohort hazards induced by heterogeneity and selective depletion.

Selection bias commonly produces curvature differences through frailty mixing and depletion. KCOR's strategy is to **estimate and remove the selection-induced depletion component**, then compare cohorts on a cumulative scale.

Figure fig. 2 provides a schematic overview of the KCOR workflow.

2.1.1 Assumptions and identifiability conditions

KCOR relies on the following assumptions for interpretable normalization and comparison. These assumptions are explicit, empirically testable, and accompanied by diagnostics that signal when they are violated.

- **Fixed cohorts at enrollment**: Individuals are assigned to cohorts at enrollment and do not switch; no censoring or as-treated switching is permitted in the primary estimand.

- **Quiet-window validity**: The fitting window must be free of major external shocks (e.g., epidemic waves, policy changes, reporting artifacts) that would induce curvature unrelated to selection-induced depletion.

- **Temporal separability (quiet-window identification)**: Frailty parameters are identified during a prespecified quiet window in which treatment effects are small relative to selection-induced depletion. This window is used solely for estimation of frailty parameters and does not constrain treatment effects at other times.
  Non-overlap refers to dominance of selection-induced depletion over treatment effects within the quiet window, not to the absence of treatment effects at other times; this condition is assessed empirically via

6

normalization diagnostics rather than imposed a priori.
When this condition is violated, KCOR trajectories fail to stabilize and diagnostic criteria indicate reduced interpretability.

- **Gamma-frailty as geometric approximation**: Gamma frailty is used not as a claim of biological truth, but as a mathematically tractable approximation whose adequacy is empirically testable.

- **Identifiability via curvature**: Frailty parameters are identified from observable curvature in cumulative-hazard space; near-linear cumulative hazards yield weak identification and naturally drive $\hat{\theta} \to 0$.

- **Baseline hazard regularity**: The underlying baseline hazard is assumed slowly varying or approximately linear over the quiet window.

- **Diagnostics required**: Poor fit, residual curvature after normalization, or parameter instability under small window perturbations signal assumption violation and limit interpretability.

Selection and treatment are not generically identifiable from two cohort hazard curves without additional structure. KCOR therefore does not claim universal identifiability. Instead, it targets a specific, testable confounding geometry—selection-induced depletion under time-invariant multiplicative heterogeneity—and relies on diagnostics, negative controls, and simulations to validate when that structure is adequate. The operating characteristics of KCOR when this temporal-separability condition holds—and when it is intentionally violated via overlap between the quiet window and treatment effects—are evaluated explicitly in Simulation S7 (Appendix B.6) and summarized in the Results (§3.4). In settings without a clear post-enrollment quiet period prior to major epidemic pressure, KCOR estimates should be interpreted as descriptive of cumulative differences rather than as isolated treatment effects, and diagnostics should be examined for evidence of assumption violation.

2.1.2 What KCOR is not: distinction from Cox and frailty regression

KCOR is **not** a Cox proportional hazards model, with or without frailty. Cox models—whether standard or augmented with gamma frailty—are regression models whose primary estimand is a coefficient vector β, typically interpreted through hazard ratios. Estimation proceeds by (penalized) partial likelihood based on risk sets, and interpretation relies on proportional hazards (at least conditional on frailty in frailty-augmented Cox models).

In contrast, KCOR does not estimate regression coefficients, does not condition on risk sets, and does not assume proportional hazards. KCOR treats cohort heterogeneity (frailty-driven selection and depletion) as the primary object of inference rather than as a nuisance random effect. Specifically, KCOR estimates cohort-specific frailty parameters from curvature in observed cumulative hazards during epidemiologically quiet periods and uses these parameters to compute **depletion-neutralized baseline cumulative hazards**. KCOR then compares cohorts cumulatively using ratios of these depletion-neutralized quantities.

Because KCOR neither targets β nor uses partial likelihood, it is not a special case of Cox frailty models and does not generalize them. KCOR addresses a different inferential problem: normalization of selection-induced hazard curvature and cumulative comparison after that normalization. Cox-type models remain appropriate when the scientific target is an instantaneous hazard ratio and proportional hazards is defensible; KCOR is designed for settings where non-proportional hazards from selection and depletion dominate and cumulative comparisons are the target.

2.1.3 Model diagnostics and falsification criteria

KCOR is not validated by goodness-of-fit alone. Instead, it implies a set of independent structural diagnostics that must be satisfied simultaneously. These diagnostics are defined ex ante and provide explicit failure modes: violation of any one constitutes evidence against model adequacy rather than a need for reinterpretation.

Specifically, KCOR implies that (i) unvaccinated cohorts drawn from heterogeneous populations exhibit non-zero gamma frailty with visible curvature in cumulative hazard; (ii) cohorts subject to strong selective uptake exhibit near-zero estimated frailty ($\hat{\theta} \approx 0$) and approximately linear cumulative hazards during epidemiologically quiet windows; (iii) fitted frailty parameters vary coherently with age, reflecting the interaction between baseline hazard and heterogeneity rather than enforcing monotonicity; and (iv) cumulative hazard ratios converge to stable asymptotes following KCOR normalization.

Importantly, none of these properties are imposed as constraints or priors. All emerge from the data. In low-information regimes (e.g., sparse events or highly selected cohorts), KCOR is expected to degrade toward instability or attenuation rather than producing spurious effects. These behaviors provide a falsifiable diagnostic framework rather than a model-tuning mechanism.

2.2 Cohort construction and estimand

KCOR is defined for **fixed cohorts** at enrollment. Required inputs are minimal: for each individual, the intervention or enrollment date(s) and the **event date** (e.g., death for mortality analyses), with birth date or year included only if age stratification is performed. Throughout, we use "event" to denote the outcome of interest, with event timing recorded relative to cohort enrollment.

- Cohorts are fixed at enrollment and defined by intervention status at the start of the enrollment week; doses administered during the enrollment week do not affect cohort assignment (i.e., dose status is determined by doses received **strictly before** the enrollment week start).
- No censoring or cohort switching is permitted in the primary estimand.
- Analysis proceeds in **event time** t (time since enrollment).

This fixed-cohort design corresponds to an intent-to-treat–like estimand under selection. It is chosen deliberately to avoid time-varying deferral bias, immortal time bias, and dynamic health-based sorting that arise when individuals change exposure status during follow-up. Dynamic "as-treated" formulations are treated as sensitivity analyses rather than primary estimands.

Failure event. The failure event analyzed in this manuscript is **all-cause mortality**, defined as death from any cause occurring after cohort enrollment. KCOR therefore does not target a cause-specific hazard and is not framed as a competing risks analysis. This choice is deliberate: selection-induced depletion operates on overall mortality risk regardless of cause, and restricting to cause-specific outcomes requires additional assumptions and introduces sensitivity to misclassification and post-treatment information. Extensions of KCOR to cause-specific outcomes are possible but are outside the scope of this methods paper.

2.2.1 KCOR data representation and reproducibility

All analyses were performed using the KCOR (Kirsch Cumulative Outcomes Ratio) framework, which operates on fixed-cohort, time-indexed observational data under explicitly defined enrollment and outcome semantics.

To ensure reproducibility across jurisdictions with differing privacy and disclosure-control regimes, KCOR supports three formally defined data representations:

1. Record-level data

2. Aggregated outcome summary data

3. Hazard summary data

These formats differ only in disclosure surface; they are analytically equivalent under the KCOR methodology.

The precise definitions of enrollment, cohort freezing, time indexing, risk sets, outcomes, stratifiers, and disclosure-control constraints are specified in the **KCOR data format specification**, which is versioned and publicly available:

> *KCOR File Format Specification*
> `documentation/specs/KCOR_file_format.md`

All KCOR results presented in this paper can be reproduced from any data representation that conforms to this specification.

2.3 Hazard estimation and cumulative hazards (discrete time)

Let t denote event time since enrollment (e.g., weeks), $D_d(t)$ deaths during interval t in cohort d, and $N_d(t)$ the number at risk at the start of interval t. In discrete time, hazards are treated as piecewise-constant and can be computed from

interval risk as

$$h_d^{\text{obs}}(t) = -\ln\left(1 - \frac{D_d(t)}{N_d(t)}\right). \tag{1}$$

We work primarily in **cumulative-hazard space**, accumulating observed hazards after an optional stabilization skip (see §2.7):

$$H_d^{\text{obs}}(t) = \sum_{s \leq t} h_d^{\text{eff}}(s)\, \Delta t, \qquad \Delta t = 1 \text{ (one time bin)}. \tag{2}$$

2.4 Selection model: gamma frailty and the cumulative-hazard identity

2.4.1 Individual hazards with multiplicative frailty

Model individual hazards in cohort d as multiplicative frailty on top of a baseline hazard:

$$h_{i,d}(t) = z_{i,d}\, h_{0,d}(t), \qquad z_{i,d} \sim \text{Gamma(mean} = 1, \text{ var} = \theta_d). \tag{3}$$

Gamma frailty is used not as a claim of biological truth, but as a mathematically minimal and widely used model for unobserved heterogeneity whose Laplace transform yields a closed-form relationship between observed and baseline cumulative hazards[1]. In KCOR, gamma frailty therefore serves as a **geometric approximation for depletion normalization**: its adequacy is treated as empirically testable and is assessed using prespecified negative controls and sensitivity analyses rather than assumed dogmatically.

Frailty $z_{i,d}$ captures latent heterogeneity in baseline risk and drives selective depletion: higher-frailty individuals die earlier, changing the cohort composition over time and inducing curvature in $h_d^{\text{obs}}(t)$ even when $h_{0,d}(t)$ is simple.

Interpretation of frailty in this work. Here "frailty" denotes unobserved, time-invariant multiplicative heterogeneity in baseline mortality risk at cohort entry. It is not interpreted as a specific biological attribute and is not treated as a causal mediator of vaccination. Rather, it is a statistical construct capturing latent heterogeneity that produces selective depletion over time and induces curvature in cohort-level hazards and cumulative hazards. KCOR uses frailty as a geometric device to model and remove selection-induced curvature prior to cohort comparison. Frailty in KCOR is not interpreted as a causal variable or mediator of the intervention; it is a statistical construct used to model and remove selection-induced depletion geometry prior to cohort comparison.

2.4.2 Gamma-frailty identity (core mathematics)

Let the baseline cumulative hazard be

$$H_{0,d}(t) = \int_0^t h_{0,d}(s)\, ds. \tag{4}$$

Integrating over gamma frailty yields a closed-form relationship between the observed cohort cumulative hazard and the baseline cumulative hazard:

$$H_d^{\text{obs}}(t) = \frac{1}{\theta_d}\, \log\left(1 + \theta_d H_{0,d}(t)\right). \tag{5}$$

This identity can be inverted exactly:

$$H_{0,d}(t) = \frac{e^{\theta_d H_d^{\text{obs}}(t)} - 1}{\theta_d}. \tag{6}$$

This inversion is the **KCOR v6 normalization step**: it transforms the observed cumulative hazard into a depletion-neutralized baseline cumulative hazard for each cohort. Figure fig. 1 illustrates this logic schematically.
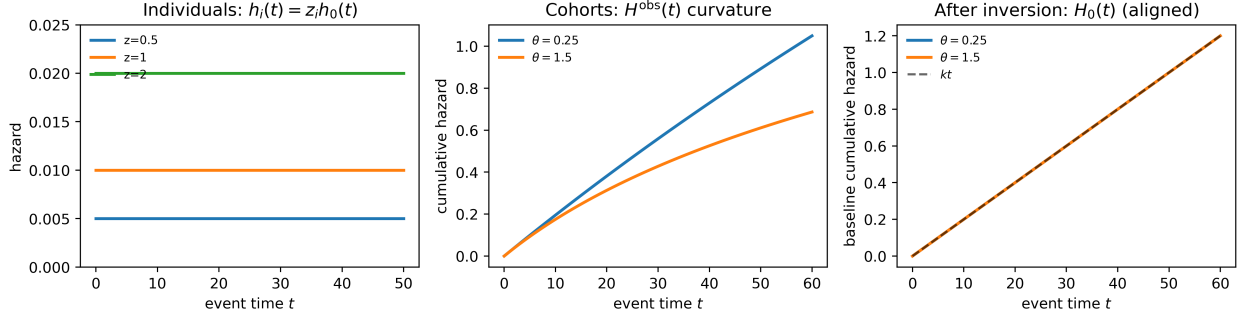
Figure 1: Three-panel schematic illustrating the KCOR v6 normalization logic. Left: individual hazards differ only by multiplicative frailty z, with no treatment effect. Middle: aggregation over heterogeneous frailty induces cohort-level curvature in observed cumulative hazards $H^{\mathrm{obs}}(t)$ despite identical baseline hazards. Right: inversion of the gamma-frailty identity recovers aligned baseline cumulative hazards $H_0(t)$, demonstrating depletion-neutralization. This figure is schematic and intended for conceptual illustration; it does not represent empirical data.

