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

Manuscript metadata

- **Article type**: Methods / Statistical method
- **Running title**: KCOR via gamma-frailty normalization
- **Author**: Steven T. Kirsch
- **Affiliations**: Independent Researcher, United States
- **Corresponding author**: stk@alum.mit.edu
- **Word count**: 12,100 (excluding Abstract, References, and Supplementary material)
- **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 and transforms observed cumulative hazards into depletion-neutralized baseline cumulative hazards. Cohorts are then compared using a **cumulative hazard ratio**, which is the estimand reported by the method. Because selection and treatment may be confounded in observational data, KCOR is presented as a normalization-and-comparison framework rather than a general causal estimator; interpretability is evaluated using prespecified diagnostics, negative controls, and simulations. KCOR requires only event timing—dates of enrollment and death, with optional age stratification—making it applicable to minimal record-level mortality datasets. Simulation studies confirm that KCOR remains near-null under selection-only regimes and reliably detects injected cumulative effects when present.

Key messages

• Selection-induced depletion under latent frailty heterogeneity precludes reliable use of proportional hazards–based methods in many retrospective cohort studies.

• KCOR enables depletion-neutralized cumulative outcome comparisons using only minimal event-time information (e.g., dates of birth, intervention, and death), without requiring covariate completeness or proportional hazards assumptions.

• Simulation studies show that KCOR remains stable under selection-only regimes, while standard Cox regression can yield strongly non-null hazard ratios despite a true null effect.

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.

The methodological problem addressed here is general. The COVID-19 period provides a natural empirical regime characterized by strong selection heterogeneity and non-proportional hazards, serving as a useful illustration for the proposed framework. 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.

1.3.1 Related work beyond proportional hazards: time-varying effects and flexible hazard modeling

A large literature relaxes proportional hazards by allowing either (i) covariate effects to vary over time or (ii) the baseline hazard to be modeled flexibly. Representative examples include time-varying coefficient Cox models, flexible parametric survival models (e.g., spline-based log-cumulative hazard models), additive hazards models (Aalen-type formulations), landmarking and dynamic prediction approaches, and weighting-based approaches such as marginal structural models (MSMs) for time-varying confounding.[4–11] These methods are most effective when the scientific target is an instantaneous contrast (e.g., a time-specific hazard ratio) and when sufficiently rich covariate histories are available to support identification. In contrast, KCOR targets cumulative contrasts under explicit depletion normalization in cumulative-hazard space, rather than instantaneous hazard ratios conditional on survival.

KCOR is motivated by a different failure mode that remains even when proportional hazards is relaxed: cohort-level *curvature* in cumulative-hazard space arising from selection-induced frailty heterogeneity and depletion of susceptibles. Allowing hazard ratios to vary over time can improve descriptive fit, but it does not by itself normalize depletion geometry or make cohorts exchangeable. Similarly, flexible baselines can absorb curvature without distinguishing whether it arises from latent selection/depletion versus genuine time-varying external hazards. KCOR therefore occupies a distinct role: it uses a mixture-model identity in cumulative-hazard space to estimate and invert selection-induced depletion dynamics during prespecified quiet periods, and then defines a cumulative comparison operator (KCOR) on the depletion-neutralized scale.

In settings where rich covariate histories and credible causal designs (e.g., target trial emulation, MSMs with well-specified treatment models, instrumental variables, or regression discontinuity) are feasible, those approaches should be preferred. KCOR is designed for a complementary setting: minimal-data retrospective registries where only event timing is available and where selection-induced depletion dominates non-proportionality. In such settings, KCOR provides a normalization-and-diagnostic framework intended to make depletion-driven non-proportionality explicit and testable rather than implicitly absorbed by flexible hazard models.

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 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)

Table 3: Positioning KCOR among retrospective methods.

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**

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.

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 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 after an optional stabilization skip period:

$$H_{\text{obs},d}(t).$$

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:

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

Under the reference constant-baseline specification over the quiet window:

$$H_{0,d}(t) = k_d\, t.$$

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:

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

5. **KCOR ratio**: Cohorts are compared via the ratio of depletion-neutralized cumulative hazards:

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

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 4 defines the notation used throughout the Methods section.

Table 4: Notation used throughout the Methods section.

Symbol	Definition
d	Cohort index
t	Event time since enrollment (discrete bins, e.g., weeks)
$h_{\mathrm{obs},d}(t)$	Observed cohort hazard
$H_{\mathrm{obs},d}(t)$	Observed cumulative hazard (after skip/stabilization)
$h_{0,d}(t)$	Baseline hazard for cohort d under the depletion-neutralized model
$H_{0,d}(t)$	Baseline cumulative hazard for cohort d under the depletion-neutralized model
$\tilde{H}_{0,d}(t)$	Depletion-neutralized baseline cumulative hazard
θ_d	Frailty variance (selection strength) for cohort d; governs curvature in the observed cumulative hazard
$\hat{\theta}_d$	Estimated frailty variance from quiet-window fitting
k_d	Baseline hazard level for cohort d under the default baseline shape
$\hat{k}_d$	Estimated baseline hazard level from quiet-window fitting

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 2 provides a schematic overview of the KCOR workflow.

2.1.1 KCOR assumptions and identifiability

KCOR relies on **five explicit assumptions (A1–A5)**. These assumptions are stated exhaustively here; **no additional assumptions are introduced elsewhere in the manuscript**. Each assumption is either enforced by study design or empirically testable via prespecified diagnostics. When an assumption is violated, KCOR does not silently produce misleading estimates; instead, the violation manifests through identifiable diagnostic failures described in Appendix D.

A1. Fixed cohorts at enrollment. Individuals are assigned to cohorts at enrollment and remain in those cohorts for the duration of follow-up. No post-enrollment switching or censoring is permitted in the primary estimand.

A2. Shared external hazard environment. All cohorts are exposed to the same calendar-time external mortality environment (e.g., seasonality, epidemic waves, reporting artifacts). External shocks may occur but must affect cohorts symmetrically.

A3. Selection operates through time-invariant latent frailty. Differences between cohorts at enrollment arise primarily through selection on latent, multiplicative frailty that remains constant over follow-up and induces depletion of susceptibles.

A4. Gamma frailty adequately approximates depletion geometry. Gamma frailty provides a sufficient approximation to the cumulative-hazard curvature induced by frailty-driven depletion over the estimation window. This assumption is not asserted a priori and is evaluated empirically.

A5. Existence of a valid quiet window for frailty identification. There exists a prespecified epidemiologically quiet period *(i.e., periods with approximately stable baseline hazard and no major external mortality shocks, rather than simply periods of low mortality)* during which selection-induced depletion dominates other sources of curvature, permitting identification of frailty parameters. This assumption is assessed empirically and is the primary dataset-specific requirement for KCOR applicability.

Of these five assumptions, **A1–A3 are structural and are typically satisfied by cohort construction in national mortality datasets**, while **A4–A5 are empirically evaluated using diagnostics**. Failure of any assumption limits interpretability and is explicitly signaled by KCOR's diagnostic outputs (Appendix D). Sanity checks on frailty estimates—including magnitude, stability, and coherence with observed cumulative-hazard curvature—are treated diagnostically rather than as assumptions (Appendix D; see also diagnostic coherence in §2.1.2). KCOR does not assume causal separability of selection and treatment effects; rather, it identifies cumulative outcome differences under explicit normalization of selection-induced depletion, making departures from the null of no net harm or benefit interpretable when diagnostic and interpretability conditions are satisfied.

KCOR is designed for settings in which selection-induced depletion of susceptibles is the dominant source of non-proportionality in cohort hazards; regimes in which timing misalignment or denominator redefinition dominate are typically identifiable through early instability, lack of a valid quiet window, or failure of post-normalization linearity and therefore fall outside the intended scope of the method.

2.1.2 Interpretability checklist for KCOR results

The assumptions in §2.1.1 specify the conditions under which KCOR is a valid and diagnostically enforceable estimator. Interpretation of a specific KCOR result—particularly as evidence of harm or benefit—requires additional prespecified checks. These checks do not constitute assumptions of the method, but determine what inferential question a given KCOR trajectory answers.

Before interpreting a KCOR result as evidence of cumulative harm or benefit, the following interpretability checks should be verified:

1. **Dynamic selection handling.** Early post-enrollment periods subject to short-horizon dynamic selection (e.g., deferral effects) are excluded from frailty identification via prespecified skip weeks.

2. **Quiet baseline anchoring.** The baseline anchoring period used for comparison lies within an epidemiologically quiet window, free of major external shocks, and exhibits approximate post-normalization linearity.

3. **Temporal alignment with hypothesized effects.** The follow-up window used for interpretation overlaps the period during which a substantive effect is hypothesized to occur; KCOR does not recover effects outside the analyzed window.

4. **Post-normalization stability.** KCOR(t) trajectories stabilize rather than drift following normalization and anchoring, consistent with successful removal of selection-induced depletion curvature.

5. **Diagnostic coherence.** Fitted frailty parameters and residual diagnostics are stable under reasonable perturbations of skip weeks and quiet-window boundaries.

Failure of any interpretability check narrows what can be inferred, but does not invalidate the KCOR estimator itself. Formal diagnostics underlying these checks, including frailty-parameter coherence and stability, are detailed in Appendix D.

2.1.2a Multiplicity and ensemble interpretation

KCOR is not intended as a mass univariate hypothesis-testing framework. When applied across multiple age strata, enrollment periods, or dose comparisons, resulting KCOR(t) trajectories should be interpreted as correlated views of a shared underlying process rather than as independent statistical tests. Accordingly, inference emphasizes diagnostic

coherence, temporal stability, and consistency of cumulative-outcome patterns across strata, rather than reliance on nominal significance thresholds or formal multiple-comparison corrections.

2.1.3 Identifiability and scope of inference

Identifiability in the KCOR framework concerns whether cumulative outcome differences can be meaningfully interpreted after explicit normalization of selection-induced depletion. KCOR is not a general causal estimator under arbitrary unmeasured confounding; it is a depletion-normalization method that yields a cumulative cohort contrast that is interpretable only conditional on the stated assumptions, controls, and diagnostics. KCOR does not seek to decompose selection and treatment effects into separate causal components; instead, it tests whether observed cumulative outcomes depart from the null hypothesis of no net harm or benefit under shared external hazards and diagnosed frailty structure. When diagnostic and interpretability conditions are satisfied, persistent departures of KCOR(t) from unity reflect differences in cumulative outcomes that cannot be attributed to selection alone. When those conditions are not met, KCOR explicitly signals loss of identifiability rather than producing misleading estimates.