2.4.3 Baseline shape for fitting (default)

To identify θ_d from data, KCOR fits the gamma-frailty model during epidemiologically quiet periods. In the reference specification, the baseline time-shape is taken to be constant over the fit window:

$$h_{0,d}(t) = k_d\, g(t), \qquad g(t) = 1, \qquad H_{0,d}(t) = k_d t. \tag{7}$$

This choice minimizes degrees of freedom and forces curvature during quiet periods to be explained by selection (frailty) rather than by an explicit time-varying baseline.

2.5 Estimation during quiet periods (cumulative-hazard least squares)

KCOR estimates (k_d, θ_d) independently for each cohort d using only time bins that fall inside a prespecified **quiet window** in calendar time (ISO week space). The quiet window is prespecified and applied consistently across cohorts within an analysis; robustness to alternate quiet-window bounds is assessed in sensitivity analyses. Quiet periods are identified diagnostically via stability of observed cumulative hazards and absence of external shocks, rather than by a fixed universal numeric threshold. Let $\mathcal{T}_d$ denote the set of event-time bins t whose corresponding calendar week lies in the quiet window, with t also satisfying $t \geq \mathrm{SKIP\_WEEKS}$.

Under the default baseline shape, the model-implied observed cumulative hazard is

$$H_d^{\mathrm{model}}(t; k_d, \theta_d) = \frac{1}{\theta_d}\, \log\!\left(1 + \theta_d k_d t\right). \tag{8}$$

Identifiability of (k_d, θ_d) comes from curvature in cumulative-hazard space: the mapping $t \mapsto H_d^{\mathrm{obs}}(t)$ is nonlinear when $\theta_d > 0$. When depletion is weak (or the quiet window is too short to show curvature), the model smoothly collapses to a linear cumulative hazard, since $H_d^{\mathrm{model}}(t; k_d, \theta_d) \to k_d t$ as $\theta_d \to 0$. Operationally, near-linear $H_d^{\mathrm{obs}}(t)$ naturally drives $\hat{\theta}_d \approx 0$; fit diagnostics such as n_{obs} and RMSE in H-space provide a practical check on whether the selection parameters are being identified from the quiet-window data. In practice, lack of identifiable curvature naturally manifests as $\hat{\theta} \to 0$, providing an internal diagnostic for non-identifiability over short or sparse follow-up.

In applied analyses, this behavior is most commonly observed in vaccinated cohorts, whose cumulative hazards during quiet periods are often close to linear. In such cases, the gamma-frailty fit collapses naturally to $\hat{\theta}_d \approx 0$, indicating minimal detectable depletion. This outcome is data-driven and reflects the absence of observable selection-induced curvature rather than a modeling assumption. When residual time-varying risk contaminates a nominally quiet window, $\hat{\theta}$ estimates naturally shrink toward zero, signaling limited identifiability rather than inducing spurious correction.

Parameters are estimated by constrained nonlinear least squares:

$$(\hat{k}_d, \hat{\theta}_d) = \arg \min_{k_d > 0,\ \theta_d \geq 0} \sum_{t \in \mathcal{T}_d} \left[H_d^{\text{obs}}(t) - H_d^{\text{model}}(t; k_d, \theta_d) \right]^2 . \tag{9}$$

We fit in cumulative-hazard space rather than maximizing a likelihood because the primary inputs are discrete-time, cohort-aggregated hazards and the objective is stable estimation of selection-induced depletion curvature during quiet periods. Least squares on $H_d^{\text{obs}}(t)$ is numerically robust under sparse events, emphasizes shape agreement over the fit window, and yields diagnostics (e.g., RMSE in H-space) that directly reflect the quality of the depletion fit. Likelihood-based fitting can be treated as a sensitivity analysis, but is not required for the normalization identity itself.

Reference implementation defaults (this repo)

The manuscript describes KCOR generically; for reproducibility, this repository's KCOR v6 defaults are:

- **Quiet window**: ISO weeks 2022–24 through 2024–16 (inclusive).
- **Skip weeks**: a fixed prespecified skip, `SKIP_WEEKS = DYNAMIC_HVE_SKIP_WEEKS` (see code), applied by setting $h_d^{\text{eff}}(t) = 0$ for $t <$ SKIP_WEEKS.
- **Observed hazard transform used to build H_d^{obs} from weekly mortality risk**: if $\text{MR}_{d,t} = D_{d,t}/Y_{d,t}$ is the weekly mortality input,

$$h_d^{\text{obs}}(t) = -\log\left(\frac{1 - 1.5\,\text{MR}_{d,t}}{1 - 0.5\,\text{MR}_{d,t}} \right). \tag{10}$$

This form provides a second-order accurate approximation to the continuous-time hazard while remaining numerically stable for small mortality-rate (MR) values.

- **Fit method**: nonlinear least squares in cumulative-hazard space (not MLE), with constraints $k_d > 0$ and $\theta_d \geq 0$.
- **Cohort indexing (implementation)**: enrollment period (sheet) × YearOfBirth group × Dose, plus an all-ages cohort (YearOfBirth = −2).

2.6 Normalization (depletion-neutralized cumulative hazards)

After fitting, KCOR computes the depletion-neutralized baseline cumulative hazard for each cohort d by applying the inversion to the full post-enrollment trajectory:

$$\tilde{H}_{0,d}(t) = \frac{e^{\hat{\theta}_d H_d^{\text{obs}}(t)} - 1}{\hat{\theta}_d}. \tag{11}$$

This normalization maps each cohort into a depletion-neutralized baseline-hazard space in which the contribution of gamma frailty parameters (θ_d, k_d) to hazard curvature has been factored out. *(Conceptually, this places all cohorts into an equivalent θ-factored comparison space.)* In this space, cumulative hazards are directly comparable across cohorts, and remaining differences reflect real differences in baseline risk rather than selection-induced depletion.

When $\hat{\theta}_d \approx 0$, this mapping leaves the observed cumulative hazard essentially unchanged, whereas cohorts with substantial depletion undergo a pronounced upward correction.

An adjusted hazard may be recovered by differencing:

$$\tilde{h}_{0,d}(t) \approx \tilde{H}_{0,d}(t) - \tilde{H}_{0,d}(t-1). \tag{12}$$

The key object for KCOR is $\tilde{H}_{0,d}(t)$; differenced hazards are optional diagnostics.

Normalization is necessary but not sufficient. The depletion-neutralized cumulative hazard $\tilde{H}_{0,d}(t)$ is not itself the estimand of interest. Its role is to place cohorts into a common comparison space in which selection-induced depletion dynamics have been removed. The substantive comparison—and therefore the inferential signal—arises only when

these normalized cumulative hazards are compared across cohorts via the KCOR estimator (§2.8). Because normalization operates in cumulative-hazard space and removes time-varying curvature rather than rescaling instantaneous hazards, applying Cox regression to normalized outputs generally re-introduces misspecification. Applying standard proportional-hazards or regression-based estimators after normalization is generally inappropriate, because the comparison is cumulative by construction and because residual non-proportionality is precisely what KCOR is designed to reveal. KCOR therefore integrates normalization and comparison into a single, internally consistent system.

2.6.1 Internal diagnostics and 'self-check' behavior

KCOR includes internal diagnostics intended to make model stress visible rather than hidden.

1. **Post-normalization linearity in quiet periods.** During a prespecified quiet window, the working model assumes that curvature in observed cumulative hazard is primarily driven by depletion under heterogeneity. After inversion, the depletion-neutralized cumulative hazard $\tilde{H}_{0,d}(t)$ should be approximately linear in event time over the same quiet window. Systematic residual curvature (e.g., sustained concavity/convexity) indicates that the quiet-window assumption is violated (external shocks, secular trends) or that the depletion geometry is misspecified for that cohort.

2. **Fit residual structure in cumulative-hazard space.** Define residuals $r_d(t) = H_d^{\mathrm{obs}}(t) - H_d^{\mathrm{model}}(t; \hat{k}_d, \hat{\theta}_d)$ over the fit set $\mathcal{T}_d$. KCOR expects residuals to be small and not systematically time-structured. Strongly patterned residuals indicate that the curvature attributed to depletion is instead being driven by unmodeled time-varying hazards.

3. **Parameter stability to window perturbations.** Under valid quiet-window selection, $(\hat{k}_d, \hat{\theta}_d)$ should be stable to small perturbations of the quiet-window boundaries (e.g., ±4 weeks). Large changes in $\hat{\theta}_d$ under small boundary shifts signal that the fitted curvature is sensitive to transient dynamics rather than stable depletion.

4. **Non-identifiability manifests as $\hat{\theta} \to 0$.** When the observed cumulative hazard is near-linear (weak curvature) or events are sparse, θ is weakly identified. In such cases, KCOR should be interpreted primarily as a diagnostic (limited evidence of detectable depletion curvature) rather than a strong correction.

These diagnostics are reported alongside KCOR curves. Importantly, the goal is not to assert that a single parametric form is always correct, but to ensure that when the form is incorrect or the window is contaminated, the method signals this explicitly rather than silently producing a misleading 'corrected' estimate.

2.7 Stabilization (early weeks)

In many applications, the first few post-enrollment intervals can be unstable due to immediate post-enrollment artifacts (e.g., rapid deferral, short-term sorting, administrative effects). KCOR supports a prespecified stabilization rule by excluding early weeks from accumulation and from quiet-window fitting. The skip-weeks parameter is prespecified and evaluated via sensitivity analysis to exclude early enrollment instability rather than to tune estimates.

In discrete time, define an effective hazard for accumulation:

$$h_d^{\mathrm{eff}}(t) = \begin{cases} 0, & t < \text{SKIP\_WEEKS} \\ h_d^{\mathrm{obs}}(t), & t \geq \text{SKIP\_WEEKS}. \end{cases} \tag{13}$$

Then compute $H_d^{\mathrm{obs}}(t)$ from $h_d^{\mathrm{eff}}(t)$ as in §2.3.

2.8 KCOR estimator (v6)

For cohorts A and B, KCOR compares depletion-neutralized cumulative hazards:

$$\mathrm{KCOR}(t) = \frac{\tilde{H}_{0,A}(t)}{\tilde{H}_{0,B}(t)}. \tag{14}$$

This is a cumulative comparison in hazard space after removing cohort-specific selection-induced depletion dynamics estimated during quiet periods.

2.9 Uncertainty quantification

KCOR can be equipped with uncertainty intervals via:

- **Analytic variance propagation** for cumulative hazards combined with delta-method approximations for the ratio, and/or
- **Monte Carlo resampling** to capture uncertainty from event realization and from estimation of (k_d, θ_d) and the resulting normalization.

Uncertainty intervals reflect stochastic event realization and model-fit uncertainty in the selection-parameter estimation. They do not assume sampling from a superpopulation and may be interpreted as uncertainty conditional on the observed risk sets and modeling assumptions.

2.10 Algorithm summary and reproducibility checklist

Table tbl. 5 summarizes the complete KCOR v6 pipeline.

Table 5: Step-by-step KCOR v6 algorithm (high-level), with recommended prespecification and diagnostics.