2.1.4 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 causal terminology, latent frailty heterogeneity functions as an unmeasured confounder; however, its defining feature in this setting is not baseline imbalance per se but the induced time-evolution of the risk set, which violates proportional hazards even under a true null effect.

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.

2.1.5 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:

$$\hat{\theta}_d \approx 0$$

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

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

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

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

2.3.1 Discrete time and tied events

KCOR is formulated in discrete time using aggregated event counts and therefore naturally accommodates tied event times. For population registry data, weekly or similarly coarse time bins are generally appropriate and reduce noise without sacrificing identifiability. Finer binning does not improve depletion normalization and may destabilize frailty estimation when event counts per bin are small. Bin width should be guided by diagnostic stability rather than temporal resolution alone.

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:

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

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 observed cohort hazards 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:

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

This identity can be inverted exactly:

$$H_{0,d}(t) = \frac{\exp\left(\theta_d\, H_{\text{obs},d}(t)\right) - 1}{\theta_d}. \tag{6}$$

This inversion is the **KCOR normalization step**: it transforms the observed cumulative hazard into a depletion-neutralized baseline cumulative hazard for each cohort. Figure 1 illustrates this logic schematically.
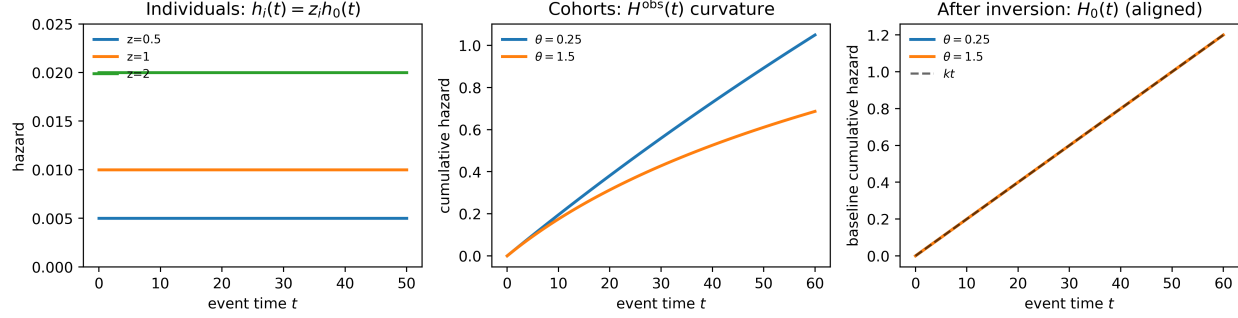

Figure 1: Three-panel schematic illustrating the KCOR normalization logic. Left: individual hazards differ only by multiplicative frailty z, with no treatment effect. Middle: aggregation over heterogeneous frailty induces cohort-level curvature in observed cumulative hazards despite identical baseline hazards. Right: inversion of the gamma-frailty identity recovers aligned baseline cumulative hazards, demonstrating depletion-neutralization. This figure is schematic and intended for conceptual illustration; it does not represent empirical data.

Observed cumulative hazard:

$$H_{\text{obs},d}(t)$$

Baseline cumulative hazard:

$$H_{0,d}(t)$$

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:

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

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 $t \geq \text{SKIP\_WEEKS}$.

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

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

Identifiability of (k_d, θ_d) comes from curvature in cumulative-hazard space: observed cumulative hazards are nonlinear in event time 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 $H_d^{\text{model}}(t; k_d, \theta_d) \to k_d t$ as $\theta_d \to 0$. Operationally, near-linear observed cumulative hazards naturally drive the fitted frailty variance toward zero; 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 fitted frailty variance estimates approaching zero, 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, 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, fitted frailty variance estimates naturally shrink toward zero, signaling limited identifiability rather than inducing spurious correction.

Parameters are estimated by constrained nonlinear least squares:

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

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 observed cumulative hazards 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 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 $t < \text{SKIP\_WEEKS}$.
- **Observed hazard transform used to build observed cumulative hazards from weekly mortality risk**: if $\text{MR}_{d,t} = D_{d,t}/Y_{d,t}$ is the weekly mortality input,

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

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:

$$\tilde{H}_{0,d}(t) = \frac{\exp\left(\hat{\theta}_d\, H_{\text{obs},d}(t)\right) - 1}{\hat{\theta}_d}. \tag{11}$$

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 measurable depletion undergo an 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 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.

In KCOR, the parametric model is used solely to estimate and invert selection-induced curvature in cumulative-hazard space; treatment comparisons are then made directly on the adjusted data. In contrast, Cox regression fits a hazard model to the observed data and derives treatment effects from model coefficients, implicitly attributing all systematic hazard divergence—including selection effects—to the exposure.

2.6.1 Computational considerations

KCOR operates on aggregated event counts in discrete time and cumulative-hazard space. Computational complexity scales linearly with the number of time bins and strata rather than the number of individuals, making the method feasible for very large population registries. In practice, KCOR analyses on national-scale datasets (millions of individuals) are memory-bound rather than CPU-bound and can be implemented efficiently using standard vectorized numerical libraries. No iterative optimization over individual-level records is required.

2.6.2 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 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 over the fit set $\mathcal{T}_d$:

$$r_d(t) = H_{\mathrm{obs},d}(t) - H_d^{\mathrm{model}}(t; \hat{k}_d, \hat{\theta}_d).$$

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 fitted frailty variance under small boundary shifts signal that the fitted curvature is sensitive to transient dynamics rather than stable depletion.

4. **Non-identifiability manifests as:**

$$\hat{\theta}_d \to 0.$$

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(t) 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:

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

Then compute observed cumulative hazards from $h_d^{\mathrm{eff}}(t)$ as in §2.3:

$$H_{\mathrm{obs},d}(t).$$

2.8 KCOR estimator

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

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

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 5 summarizes the complete KCOR pipeline.

Table 5: Step-by-step KCOR 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 (observed hazards)	Hazard curves	Yes (binning/transform)	Check for zeros/sparsity
3	Apply stabilization skip and accumulate observed cumulative hazards	Observed cumulative hazards	Yes (skip rule)	Plot observed cumulative hazards
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	Fitted parameters	Yes	RMSE, residuals, fit stability
6	Normalize: invert gamma-frailty identity to depletion-neutralized cumulative hazards	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

Step	Operation	Output	Prespecify?	Diagnostics
8	Uncertainty	CI / intervals	Yes	Coverage on positive controls

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 $\hat{H}_0(t)$
- Depletion effects neutralized
- Approximately linearized hazard scale
- Directly comparable across cohorts

KCOR(t) = $\hat{H}_1(t) / \hat{H}_0(t)$
- 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.

2.11 Relationship to Cox proportional hazards

Cox proportional hazards models estimate an instantaneous hazard ratio under the assumption that hazards differ by a time-invariant multiplicative factor. Under selective uptake with latent frailty heterogeneity, this assumption is typically violated, yielding time-varying hazard ratios induced purely by depletion dynamics. Accordingly, Cox results are presented here only as a diagnostic demonstration of assumption failure, not as a competing causal estimator.

Conceptually, Cox regression estimates a treatment effect by fitting a hazard model to the observed data and reading the effect from the fitted coefficients, whereas KCOR uses a parametric model only to adjust the observed data for selection-induced depletion and then computes cohort contrasts directly from the adjusted cumulative hazards themselves.

2.11.1 Demonstration: Cox bias under frailty heterogeneity with no treatment effect

We conducted a controlled synthetic experiment in which the **true causal effect is known to be zero by construction**, isolating latent frailty heterogeneity as the sole driver of depletion-induced non-proportional hazards. Cox and KCOR were applied to the same simulated datasets under identical information constraints.

Data-generating process.

Two cohorts of equal size were simulated under the same baseline hazard $h_0(t)$ over time (constant or Gompertz). Individual hazards were generated as $z\,h_0(t)$, with frailty

$$z \sim \mathrm{Gamma}(\theta^{-1}, \theta^{-1}),$$

with mean 1 and variance θ.

Cohort A was generated with $\theta = 0$ (no frailty heterogeneity), while Cohort B was generated with $\theta > 0$. **No treatment or intervention effect was applied**: conditional on frailty, the two cohorts have identical hazards at all times. Thus, the true causal hazard ratio between cohorts is exactly 1 for all t.

Simulations were repeated over a grid of frailty variances $\theta \in \{0, 0.5, 1, 2, 5, 10, 20\}$.

Cox analysis.

For each simulated dataset, we fitted a standard Cox proportional hazards model using partial likelihood (statsmodels `PHReg`), with cohort membership as the sole covariate (no time-varying covariates or interactions). The resulting hazard ratio estimates and confidence intervals therefore reflect **only differences induced by frailty-driven depletion**, not any causal effect.

KCOR analysis.

The same simulated datasets were analyzed using KCOR. For the synthetic datasets, cohort-specific observed cumulative hazards were estimated nonparametrically using the Nelson–Aalen estimator, then mapped to depletion-neutralized baseline cumulative hazards via the gamma-frailty inversion prior to computing $\mathrm{KCOR}(t)$. Although the data-generating process specifies individual hazards, cumulative hazards were estimated from simulated event-time data using Nelson–Aalen to mirror the information available in observational registry studies, rather than exploiting simulator-only knowledge. Frailty parameters were estimated during a prespecified quiet window, followed by cumulative-hazard normalization and computation of $\mathrm{KCOR}(t)$. Post-normalization slope and asymptotic KCOR(t) values were examined to assess departure from the null.

Expected behavior under the null.

Because the data-generating process includes **no treatment effect**, any valid estimator should return a null result. In this setting:

- **Cox regression** is expected to produce apparent non-null hazard ratios as θ increases, reflecting differential depletion of susceptibles and violation of proportional hazards induced by frailty heterogeneity.
- **KCOR** is expected to remain centered near unity with negligible post-normalization slope across all θ, consistent with correct null behavior after depletion normalization.

Summary of findings.

Across increasing values of θ, Cox regression produced progressively larger apparent deviations from a hazard ratio of 1. The direction and magnitude of the apparent effect depended on the follow-up horizon and degree of frailty heterogeneity. In contrast, KCOR(t) trajectories remained stable and centered near unity, with post-normalization slopes approximately zero across all simulated conditions.

These results demonstrate that **frailty heterogeneity alone is sufficient to induce spurious hazard ratios in Cox regression**, while KCOR correctly returns a null result under the same conditions.