Step	Operation	Output	Prespecify?	Diagnostics
1	Choose enrollment date and define fixed cohorts	Cohort labels	Yes	Verify cohort sizes/risk sets
2	Compute discrete-time hazards $h_d^{\text{obs}}(t)$	Hazard curves	Yes (binning/transform)	Check for zeros/sparsity
3	Apply stabilization skip and accumulate $H_d^{\text{obs}}(t)$	Observed cumulative hazards	Yes (skip rule)	Plot $H_d^{\text{obs}}(t)$
4	Select quiet-window bins in calendar ISO-week space	Fit points $\mathcal{T}_d$	Yes	Overlay quiet window on hazard plots
5	Fit (k_d, θ_d) via cumulative-hazard least squares	$\hat{k}_d, \hat{\theta}_d$	Yes	RMSE, residuals, fit stability
6	Normalize: invert gamma-frailty identity to $\tilde{H}_{0,d}(t)$	Depletion-neutralized cumulative hazards	Yes	Compare pre/post shapes; sanity checks
7	Cumulate and ratio: compute KCOR(t)	KCOR(t) curve	Yes (horizon)	Flat under negative controls
8	Uncertainty	CI / intervals	Yes	Coverage on positive controls

2.11 Relationship to Cox proportional hazards

Cox proportional hazards models estimate instantaneous hazard under the assumption of time-invariant hazard ratios. In observational cohorts with selective uptake and frailty heterogeneity, this assumption is structurally violated, leading to time-varying hazard ratios and cumulative hazard trajectories inconsistent with observed data. Cox estimates are
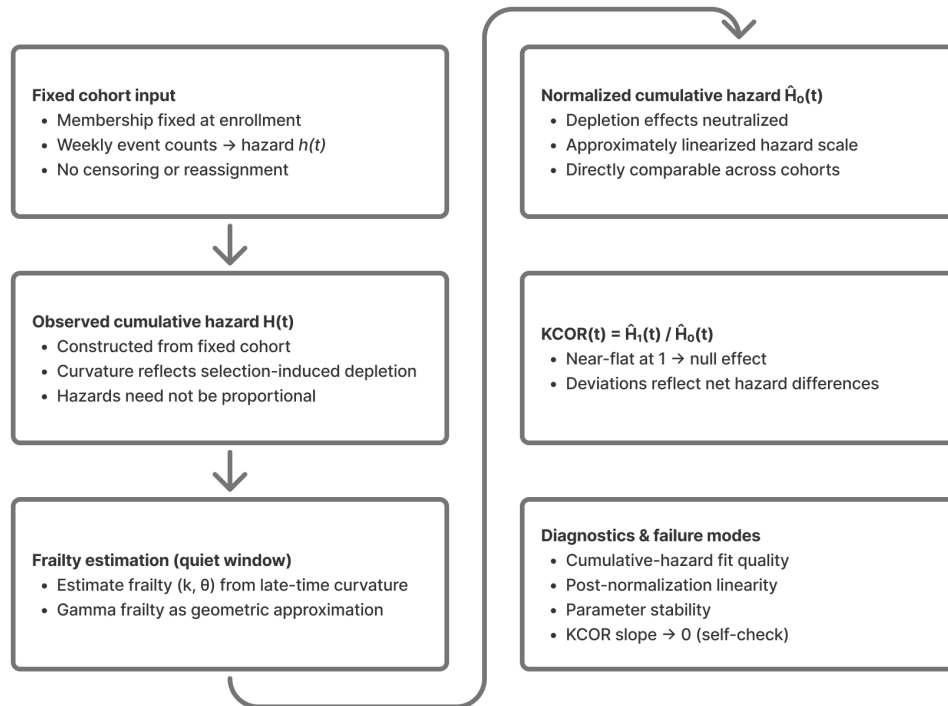
Fixed cohort input
- Membership fixed at enrollment
- Weekly event counts → hazard $h(t)$
- No censoring or reassignment

Observed cumulative hazard H(t)
- Constructed from fixed cohort
- Curvature reflects selection-induced depletion
- Hazards need not be proportional

Frailty estimation (quiet window)
- Estimate frailty (k, θ) from late-time curvature
- Gamma frailty as geometric approximation

Normalized cumulative hazard $\hat{H}_o(t)$
- Depletion effects neutralized
- Approximately linearized hazard scale
- Directly comparable across cohorts

KCOR(t) = $\hat{H}_1(t) / \hat{H}_o(t)$
- Near-flat at 1 → null effect
- Deviations reflect net hazard differences

Diagnostics & failure modes
- Cumulative-hazard fit quality
- Post-normalization linearity
- Parameter stability
- KCOR slope → 0 (self-check)

Figure 2: KCOR as an integrated depletion-neutralized comparison system. The KCOR pipeline operates as a single, end-to-end system. Observed cohort cumulative hazards are first mapped into depletion-neutralized hazard space via gamma-frailty inversion. This normalization step alone does not constitute inference. The KCOR estimator then compares normalized cumulative hazards via a cumulative ratio, which is the estimand that determines whether one cohort experienced higher or lower cumulative event risk than another under the stated assumptions. Diagnostic checks shown alongside the workflow indicate when depletion-neutralization is valid and when results should be interpreted cautiously.

therefore presented here solely for diagnostic illustration to demonstrate assumption failure, not as a competing causal estimator.

3. Validation and control tests

This section is the core validation claim of KCOR:

- **Negative controls (null under selection):** under a true null effect, KCOR remains approximately flat at 1 even when selection induces large curvature differences.
- **Positive controls (detect injected effects):** when known harm/benefit is injected into otherwise-null data, KCOR reliably detects it.

Age band (years)	$\hat{\theta}$ Dose 0 (median)	$\hat{\theta}$ Dose 2 (median)
40–49	16.79	2.66×10^{-6}
50–59	23.02	1.87×10^{-4}
60–69	13.13	7.01×10^{-18}
70+	1.89	4.50×10^{-16}

Table 2. Estimated gamma-frailty variance ($\hat{\theta}$) by age band and vaccination status for Czech cohorts enrolled in 2021_24. {#tbl:frailty_diagnostics}

$\hat{\theta}$ quantifies unobserved frailty heterogeneity and depletion of susceptibles within cohorts. Near-zero values indicate effectively linear cumulative hazards over the quiet window and are typical of strongly pre-selected cohorts. Values are summarized as medians across enrollment subcohorts within 2021_24.

As shown in Table 2, unvaccinated cohorts exhibit substantial frailty heterogeneity ($\hat{\theta} > 0$), while Dose 2 cohorts show near-zero estimated frailty ($\hat{\theta} \approx 0$) across all age bands, consistent with strong selective uptake prior to follow-up. Frailty variance is largest at younger ages, where low baseline mortality amplifies the impact of heterogeneity on cumulative hazard curvature, and declines at older ages where mortality is compressed and survivors are more homogeneous. No diagnostic reversals or instabilities are observed.

Given the strong frailty heterogeneity shown in Table 2, raw cumulative outcome contrasts (Table 3) are expected to reflect both selection-induced depletion effects and any underlying treatment differences. KCOR normalization removes the depletion component, enabling interpretable comparison of the remaining differences.

Age band (years)	Dose 0 cumulative hazard	Dose 2 cumulative hazard	Ratio
40–49	0.022731	0.012510	1.8171
50–59	0.060750	0.035750	1.6993
60–69	0.204957	0.083450	2.4561
70+	1.347411	0.869921	1.5489

Table 3. Ratio of observed cumulative mortality hazards for unvaccinated (Dose 0) versus fully vaccinated (Dose 2) Czech cohorts enrolled in 2021_24. {#tbl:raw_cumulative_outcomes}

Values reflect raw cumulative outcome differences prior to KCOR normalization and are not interpreted causally due to cohort non-exchangeability. Cumulative hazards were integrated from cohort enrollment through the end of available follow-up (week 2024-16), identically for Dose 0 and Dose 2 cohorts.

3.0 Frailty normalization behavior under empirical validation

Across examined age strata in the Czech Republic mortality dataset, fitted frailty parameters exhibit a pronounced asymmetry across cohorts. Some cohorts show negligible estimated frailty variance ($\hat{\theta} \approx 0$), while others exhibit substantial frailty-driven depletion. This pattern reflects differences in selection-induced hazard curvature at cohort entry rather than any prespecified cohort identity.

As a consequence, KCOR normalization leaves some cohorts' cumulative hazards nearly unchanged, while substantially increasing the depletion-neutralized cumulative hazard for others. This behavior is consistent with curvature-driven normalization rather than cohort identity. This pattern is visible directly in adjusted versus unadjusted cumulative hazard plots and is summarized quantitatively in the fitted-parameter logs (see `KCOR_summary.log`).

After frailty normalization, the adjusted cumulative hazards $\tilde{H}_{0,d}(t)$ are approximately linear in event time. Residual deviations from linearity reflect real time-varying risk—such as seasonality or epidemic waves—rather than selection-induced depletion. This linearization is a diagnostic consistent with successful removal of depletion-driven curvature under the working model; persistent nonlinearity or parameter instability indicates model stress or quiet-window contamination.

3.1 Negative controls: null under selection-induced curvature

3.1.1 Fully synthetic negative control (recommended)

Design a simulation where:

- Individual hazards follow a baseline hazard $h_0(t)$ multiplied by frailty z.
- Two cohorts have different frailty variance θ (or different selection rules), creating different cohort-level curvature in $h_d^{\mathrm{obs}}(t)$.
- **No treatment effect is applied** (both cohorts share the same baseline hazard $h_0(t)$).

After estimating (k_d, θ_d) during quiet periods and applying gamma-frailty inversion, KCOR should remain approximately constant at 1 over follow-up.

In-model "gamma-frailty" stress test (highly convincing null)

One especially clear falsification test is an **in-model gamma-frailty null**: simulate data directly from the gamma-frailty model with the same $h_0(t)$ but different θ between cohorts. This induces strong, visibly different hazard curvature from depletion alone. Because the data-generating process matches the model, the fitted normalization is exact up to sampling noise, and KCOR should be flat at 1.

Suggested construction (example):

- Time unit: weeks.
- Baseline hazard shape: $g(t) = 1$ during quiet periods; choose a baseline level k in the chosen time units.
- Cohort A: $\theta_A > 0$ (stronger depletion).
- Cohort B: $\theta_B > 0$ (weaker depletion).

Figure fig. 3 shows this construction.

3.1.2 Empirical "within-category" negative control (already implemented in repo)

The repository includes a pragmatic negative control construction that repurposes a real dataset by comparing "like with like" while inducing large composition differences (e.g., age band shifts). In this construction, age strata are remapped into pseudo-doses so that comparisons are, by construction, within the same underlying category; the expected differential effect is near zero, but the baseline hazards differ strongly.

These age-shift negative controls deliberately induce extreme baseline mortality differences (10–20 year age gaps) while preserving a true null effect by construction, since all vaccination states are compared symmetrically. The near-flat KCOR trajectories are consistent with the estimator normalizing selection-induced depletion curvature without introducing spurious time trends or cumulative drift.

For the empirical age-shift negative control (Figures fig. 4 and fig. 5), we use aggregated weekly cohort summaries derived from the Czech Republic administrative mortality and vaccination dataset and exported in KCOR_CMR format.

Notably, KCOR estimates frailty parameters independently for each cohort without knowledge of exposure status; the observed asymmetry in depletion correction arises entirely from differences in hazard curvature rather than from any vaccination-specific assumptions.