Table 6: Cox vs KCOR under a synthetic null with increasing frailty heterogeneity. Two cohorts are simulated with identical baseline hazards and no treatment effect *(null by construction)*; cohorts differ only in gamma frailty variance (θ). Despite the true hazard ratio being 1 by construction, Cox regression produces increasingly non-null hazard ratios as θ increases, reflecting depletion-induced non-proportional hazards. KCOR(t) remains centered near unity with negligible post-normalization slope across θ values. (Exact values depend on simulation seed and follow-up horizon.)

Frailty variance (θ)	Cox HR	95% CI	Cox p-value	KCOR asymptote	KCOR post-norm slope
0.0	0.988	[0.969, 1.008]	0.234	0.988	7.6×10^{-4}
0.5	0.965	[0.946, 0.985]	4.9×10^{-4}	0.990	-3.8×10^{-5}
1.0	0.944	[0.926, 0.963]	1.7×10^{-8}	0.992	-3.0×10^{-4}
2.0	0.902	[0.884, 0.921]	2.4×10^{-23}	0.991	3.7×10^{-4}
5.0	0.804	[0.787, 0.820]	1.5×10^{-93}	0.993	-5.3×10^{-4}
10.0	0.701	[0.686, 0.717]	$< 10^{-200}$	1.020	3.2×10^{-4}
20.0	0.551	[0.539, 0.564]	$< 10^{-300}$	1.024	-1.6×10^{-4}

Interpretation.

This demonstration shows that Cox proportional hazards regression can report highly statistically significant non-null hazard ratios—even when the true treatment effect is identically zero—solely due to frailty-induced depletion (e.g., $p < 10^{-300}$ at $\theta = 20$), with the magnitude and direction of the apparent effect depending on follow-up horizon, not any causal signal. KCOR, by explicitly normalizing depletion geometry in cumulative-hazard space, correctly returns a null result under the same conditions, with KCOR(t) remaining centered near unity with negligible post-normalization slope across all frailty variance values. This controlled example motivates the use of KCOR for retrospective vaccine studies, where frailty heterogeneity and non-proportional hazards are expected to be substantial.

This synthetic null isolates depletion-driven non-proportionality as a sufficient cause of strongly non-null Cox estimates, motivating cumulative-hazard–based normalization when selection-induced curvature is present.