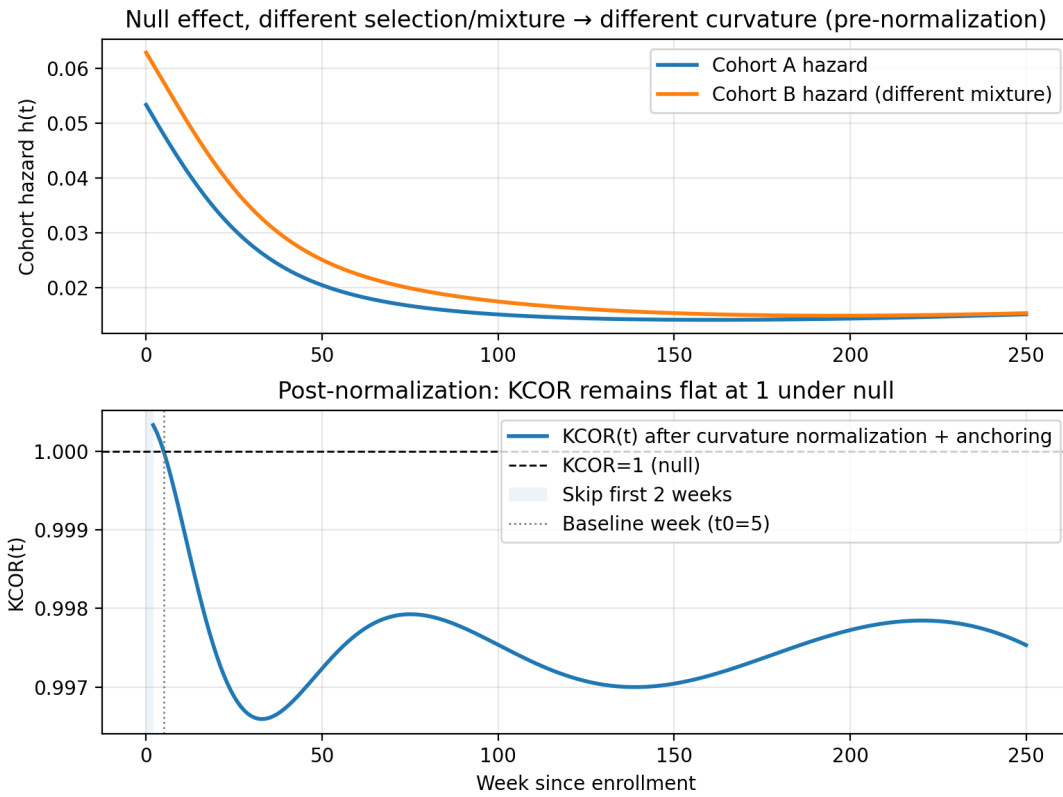Two snapshots illustrate that KCOR is near-flat even under 10–20 year age differences:

Figure 3: Synthetic negative control under strong selection (different curvature) but no effect: KCOR remains flat at 1. Top panel shows cohort hazards with different frailty-mixture weights inducing different curvature. Bottom panel shows KCOR(t) remaining near 1.0 after normalization, demonstrating successful depletion-neutralization under the null.
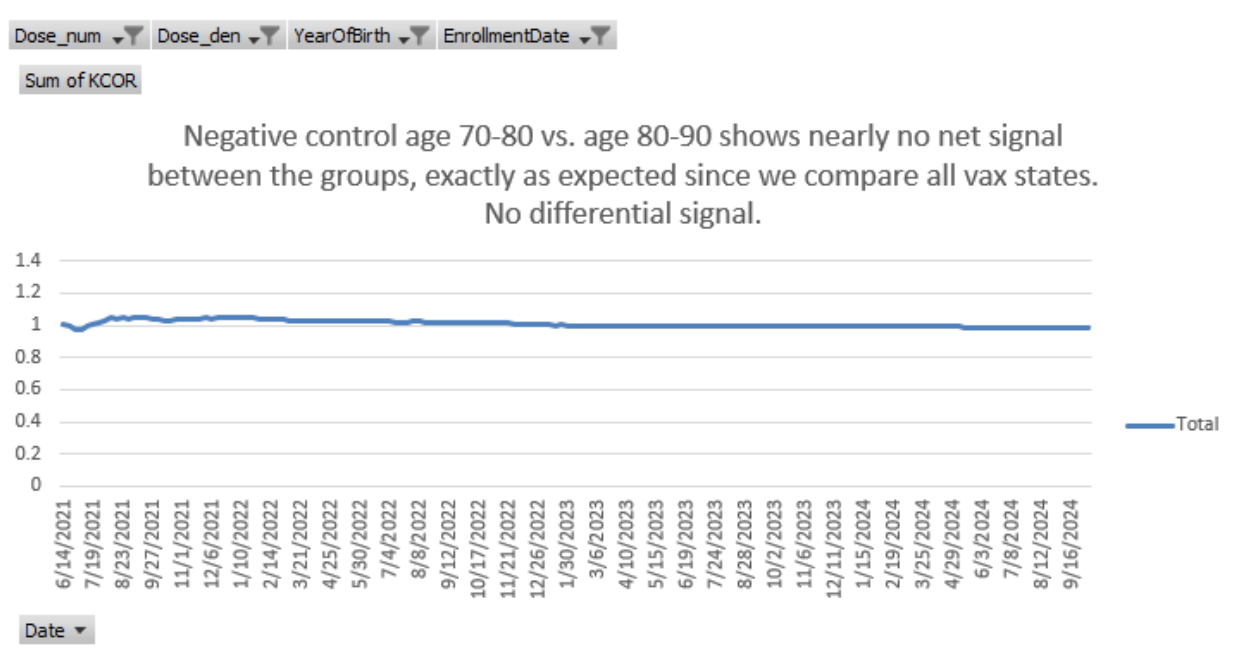
Figure 4: Empirical negative control with approximately 10-year age difference between cohorts. Despite large baseline mortality differences, KCOR remains near-flat at 1 over follow-up, consistent with a true null effect. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Appendix B.2)[4].

Figure 5: Empirical negative control with approximately 20-year age difference between cohorts. Even under extreme composition differences, KCOR exhibits no systematic drift, demonstrating robustness to selection-induced curvature. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Appendix B.2).

Table tbl. 8 provides numeric summaries.

Table 8: Example end-of-window KCOR values from the empirical negative control (pooled/ASMR summaries), showing near-null behavior under large composition differences. (Source: `test/negative_control/out/KCOR_summary.log`)

Enrollment	Dose comparison	KCOR (pooled/ASMR)	95% CI
2021_24	1 vs 0	1.0097	[0.992, 1.027]
2021_24	2 vs 0	1.0213	[1.000, 1.043]
2021_24	2 vs 1	1.0115	[0.991, 1.033]
2022_06	1 vs 0	0.9858	[0.970, 1.002]
2022_06	2 vs 0	1.0756	[1.055, 1.097]
2022_06	2 vs 1	1.0911	[1.070, 1.112]

3.2 Positive controls: detect injected harm/benefit

The effect window is a simulation construct used solely for positive-control validation and does not represent a real-world intervention period or biological effect window.

Positive controls are constructed by starting from a negative-control dataset and injecting a known effect into the data-generating process for one cohort, for example by multiplying the *baseline* hazard by a constant factor r over a prespecified interval:

$$h_{0,\text{treated}}(t) = r \cdot h_{0,\text{control}}(t) \quad \text{for } t \in [t_1, t_2], \tag{15}$$

with $r > 1$ for harm and $0 < r < 1$ for benefit.

After gamma-frailty normalization (inversion), KCOR should deviate from 1 in the correct direction and with magnitude consistent with the injected effect (up to discretization and sampling noise). Figure fig. 6 and Table tbl. 9 confirm this behavior.

Table 9: Positive control results comparing injected hazard multipliers to detected KCOR deviations. Both scenarios show KCOR deviating from 1.0 in the expected direction, validating that the estimator can detect true effects.

Scenario	Effect window	Hazard multiplier r	Expected direction	Observed KCOR at week 80
Benefit	week 20–80	0.8	< 1	0.825
Harm	week 20–80	1.2	> 1	1.107

3.3 Sensitivity analyses (robustness checks)

The primary analysis uses a prespecified quiet window applied uniformly across cohorts; sensitivity analyses explicitly vary quiet-window bounds and related prespecified choices to assess robustness.

KCOR results should be robust (up to numerical tolerance) to reasonable variations in:

- Quiet-window selection (calendar ISO-week bounds)
- Stabilization skip (early-bin handling)
- Time-binning resolution
- Age stratification and/or stratified analyses where appropriate
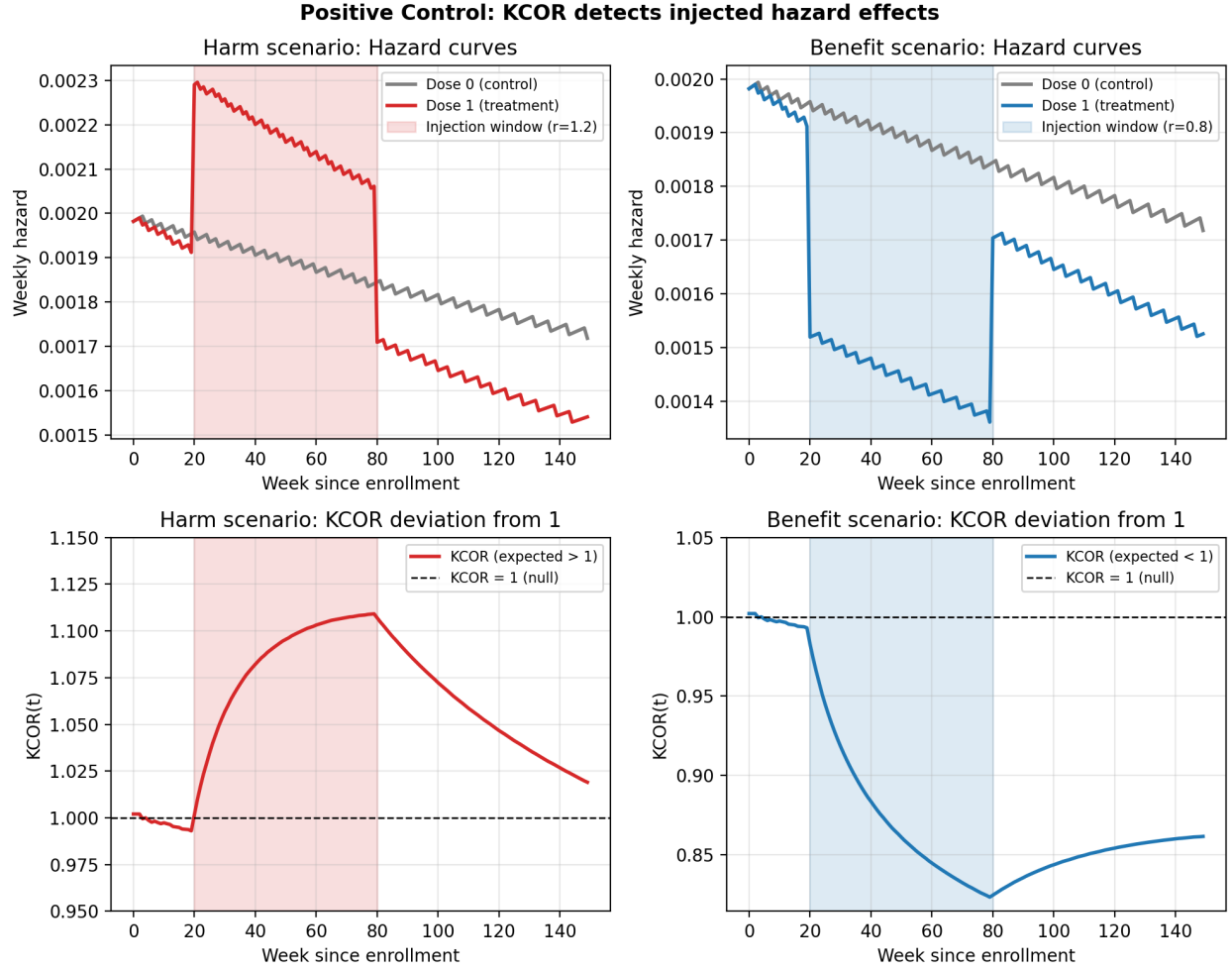- Baseline shape choice $g(t)$ (default $g(t) = 1$; alternatives can be assessed as sensitivity)

Figure 6: Positive control validation: KCOR correctly detects injected effects. Left panels show harm scenario (r=1.2), right panels show benefit scenario (r=0.8). Top row displays cohort hazard curves with effect window shaded. Bottom row shows KCOR(t) deviating from 1.0 in the expected direction during the effect window.
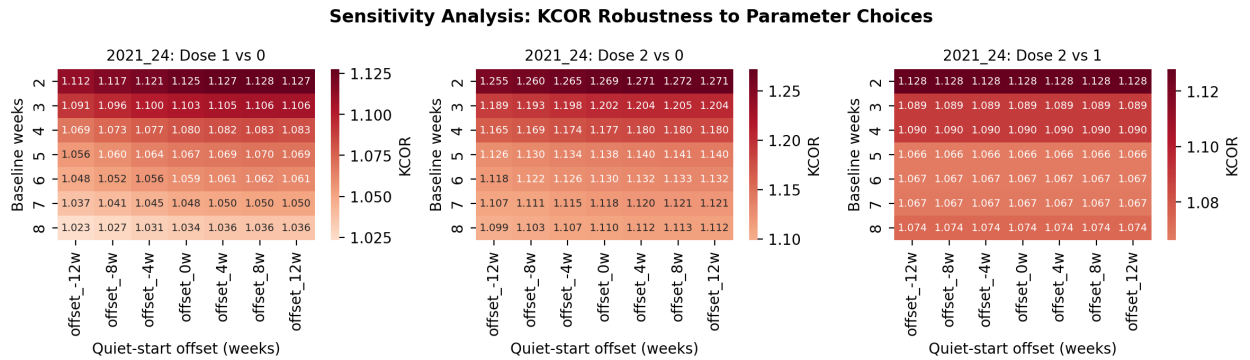

Figure 7: Sensitivity analysis summary showing KCOR values across parameter grid. Heatmaps display KCOR estimates for different combinations of baseline weeks (rows) and quiet-window start offsets (columns). Color scale centered at 1.0 shows stability of estimates across parameter choices, with values remaining close to 1.0 across the grid.