2.11.2 Relation to other approaches addressing selection effects

Several existing approaches, including inverse-probability weighting, marginal structural models, and joint frailty formulations, aim to mitigate selection bias in observational survival analyses. These methods rely on explicit modeling of treatment assignment or hazard structure and generally operate in instantaneous-hazard space. KCOR differs in focusing directly on depletion geometry in cumulative-hazard space and enforcing diagnostic checks that signal loss of identifiability when normalization fails. A full head-to-head empirical comparison across all model families is beyond scope of this methods manuscript; however, we explicitly contrast estimands, assumptions, and failure modes in §1.5 and we provide controlled demonstrations (including a synthetic null) showing that Cox-type estimators can return strongly non-null hazard ratios under frailty-driven depletion with no treatment effect, whereas
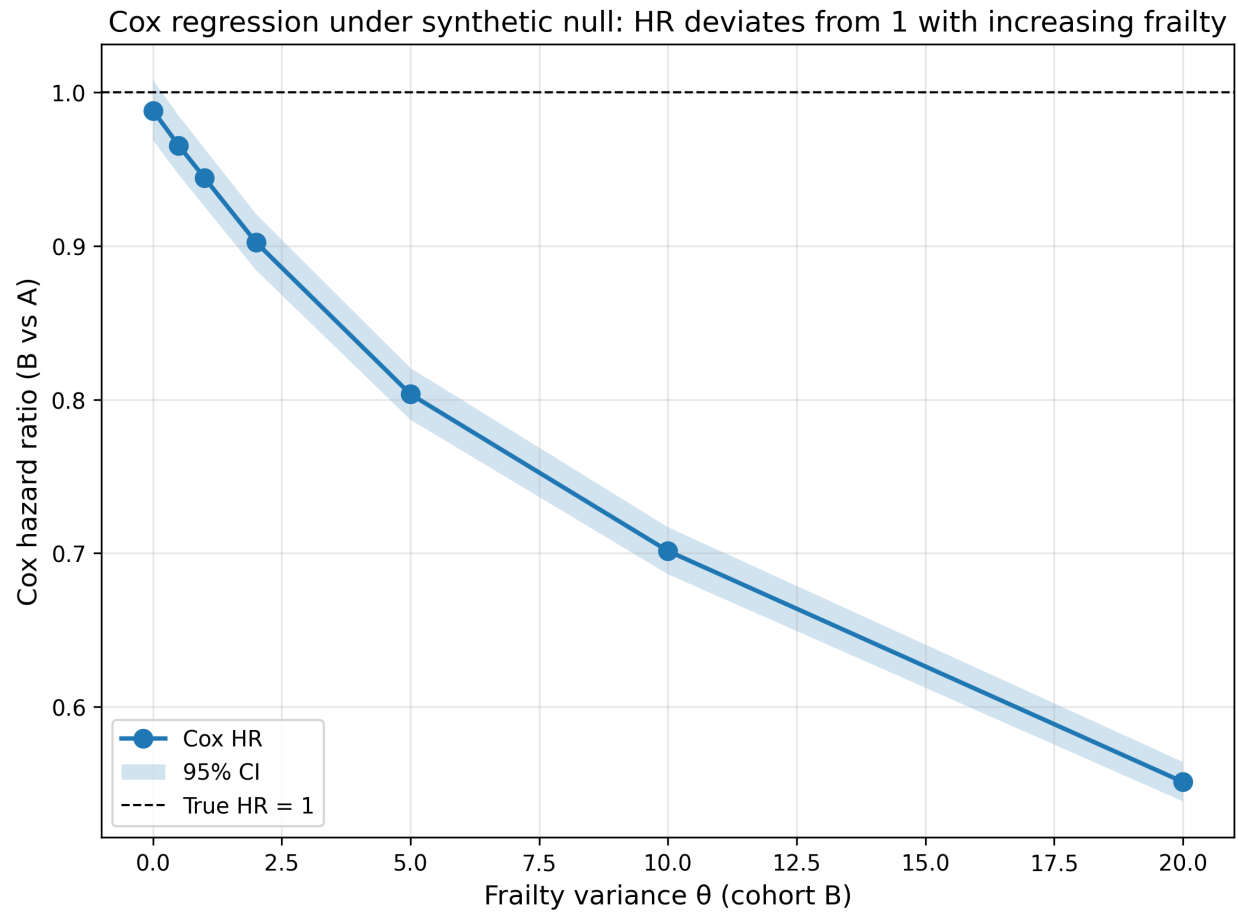
Figure 3: Cox regression produces spurious non-null hazard ratios under a *synthetic null* as frailty heterogeneity increases. Hazard ratios (with 95% confidence intervals) from Cox proportional hazards regression comparing cohort B to cohort A in simulations where the true treatment effect is identically zero and cohorts differ only in frailty variance (θ). Deviations from HR=1 arise solely from frailty-driven depletion and associated non-proportional hazards.
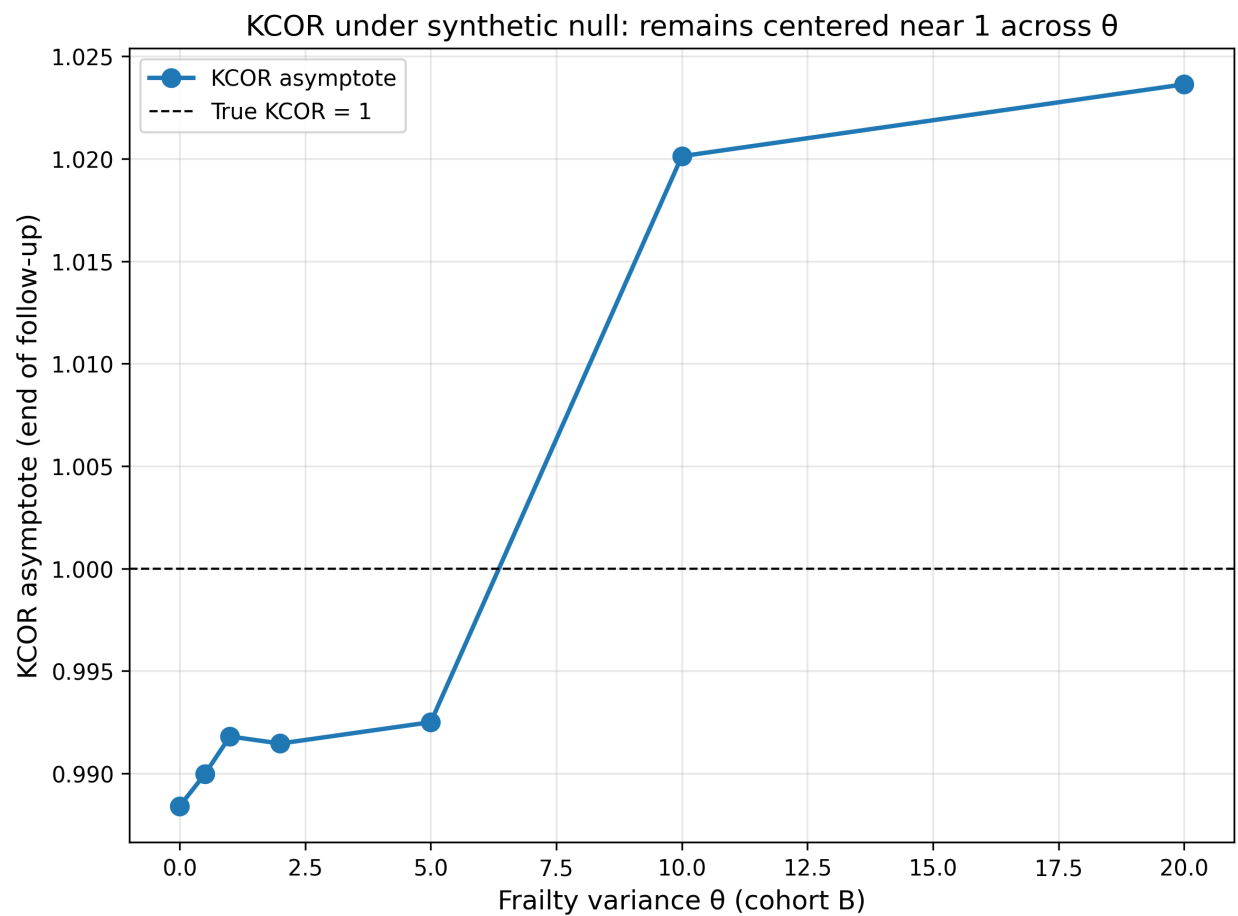
Figure 4: KCOR(t) remains null under a synthetic null across increasing frailty heterogeneity. KCOR(t) asymptotes remain near 1 across θ in the same simulations, consistent with correct null behavior after depletion normalization.

KCOR remains near-null under the same conditions. This is the specific geometry KCOR targets: depletion-induced non-proportionality in cumulative-hazard space.

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.

Throughout, curvature in cumulative hazard plots reflects selection-induced depletion, while linearity after normalization indicates successful removal of that curvature.

Table 7: Estimated gamma-frailty variance (fitted frailty variance) by
age band and vaccination status for Czech cohorts enrolled in 2021_24.

Age band (years)	Fitted frailty variance (Dose 0)	Fitted frailty variance (Dose 2)
40–49	16.79	2.66×10^{-6}
50–59	23.02	1.87×10^{-4}
60–69	13.13	7.01×10^{-18}
70–79	6.98	3.46×10^{-17}
80–89	2.97	2.03×10^{-11}
90–99	0.80	8.66×10^{-16}
All ages (full population)	4.98	1.02×10^{-11}

Notes: - The fitted frailty variance 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. - Each entry reports a single fitted gamma-frailty variance for the specified age band and vaccination status within the 2021_24 enrollment cohort. - The "All ages (full population)" row corresponds to an independent fit over the full pooled age range, included as a global diagnostic. - Table 8 reports raw outcome contrasts for ages 40+ (YOB ≤ 1980) where event counts are stable.

Diagnostic checks: - **Dose ordering:** the fitted frailty variance is positive for Dose 0 and collapses toward zero for Dose 2 across all age strata, consistent with selective uptake. - **Magnitude separation:** Dose 2 estimates are effectively zero relative to Dose 0, indicating near-linear cumulative hazards rather than forced curvature. - **Age coherence:** the fitted frailty variance decreases at older ages as baseline mortality rises and survivor populations become more homogeneous; monotonicity is not imposed. - **Stability:** No sign reversals, boundary pathologies, or numerical instabilities are observed. - **Falsifiability:** Failure of any one of these checks would constitute evidence against model adequacy.

Interpretation: Unvaccinated cohorts exhibit frailty heterogeneity, while Dose 2 cohorts show near-zero estimated frailty across all age bands, consistent with selective uptake prior to follow-up:

$$\hat{\theta}_d > 0$$

for Dose 0 cohorts and

$$\hat{\theta}_d \approx 0$$

for Dose 2 cohorts. 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. Because Table 7 demonstrates selection-induced heterogeneity, unadjusted

cumulative outcome contrasts are expected to conflate depletion effects with any true treatment differences; Table 8 therefore reports raw cumulative hazards solely as a pre-normalization diagnostic. KCOR normalization removes the depletion component, enabling interpretable comparison of the remaining differences.

Table 8: Ratio of observed cumulative mortality hazards for unvaccinated (Dose 0) versus fully vaccinated (Dose 2) Czech cohorts enrolled in 2021_24.

Age band (years)	Dose 0 cumulative hazard	Dose 2 cumulative hazard	Ratio
40–49	0.005260	0.004117	1.2776
50–59	0.014969	0.009582	1.5622
60–69	0.045475	0.023136	1.9655
70–79	0.123097	0.057675	2.1343
80–89	0.307169	0.167345	1.8355
90–99	0.776341	0.517284	1.5008
All ages (full population)	0.023160	0.073323	0.3159

This table reports unadjusted cumulative hazards derived directly from the raw data, prior to any frailty normalization or depletion correction, and is shown to illustrate the magnitude and direction of selection-induced curvature addressed by KCOR.

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 for the 2021_24 enrollment window (through week 2024-16), identically for Dose 0 and Dose 2 cohorts.

These raw contrasts reflect both selection and depletion effects and are not interpreted causally.

3.0.1 Illustrative empirical context: COVID-19 mortality data

The COVID-19 vaccination period provides a natural empirical regime characterized by strong selection heterogeneity and non-proportional hazards, making it a useful illustration for the KCOR framework. During this period, vaccine uptake was voluntary, rapidly time-varying, and correlated with baseline health status, creating clear examples of selection-induced non-proportional hazards. The Czech Republic national mortality registry data used in Tables 2 and 3 exemplify this regime: voluntary uptake led to asymmetric selection at enrollment, with vaccinated cohorts exhibiting minimal frailty heterogeneity while unvaccinated cohorts retained substantial heterogeneity. This asymmetric pattern reflects the healthy vaccinee effect operating through selective uptake rather than treatment. KCOR normalization removes this selection-induced curvature, enabling interpretable cumulative comparisons. While these examples illustrate KCOR's application, the method is general and applies to any retrospective cohort comparison where selection induces differential depletion dynamics.