Figure fig. 7 summarizes KCOR stability across the parameter grid.

Across all tested parameter ranges, KCOR values remained within approximately ±5% of unity, indicating stability under reasonable variations in fitting choices.

3.4 Simulation grid: operating characteristics and failure-mode diagnostics

We further evaluate KCOR using a compact simulation grid designed to (i) confirm near-null behavior under selection-induced curvature, (ii) confirm detection of injected effects, and (iii) characterize failure modes and diagnostics under model misspecification and adverse data regimes. Each scenario generates cohort-level weekly counts in KCOR_CMR format. KCOR is then fit using the same prespecified quiet-window procedure as in the empirical analyses, and we report both KCOR(t) trajectories and diagnostic summaries, including cumulative-hazard fit error and post-normalization linearity. The scenarios isolate specific stresses, including non-gamma frailty, contamination of the quiet window by an external shock, sparse events, **joint frailty and treatment effects (S7)**, and **tail-sampling / bimodal selection** (cohorts drawn from different parts of the same underlying frailty distribution, e.g., vaccinated sampled from mid-quantiles; un-vaccinated from low+high tails, producing non-gamma mixture geometry at the cohort level). Code to reproduce all simulations and figures is included in the repository. *Near-flat* is defined operationally as median KCOR(t) remaining within ±5% of unity over the diagnostic window (weeks 20–100), excluding early transients. See Figures fig. 8 and fig. 9 below.

Joint frailty and treatment-effect simulation (S7).
Figure fig. 10 summarizes results from the S7 simulation, in which cohorts differ in frailty-driven depletion dynamics and a known treatment effect is introduced in a separate time window. Frailty parameters are estimated exclusively during a prespecified quiet window, and KCOR normalization is then applied to the full follow-up period.

When the quiet window and treatment window are temporally separable, KCOR exhibits the expected behavior under identifiable conditions: KCOR(t) remains approximately flat and near unity throughout the quiet window, indicating successful identification and removal of selection-induced depletion curvature, and deviates from unity only during the treatment window, in the correct direction and with magnitude consistent with the injected effect (harm or benefit). This behavior holds across multiple effect shapes (step, ramp, smooth pulse) and effect magnitudes.

Importantly, S7 also includes an intentional violation of temporal separability in which the treatment window overlaps the quiet window. In this overlap variant, KCOR trajectories no longer stabilize during the nominal quiet period, and fit diagnostics degrade (Figure fig. 11), including increased cumulative-hazard fit error and reduced post-normalization linearity. In these cases, KCOR does not produce a spurious treatment signal; instead, diagnostics correctly indicate that the assumptions required for interpretable normalization are violated.

Together, these results demonstrate KCOR's operating characteristics under joint selection and treatment dynamics: when selection-induced depletion and treatment effects are sufficiently separable in time, KCOR can disentangle the two mechanisms; when they are not, the method fails transparently via its diagnostics rather than silently misattributing curvature.

The tail-sampling scenario is included because it can confound frailty-driven depletion with cohort construction in ways not captured by a single gamma frailty distribution. The goal is not to force KCOR to 'succeed' under arbitrary misspecification, but to quantify operating characteristics: when the gamma depletion model is misspecified, KCOR should either (i) remain approximately unbiased in later windows (if the misspecification is mild in cumulative-hazard geometry), or (ii) visibly degrade via its diagnostics (poor H-space fit, post-normalization nonlinearity, parameter instability), flagging that depletion-neutralization is unreliable without model generalization.

3.5 Dynamic HVE diagnostic tests

Dynamic HVE refers to transient hazard suppression immediately after enrollment driven by short-horizon selection around intervention timing (e.g., deferral during illness). It produces a characteristic early-time pattern: an abrupt early reduction in observed hazard that decays over several weeks and is not explained by stable depletion curvature.

Empirical signature in multi-dose settings (diagnostic, not proof). When multiple 'treatment intensities' exist (e.g., dose-2 and dose-3 cohorts defined at enrollment), dynamic HVE should affect adjacent-dose cohorts similarly at early
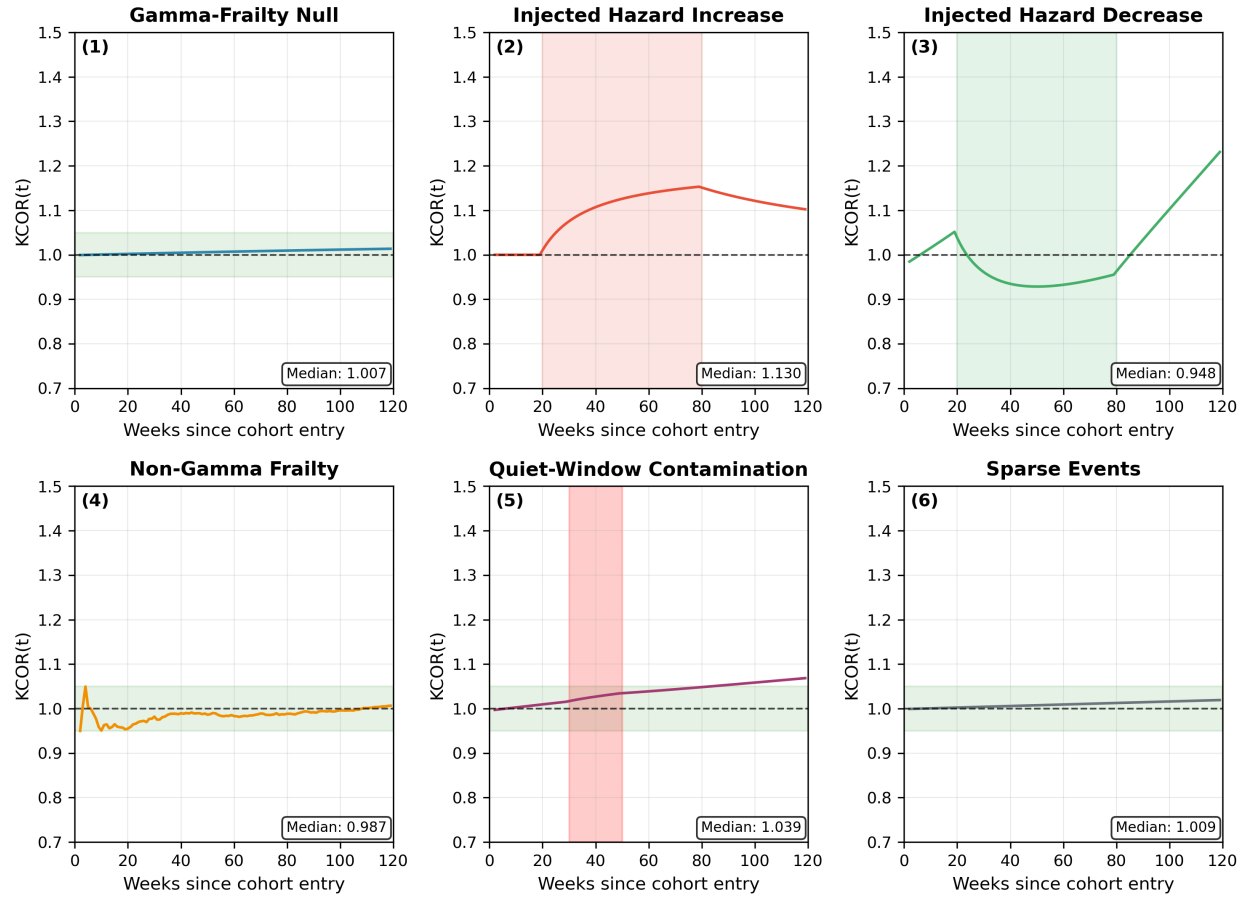
Figure 8: Simulation grid overview: KCOR(t) trajectories across prespecified scenarios, including gamma-frailty null with strong selection, injected hazard increase and decrease, non-gamma frailty, quiet-window contamination, and sparse-event regimes. Under true null, KCOR remains near-flat at 1; injected effects are detected in the expected direction; adverse regimes are accompanied by degraded diagnostics and reduced interpretability.
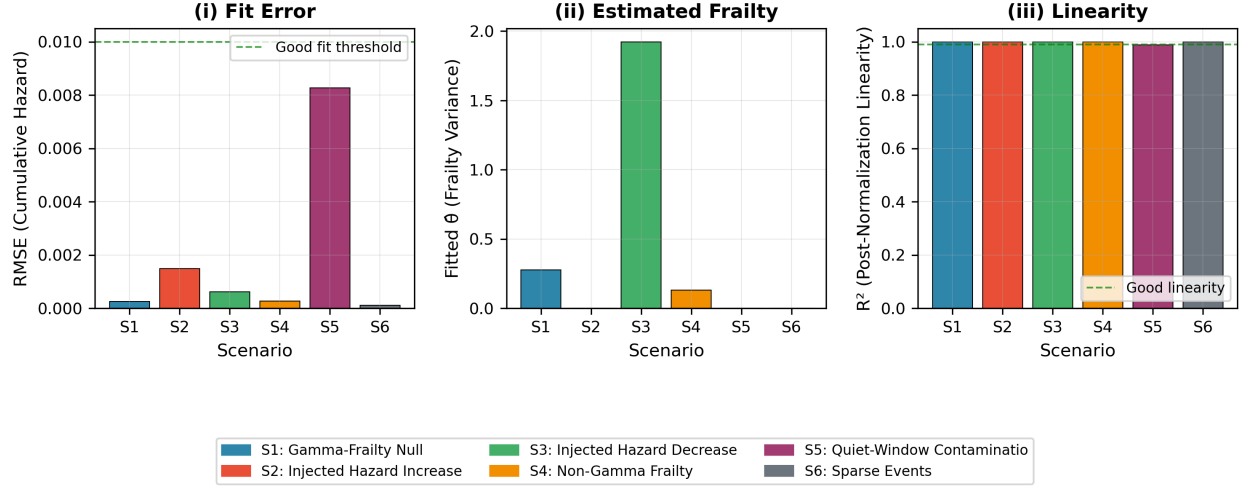
KCOR Simulation Grid: Diagnostic Summary

Figure 9: Simulation diagnostics across scenarios: (i) cumulative-hazard fit RMSE over the quiet window, (ii) fitted frailty variance estimates, and (iii) a post-normalization linearity metric for normalized cumulative hazards. Diagnostics identify regimes in which frailty normalization is well identified versus weakly identified.
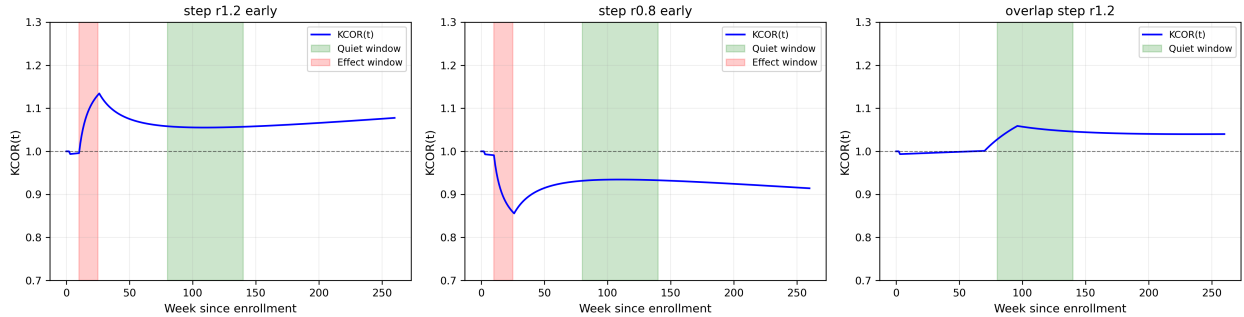

Figure 10: S7 simulation results: KCOR(t) trajectories demonstrating temporal separability. Left panel shows harm scenario (r=1.2) with effect window (weeks 10-25) and quiet window (weeks 80-140) non-overlapping. Middle panel shows benefit scenario (r=0.8). Right panel shows overlap variant where effect window intersects quiet window, demonstrating diagnostic degradation. KCOR remains approximately flat during the quiet window and deviates only during the effect window when temporal separability holds. The overlap variant is included to demonstrate failure-mode behavior and should not be interpreted as a valid application regime for KCOR.
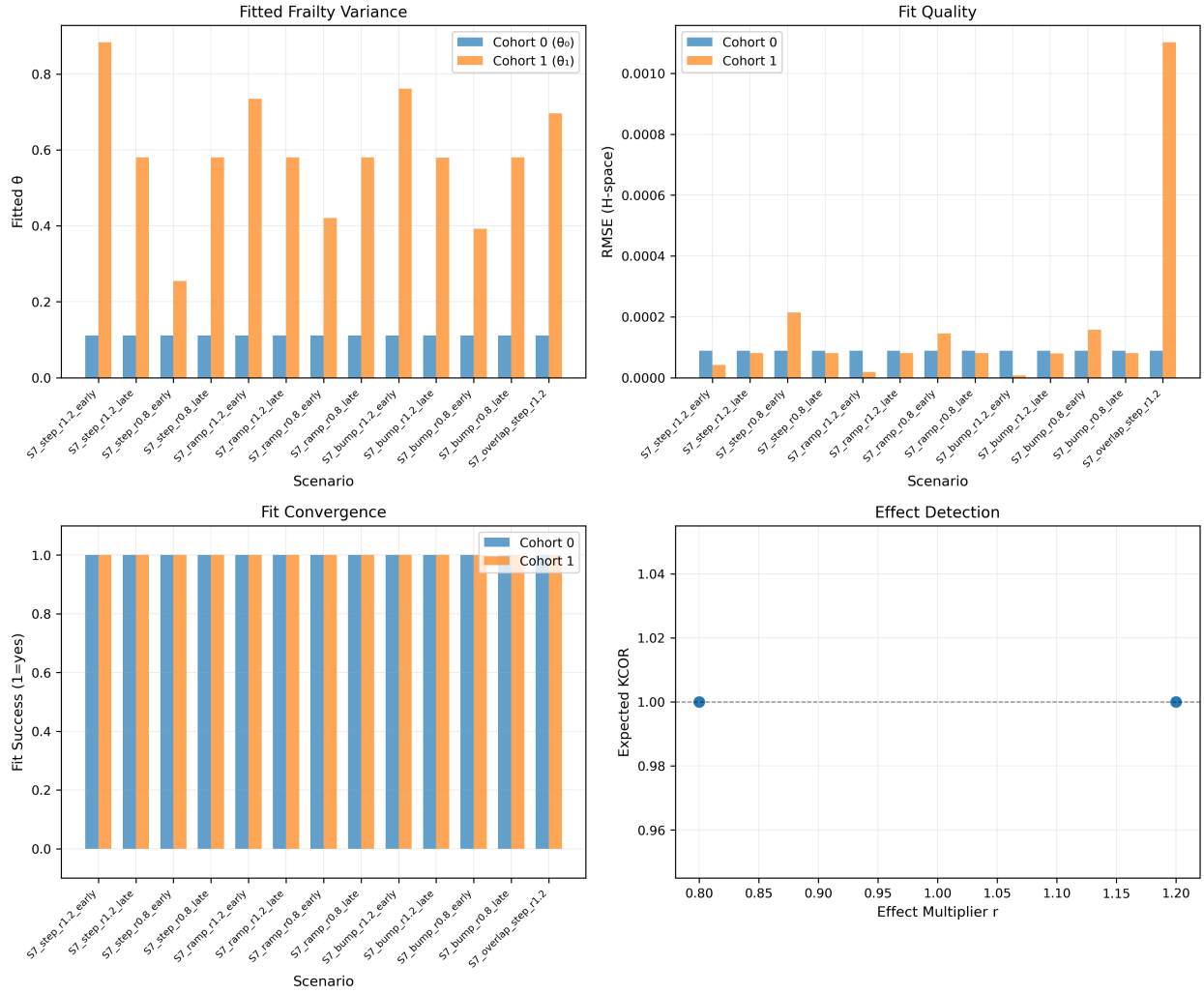
Figure 11: S7 simulation diagnostics: Fitted frailty variance parameters (θ_0, θ_1), fit quality (RMSE), and convergence status across S7 scenarios. The overlap variant shows degraded fit quality, correctly signaling violation of temporal separability assumptions.

times because both enrollments are subject to the same short-horizon deferral mechanisms. Therefore, if early post-enrollment curvature is dominated by dynamic HVE, then early-time deviations in KCOR(t) versus the same comparator should show similar transient shapes across adjacent-dose cohorts. Conversely, if early-time behavior differs substantially across adjacent-dose cohorts while post-normalization quiet-window linearity holds, it is less consistent with a single shared dynamic deferral artifact.

Simulation check. We include simulations where a transient early hazard suppression is injected around enrollment (multiplying hazard by factor $q < 1$ for weeks 0–S), separately from gamma frailty depletion, and confirm that (i) the effect is attenuated/removed by prespecified skip weeks, and (ii) remaining KCOR trajectories in later windows behave as expected under negative and positive controls.

3.6 Illustrative non-COVID example (synthetic)

To emphasize that KCOR is not specific to COVID-19 vaccination, we include a synthetic illustration motivated by elective intervention timing. Consider two cohorts defined by the timing of an elective medical procedure, where short-term deferral during acute illness induces selection into the later-treated cohort. Although no treatment effect is present by construction, the observed cumulative hazards differ due to selection-induced depletion.

Applying KCOR to this setting removes curvature attributable to depletion and yields a flat post-normalization trajectory, with KCOR(t) asymptoting to unity as expected under the null. This example demonstrates that KCOR applies generally to retrospective cohort comparisons affected by selection-induced hazard curvature, independent of disease area or intervention type.

---

4. Discussion

What KCOR does not provide

KCOR is designed to resolve a specific and otherwise unaddressed failure mode in retrospective analyses—selection-induced depletion under latent heterogeneity. Accordingly, KCOR does **not** by itself provide:

• Policy optimization or cost-benefit analysis • Transportability of effects across populations without additional assumptions • Identification under unmeasured time-varying confounding unrelated to depletion dynamics

These limitations are intrinsic to the data constraints KCOR is designed to operate under and do not detract from its role as a depletion-neutralized cohort comparison system.

4.1 What KCOR estimates

KCOR operates at a specific but critical layer of the retrospective inference stack: it both neutralizes selection-induced depletion dynamics and defines how the resulting depletion-neutralized hazards must be compared. The method's strength is not the frailty inversion in isolation, but the fact that inversion, diagnostics, and cumulative comparison are mathematically and operationally coupled. Once cohorts are mapped into depletion-neutralized hazard space, KCOR(t) directly answers whether one cohort experienced higher or lower cumulative event risk than another over follow-up, conditional on the stated assumptions. Interpreting normalized hazards without this comparison step discards the central inferential content of the method. Like all causal methods, KCOR relies on identifiable structure; its contribution is to make that structure explicit, testable, and diagnostically enforced rather than implicit. Accordingly, KCOR should not be interpreted as a standalone estimator of vaccine efficacy; it is a cumulative normalization-and-comparison framework whose causal interpretation depends on additional design assumptions, controls, and diagnostics.

As emphasized above, the frailty term is not causal and does not represent a treatment mechanism; it functions solely as a geometric normalization for selection-induced depletion.

KCOR is a **cumulative** comparison of depletion-neutralized cumulative hazards; it does not estimate instantaneous hazard ratios. It is designed for settings where selection induces non-proportional hazards such that conventional proportional-hazards estimators can be difficult to interpret. We did not pursue model selection among Cox-based

specifications (with or without frailty) because these models target instantaneous hazard ratios under proportional-hazards assumptions, whereas KCOR targets cumulative, depletion-neutralized outcomes; BIC comparisons across models with different estimands are therefore not informative for the question addressed here.

Under the working assumptions that:

1. selection-induced depletion dynamics can be estimated during quiet periods using a gamma-frailty mixture model, and
2. the fitted selection parameters can be used to invert observed cumulative hazards into depletion-neutralized baseline cumulative hazards,

then the remaining differences between cohorts are interpretable, **conditional on the stated selection model and quiet-window validity**, as differences in baseline hazard level (on a cumulative scale), summarized by $KCOR(t)$.

The observation that frailty correction is negligible for vaccinated cohorts but substantial for the unvaccinated cohort is not incidental. It reflects the asymmetric action of healthy-vaccinee selection, which concentrates lower-frailty individuals into vaccinated cohorts at enrollment while leaving the unvaccinated cohort heterogeneous. KCOR explicitly detects and removes this asymmetry by mapping cohorts into a depletion-neutralized comparison space rather than assuming proportional hazards.

Because the normalization targets selection-induced depletion curvature, KCOR results alone do not justify claims about net lives saved or lost by a particular intervention. Such claims require (i) clearly specified causal estimands, (ii) validated control outcomes, (iii) sensitivity analyses for remaining time-varying selection mechanisms and external shocks, and (iv) preferably replication across settings and outcomes. Accordingly, this manuscript focuses on method definition, diagnostics, and operating characteristics; applied causal conclusions are deferred to separate intervention-specific analyses.

Although cumulative hazards and survival functions are in one-to-one correspondence, KCOR operates in cumulative-hazard space because curvature induced by frailty depletion is additive and more readily diagnosed there. While survival-based summaries such as restricted mean survival time may be derived from normalized hazards, KCOR's primary estimand remains cumulative by construction.

4.2 Relationship to negative control methods

Negative control outcomes/tests are widely used to *detect* confounding. KCOR's objective is different: it is an estimator intended to *normalize away a specific confounding structure*—selection-induced depletion dynamics—prior to comparison. Negative and positive controls are nevertheless central to validating the estimator's behavior.

This asymmetry helps explain why standard observational analyses often report large apparent mortality benefits during periods lacking a plausible causal mechanism: vaccinated cohorts are already selection-filtered, while unvaccinated hazards are suppressed by ongoing frailty depletion. Unadjusted comparisons therefore systematically understate unvaccinated baseline risk and exaggerate apparent benefit.

4.3 Practical guidance for use

Recommended reporting includes:

- Enrollment definition and justification
- Risk set definitions and event-time binning
- Quiet-window definition and justification
- Baseline-shape choice $g(t)$ and fit diagnostics
- Skip/stabilization rule and robustness to nearby values
- Predefined negative/positive controls used for validation
- Sensitivity analysis plan and results

KCOR should therefore be applied and reported as a complete pipeline—from cohort freezing, through depletion normalization, to cumulative comparison and diagnostics—rather than as a standalone adjustment step.

5. Limitations

- **Model dependence**: Normalization relies on the adequacy of the gamma-frailty model and the baseline-shape assumption during the quiet window.
- **θ estimation is data-driven**: KCOR does not impose $\theta = 0$ for any cohort. The frequent observation that $\hat{\theta} \approx 0$ for vaccinated cohorts is a data-driven result of the frailty fit and should not be interpreted as an assumption of homogeneity.
- **Sparse events**: When event counts are small, hazard estimation and parameter fitting can be unstable.
- **Contamination of quiet periods**: External shocks (e.g., epidemic waves) overlapping the quiet window can bias selection-parameter estimation.
- **Causal interpretation**: KCOR supports interpretable cohort comparison under stated assumptions, but it is not a substitute for randomization; causal claims require explicit causal assumptions and careful validation.
- **Non-gamma frailty**: The KCOR framework assumes that selection acts approximately multiplicatively through a time-invariant frailty distribution, for which the gamma family provides a convenient and empirically testable approximation. In settings where depletion dynamics are driven by more complex mechanisms—such as time-varying frailty variance, interacting risk factors, or shared frailty correlations within subgroups—the curvature structure exploited by KCOR may be misspecified. In such cases, KCOR diagnostics (e.g., poor curvature fit or unstable $\hat{\theta}$ estimates) serve as indicators of model inadequacy rather than targets for parameter tuning. Extending the framework to accommodate dynamic or correlated frailty structures would require explicit model generalization rather than modification of KCOR normalization steps and is left to future work. Empirically, KCOR's validity depends on curvature removal rather than the specific parametric form; alternative frailty distributions that generate similar depletion geometry would yield equivalent normalization.

5.1 Failure modes and diagnostics (recommended)

KCOR is designed to normalize selection-induced depletion curvature under its stated model and windowing assumptions. Reviewers and readers should expect the method to degrade when those assumptions are violated. Common failure modes include:

- **Mis-specified quiet window**: If the quiet window overlaps major external shocks (epidemic waves, policy changes, reporting artifacts), the fitted $(\hat{k}_d, \hat{\theta}_d)$ may absorb non-selection dynamics, biasing normalization.
- **External time-varying hazards masquerading as frailty depletion**: Strong secular trends, seasonality, or outcome-definition changes can introduce curvature that is not well captured by gamma-frailty depletion alone. For example, COVID-19 waves disproportionately increase mortality among frail individuals; if one cohort has higher baseline frailty, such a wave can preferentially deplete that cohort, producing the appearance of a benefit in the lower-frailty cohort that is actually due to differential frailty-specific mortality from the external hazard rather than from the intervention under study.
- **Extremely sparse cohorts**: When events are rare, $H_d^{\text{obs}}(t)$ becomes noisy and (k_d, θ_d) can be weakly identified, often manifesting as unstable $\hat{\theta}_d$ or wide uncertainty.
- **Non-frailty-driven curvature**: Administrative censoring, cohort-definition drift, changes in risk-set construction, or differential loss can induce curvature unrelated to latent frailty.

Practical diagnostics to increase trustworthiness include:

- **Quiet-window overlays** on hazard/cumulative-hazard plots to confirm the fit window is epidemiologically stable.
- **Fit residuals in H-space** (RMSE, residual plots) and stability of $(\hat{k}_d, \hat{\theta}_d)$ under small perturbations of the quiet-window bounds.
- **Sensitivity analyses** over plausible quiet windows and skip-weeks values.
- **Prespecified negative controls**: KCOR curves should remain near-flat at 1 under control constructions designed to induce composition differences without true effects.

In practice, prespecified negative controls—such as the age-shift controls presented in §3.1.2—provide a direct empirical check that KCOR does not generate artifactual cumulative effects under strong selection-induced curvature.

5.2 Conservativeness and edge-case detection limits

Because KCOR compares fixed enrollment cohorts, subsequent uptake of the intervention among initially unexposed individuals (or additional dosing among exposed cohorts) introduces treatment crossover over time. Such crossover attenuates between-cohort contrasts and biases KCOR(t) toward unity, making the estimator conservative with respect to detecting sustained net benefit or harm. Analyses should therefore restrict follow-up to periods before substantial crossover or stratify by dosing state when the data permit.

Because KCOR defines explicit diagnostic failure modes—instability, dose reversals, age incoherence, or absence of asymptotic convergence—the absence of such failures in the Czech 2021_24 Dose 0 versus Dose 2 cohorts provides stronger validation than goodness-of-fit alone.

Conservativeness under overlap.
When treatment effects overlap temporally with the quiet window used for frailty estimation, KCOR does not attribute the resulting curvature to treatment nor amplify it into a spurious cumulative effect. Instead, overlap manifests as degraded quiet-window fit, reduced post-normalization linearity, and instability of estimated frailty parameters, all of which are explicitly surfaced by KCOR's diagnostics. In these regimes, KCOR trajectories tend to attenuate toward unity rather than diverge, reflecting loss of identifiability rather than false detection. This behavior is illustrated in the S7 overlap variant, where treatment and selection are deliberately confounded in time: KCOR does not recover a clean effect signal, and diagnostic criteria correctly indicate that the assumptions required for interpretable normalization are violated. As a result, KCOR is conservative under temporal overlap—preferring diagnostic failure and attenuation over over-interpretation—rather than producing misleading treatment effects when separability is not supported by the data. This design choice reflects an intentional bias toward false negatives rather than false positives in ambiguous regimes. See §2.1.1 and Simulation S7 (Appendix B.6) for the corresponding identifiability assumptions and stress tests.

KCOR analyses commonly exclude an initial post-enrollment window to avoid dynamic Healthy Vaccinee Effect artifacts. If an intervention induces an acute mortality effect concentrated entirely within this skipped window, that transient signal will not be captured by the primary analysis. This limitation is addressed by reporting sensitivity analyses with reduced or zero skip-weeks and/or by separately evaluating a prespecified acute-risk window.

In degenerate scenarios where an intervention induces a purely proportional level-shift in hazard that remains constant over time and does not alter depletion-driven curvature, KCOR's curvature-based contrast may have limited ability to distinguish such effects from residual baseline level differences under minimal-data constraints. Such cases are pathological in the sense that they produce no detectable depletion signature; in practice, KCOR diagnostics and control tests help identify when curvature-based inference is not informative.

Simulation results in §3.4 illustrate that when key assumptions are violated—such as non-gamma frailty geometry, contamination of the quiet window by external shocks, or extreme event sparsity—frailty normalization may become weakly identified. In such regimes, KCOR's diagnostics, including poor cumulative-hazard fit and reduced post-normalization linearity, explicitly signal that curvature-based inference is unreliable without model generalization or revised window selection.

Importantly, increasing model complexity within the Cox regression framework—via random effects, cohort-specific frailty, or information-criterion–based selection—does not resolve this limitation, because these models continue to target instantaneous hazard ratios conditional on survival rather than cumulative counterfactual outcomes. Model-selection criteria applied within the Cox regression family favor specifications that improve likelihood fit of instantaneous hazards, but such criteria do not validate cumulative counterfactual interpretation under selection-induced non-proportional hazards.

5.3 Data requirements and external validation

External validation across interventions. A natural next step is to apply KCOR to other vaccines and interventions where large-scale individual-level event timing data are available. Many RCTs are underpowered for all-cause mortality and typically do not provide record-level timing needed for KCOR-style hazard-space normalization, while large observational studies often publish only aggregated effect estimates. Where sufficiently detailed time-to-event data exist (registries, integrated health systems, or open individual-level datasets), cross-intervention comparisons can help characterize how often selection-induced depletion dominates observed hazard curvature and how frequently post-normalization trajectories remain stable under negative controls.

6. Conclusion

KCOR provides a principled approach to retrospective cohort comparison under selection-induced hazard curvature by estimating and inverting a gamma-frailty mixture model to remove cohort-specific depletion dynamics prior to comparison. Validation via negative and positive controls supports that KCOR remains near-null under selection without effect and detects injected effects when present. Applied analyses on specific datasets are best reported separately from this methods manuscript.

Declarations (journal requirements)

Ethics approval and consent to participate

Not applicable. This is a methods-only manuscript. The primary validation results use synthetic data. Empirical negative-control figures (Figures fig. 4 and fig. 5) use aggregated cohort summaries derived from Czech Republic administrative data; no record-level data are shared in this manuscript.[4]

Consent for publication

Not applicable.

Data availability

- Synthetic validation data (negative and positive control datasets) and generation scripts are available in the project repository under `test/negative_control/` and `test/positive_control/`.
- Sensitivity analysis outputs are available under `test/sensitivity/out/`.
- The reference implementation includes example datasets in KCOR_CMR format for reproducibility.
- A formal specification of the KCOR data formats is provided in `documentation/specs/KCOR_file_format.md`, including schema definitions and disclosure-control semantics.

Code availability

- The KCOR v6 reference implementation and complete validation suite are available in the project repository.
- Repository URL: https://github.com/skirsch/KCOR
- Zenodo DOI: 10.5281/zenodo.18050329

Supplementary material

Supplementary appendices provide mathematical derivations and full control-test specifications.

Appendix A. Mathematical derivations

A.1 Frailty mixing induces hazard curvature

Consider a cohort where individual i has hazard $h_i(t) = z_i \cdot h_0(t)$, with frailty z_i drawn from a distribution with mean 1 and variance $\theta > 0$. Let $S_i(t) = \exp(-z_i H_0(t))$ be the individual survival function, where $H_0(t) = \int_0^t h_0(s) \, ds$.

The cohort survival function is the expectation over frailty:

$$S^{\text{cohort}}(t) = E_z[S_i(t)] = E_z[\exp(-z H_0(t))] = \mathscr{L}_z(H_0(t)),$$

where $\mathscr{L}_z(\cdot)$ is the Laplace transform of the frailty distribution. The cohort hazard is then:

$$h^{\text{cohort}}(t) = -\frac{d}{dt} \log S^{\text{cohort}}(t).$$

Even when $h_0(t) = k$ is constant (so $H_0(t) = kt$), the cohort hazard $h^{\text{cohort}}(t)$ is generally time-varying because high-frailty individuals die earlier, shifting the surviving population toward lower frailty over time. This is the mechanism by which frailty heterogeneity induces **curvature** in cohort-level hazards.

A.2 Gamma-frailty identity derivation

For gamma-distributed frailty $z \sim \text{Gamma}(\alpha = 1/\theta, \beta = 1/\theta)$ with mean 1 and variance θ, the Laplace transform is:

$$\mathscr{L}_z(s) = (1 + \theta s)^{-1/\theta}.$$

The cohort survival function becomes:

$$S^{\text{cohort}}(t) = \left(1 + \theta H_0(t)\right)^{-1/\theta}.$$

The observed cumulative hazard is $H^{\text{obs}}(t) = -\log S^{\text{cohort}}(t)$, giving:

$$H^{\text{obs}}(t) = \frac{1}{\theta} \log\left(1 + \theta H_0(t)\right).$$

This is the gamma-frailty identity (Equation eq. 5 in the main text).

A.3 Inversion formula

Solving for $H_0(t)$ from the gamma-frailty identity:

$$\theta H^{\text{obs}}(t) = \log\left(1 + \theta H_0(t)\right)$$

$$e^{\theta H^{\text{obs}}(t)} = 1 + \theta H_0(t)$$

$$H_0(t) = \frac{e^{\theta H^{\text{obs}}(t)} - 1}{\theta}.$$

This inversion recovers the baseline cumulative hazard from the observed cumulative hazard, conditional on the frailty variance θ.

A.3a Relationship to the Vaupel–Manton–Stallard gamma frailty framework

KCOR's normalization step is grounded in the classical demographic frailty framework (e.g., Vaupel–Manton–Stallard), in which individual hazards are multiplicatively scaled by latent frailty and cohort-level hazards decelerate due to depletion of susceptibles. Under gamma frailty, the Laplace-transform identity yields a closed-form relationship between observed cohort cumulative hazard and baseline cumulative hazard, and the inversion in §A.3 recovers the baseline cumulative hazard from observed cumulative hazards given θ.

The distinction in KCOR is not the frailty identity itself, but the **direction of inference** and the **estimand**. Frailty-augmented Cox and related regression approaches embed gamma frailty within a regression model to estimate covariate effects (hazard ratios). KCOR instead uses quiet-window curvature to estimate cohort-specific frailty parameters and then inverts the frailty identity to obtain depletion-neutralized baseline cumulative hazards, defining KCOR as a ratio of these cumulative quantities. Thus, KCOR solves an inverse normalization problem and targets cumulative comparisons under selection-induced non-proportional hazards rather than instantaneous hazard-ratio regression parameters.

A.4 Variance propagation (sketch)

For uncertainty quantification, variance in $\mathrm{KCOR}(t) = \tilde{H}_{0,A}(t)/\tilde{H}_{0,B}(t)$ can be approximated via the delta method. If $\mathrm{Var}(\tilde{H}_{0,d})$ is available (e.g., from bootstrap or analytic propagation through the inversion), then:

$$\mathrm{Var}(\mathrm{KCOR}) \approx \mathrm{KCOR}^2 \left[\frac{\mathrm{Var}(\tilde{H}_{0,A})}{\tilde{H}_{0,A}^2} + \frac{\mathrm{Var}(\tilde{H}_{0,B})}{\tilde{H}_{0,B}^2} - 2\frac{\mathrm{Cov}(\tilde{H}_{0,A}, \tilde{H}_{0,B})}{\tilde{H}_{0,A}\tilde{H}_{0,B}} \right].$$

In practice, Monte Carlo resampling provides a more robust approach that captures uncertainty from both event realization and parameter estimation.

Appendix B. Control-test specifications

B.1 Negative control: synthetic gamma-frailty null

The synthetic negative control (Figure fig. 3) is generated using:

- **Data source**: `example/Frail_cohort_mix.xlsx` (pathological frailty mixture)
- **Generation script**: `code/generate_pathological_neg_control_figs.py`
- **Cohort A weights**: Equal weights across 5 frailty groups (0.2 each)
- **Cohort B weights**: Shifted weights [0.30, 0.20, 0.20, 0.20, 0.10]
- **Frailty values**: [1, 2, 4, 6, 10] (relative frailty multipliers)
- **Base weekly probability**: 0.01
- **Weekly log-slope**: 0.0 (constant baseline during quiet periods)
- **Skip weeks**: 2
- **Normalization weeks**: 4
- **Time horizon**: 250 weeks

Both cohorts share identical per-frailty-group death probabilities; only the mixture weights differ. This induces different cohort-level curvature under the null.

B.2 Negative control: empirical age-shift construction

The empirical negative control (Figures fig. 4 and fig. 5) is generated using:

- **Data source**: Czech Republic administrative mortality and vaccination data, aggregated into KCOR_CMR format
- **Generation script**: `test/negative_control/code/generate_negative_control.py`
- **Construction**: Age strata remapped to pseudo-doses within same vaccination category
- **Age mapping**:
    - Dose 0 → YoB {1930, 1935}
    - Dose 1 → YoB {1940, 1945}

– Dose 2 → YoB {1950, 1955}
- **Output YoB**: Fixed at 1950 (unvax cohort) or 1940 (vax cohort)
- **Sheets processed**: 2021_24, 2022_06

This construction ensures that dose comparisons are within the same underlying vaccination category, preserving a true null while inducing 10–20 year age differences.

B.3 Positive control: injected effect

The positive control (Figure fig. 6 and Table tbl. 9) is generated using:

- **Generation script**: `test/positive_control/code/generate_positive_control.py`
- **Initial cohort size**: 100,000 per cohort
- **Baseline hazard**: 0.002 per week
- **Frailty variance**: $\theta_0 = 0.5$ (control), $\theta_1 = 1.0$ (treatment)
- **Effect window**: weeks 20–80
- **Hazard multipliers**:
  – Harm scenario: $r = 1.2$
  – Benefit scenario: $r = 0.8$
- **Random seed**: 42
- **Enrollment date**: 2021-06-14 (ISO week 2021_24)

The injection multiplies the treatment cohort's baseline hazard by factor r during the effect window, while leaving the control cohort unchanged.

B.4 Sensitivity analysis parameters

The sensitivity analysis (Figure fig. 7) varies:

- **Baseline weeks**: [2, 3, 4, 5, 6, 7, 8]
- **Quiet-start offsets**: [-12, -8, -4, 0, +4, +8, +12] weeks from 2022-24
- **Quiet-window end**: Fixed at 2024-16
- **Dose pairs**: 1 vs 0, 2 vs 0, 2 vs 1
- **Cohorts**: 2021_24

Output grids show KCOR values for each parameter combination.

B.5 Tail-sampling / bimodal selection (adversarial selection geometry)

We generate a base frailty population distribution with mean 1. Cohort construction differs by selection rule:

- **Mid-sampled cohort**: frailty restricted to central quantiles (e.g., 25th–75th percentile) and renormalized to mean 1.
- **Tail-sampled cohort**: mixture of low and high tails (e.g., 0–15th and 85th–100th percentiles) with mixture weights chosen to yield mean 1.

Both cohorts share the same baseline hazard $h_0(t)$ and no treatment effect (negative-control version). We also generate positive-control versions by applying a known hazard multiplier in a prespecified window. We evaluate (i) KCOR drift, (ii) quiet-window fit RMSE, (iii) post-normalization linearity, and (iv) parameter stability under window perturbation.

- **Generation script**: `test/sim_grid/code/generate_tail_sampling_sim.py`
- **Base frailty distribution**: Log-normal with mean 1, variance 0.5
- **Mid-quantile cohort**: 25th–75th percentile
- **Tail-mixture cohort**: [0–15th] + [85th–100th] percentiles, equal weights
- **Baseline hazard**: 0.002 per week (constant)
- **Positive-control hazard multiplier**: $r = 1.2$ (harm) or $r = 0.8$ (benefit)
- **Effect window**: weeks 20–80
- **Random seed**: 42

B.6 Joint frailty and treatment-effect simulation (S7)

This simulation evaluates KCOR under conditions in which **both selection-induced depletion (frailty heterogeneity)** and a **true treatment effect (harm or benefit)** are present simultaneously. The purpose is to assess whether KCOR can (i) correctly identify and neutralize frailty-driven curvature using a quiet period and (ii) detect a true treatment effect outside that period without confounding the two mechanisms.

Design

Two fixed cohorts are generated with identical baseline hazards but differing frailty variance. Individual hazards are multiplicatively scaled by a latent frailty term drawn from a gamma distribution with unit mean and cohort-specific variance. A treatment effect is then injected over a prespecified time window that does not overlap the quiet period used for frailty estimation.

Formally, individual hazards are generated as

$$h_i(t) = z_i \cdot h_0(t) \cdot r(t),$$

where z_i is individual frailty, $h_0(t)$ is a shared baseline hazard, and $r(t)$ is a time-localized multiplicative treatment effect applied to one cohort only.

Frailty structure

- Cohort 0: $z \sim \text{Gamma}(\theta_0)$
- Cohort 1: $z \sim \text{Gamma}(\theta_1)$, with $\theta_1 \neq \theta_0$

Frailty distributions are normalized to unit mean, differing only in variance, thereby inducing different depletion dynamics and cumulative-hazard curvature across cohorts in the absence of any treatment effect.

Treatment effect

A known treatment effect is applied to Cohort 1 during a finite window $[t_{\text{on}}, t_{\text{off}}]$. Three effect shapes are considered:

1. Step change (constant multiplicative factor),
2. Linear ramp,
3. Smooth pulse ("bump").

Both harmful ($r(t) > 1$) and protective ($r(t) < 1$) effects are evaluated. The treatment window is chosen to lie strictly outside the quiet period used for frailty estimation.

Quiet period and estimation

Frailty parameters are estimated independently for each cohort using observed cumulative hazards over a prespecified quiet window $[t_q^{\text{start}}, t_q^{\text{end}}]$ during which $r(t) = 1$ by construction. KCOR normalization is then applied to the full time horizon using these estimated parameters.

This design enforces **temporal separability** between selection-induced depletion and treatment effects.

Evaluation criteria

The simulation is considered successful if:

1. KCOR remains approximately flat and near unity during the quiet window,
2. KCOR deviates in the correct direction and magnitude during the treatment window,
3. Fit diagnostics (e.g., residual curvature, post-normalization linearity) remain stable outside intentionally violated scenarios.

An additional stress-test variant intentionally overlaps the treatment window with the quiet period. In this case, KCOR diagnostics degrade and normalized trajectories fail to stabilize, correctly signaling violation of the identifiability assumptions rather than producing spurious treatment effects.

Interpretation

This simulation demonstrates that when selection-induced depletion and treatment effects are temporally separable, KCOR can disentangle the two mechanisms: frailty parameters are identified from quiet-period curvature, and true treatment effects manifest as deviations from unity outside that window. When separability is violated, KCOR does not silently misattribute effects; instead, diagnostics flag reduced interpretability.

Appendix C. Additional figures and diagnostics

C.1 Fit diagnostics

For each cohort d, the gamma-frailty fit produces diagnostic outputs including:

- **RMSE in H-space**: Root mean squared error between observed and model-predicted cumulative hazards over the quiet window. Values < 0.01 indicate excellent fit; values > 0.05 may warrant investigation.
- **Fitted parameters**: $\hat{k}_d$ (baseline hazard level) and $\hat{\theta}_d$ (frailty variance). Very small $\hat{\theta}_d$ (< 0.01) indicates minimal detected depletion; very large values (> 5) may indicate model stress.
- **Number of fit points**: n_{obs} observations in quiet window. Larger n_{obs} provides more stable estimates.

Example diagnostic output from the reference implementation:

```
KCOR6_FIT,EnrollmentDate=2021_24,YoB=1950,Dose=0,
  k_hat=4.29e-03,theta_hat=8.02e-01,
  RMSE_Hobs=3.37e-03,n_obs=97,success=1
```

C.2 Residual analysis

Fit residuals $r_t = H_d^{\text{obs}}(t) - H_d^{\text{model}}(t; \hat{k}_d, \hat{\theta}_d)$ should be examined for:

- **Systematic patterns**: Residuals should be approximately random around zero. Systematic curvature in residuals suggests model inadequacy.
- **Outliers**: Individual weeks with large residuals may indicate data quality issues or external shocks.
- **Autocorrelation**: Strong autocorrelation in residuals suggests the model is missing time-varying structure.

C.3 Parameter stability checks

Robustness of $(\hat{k}_d, \hat{\theta}_d)$ should be assessed by:

- **Quiet-window perturbation**: Shift the quiet-window start/end by ±4 weeks and re-fit. Stable parameters should vary by < 10%.
- **Skip-weeks sensitivity**: Vary SKIP_WEEKS from 0 to 8 and verify KCOR trajectories remain qualitatively similar.
- **Baseline-shape alternatives**: Compare default $g(t) = 1$ to mild linear trends and verify normalization is not sensitive to this choice.

The sensitivity analysis (§3.3 and Figure fig. 7) provides a systematic assessment of parameter stability.

C.4 Quiet-window overlay plots

Recommended diagnostic: overlay the prespecified quiet window on hazard and cumulative-hazard time series plots. The fit window should:

- Avoid major epidemic waves or external mortality shocks
- Contain sufficient event counts for stable estimation
- Span a time range where baseline mortality is approximately stationary

Visual inspection of quiet-window placement relative to mortality dynamics is an essential diagnostic step.

References

1. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979;16(3):439-454. doi:10.2307/2061224

2. Obel N, Fox M, Tetens M, et al. Confounding and Negative Control Methods in Observational Study of SARS-CoV-2 Vaccine Effectiveness: A Nationwide, Population-Based Danish Health Registry Study. *Clinical Epidemiology*. 2024;Volume 16:501-512. doi:10.2147/CLEP.S468572

3. Chemaitelly H, Ayoub HH, Coyle P, et al. Assessing healthy vaccinee effect in COVID-19 vaccine effectiveness studies: A national cohort study in Qatar. Schiffer JT, Henry D, eds. *eLife*. 2025;14:e103690. doi:10.7554/eLife.103690

4. Šanca O, Jarkovský J, Klimeš D, et al. Vaccination, positivity, hospitalization for COVID-19, deaths, long covid and comorbidities in people in the Czech Republic. *National Health Information Portal*. Published online 2024.