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:

$$\hat{\theta}_d \approx 0$$

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 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 observed cohort hazards.
- **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 is constant 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 5 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(t) 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 6 and 7), 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:

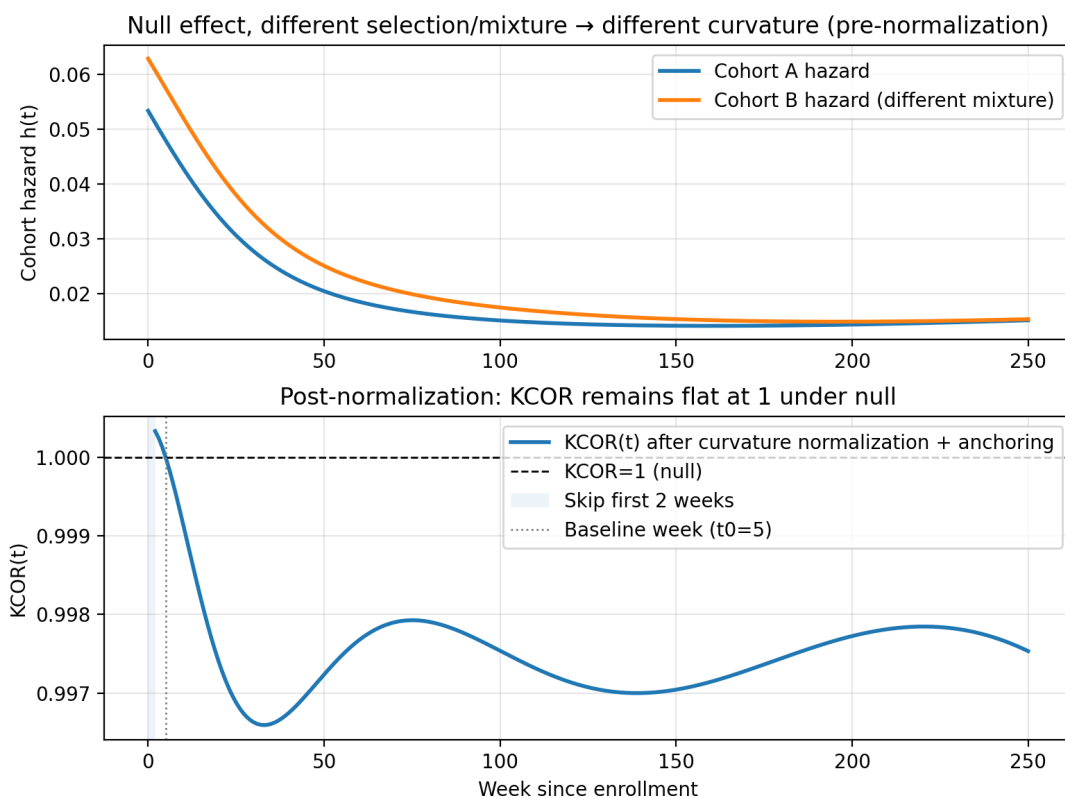Table 9 provides numeric summaries.

Figure 5: Synthetic negative control under strong selection (different curvature) but no effect: KCOR(t) 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.
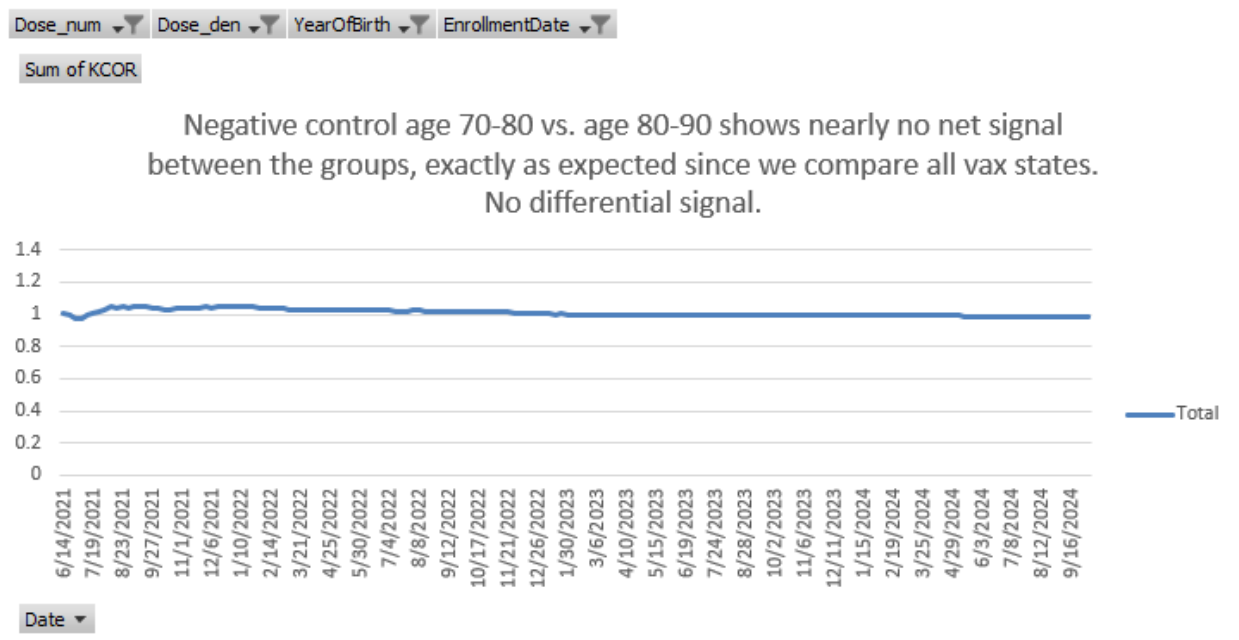
Figure 6: Empirical negative control with approximately 10-year age difference between cohorts. Despite large baseline mortality differences, KCOR(t) remains near-flat at 1 over follow-up, consistent with a true null effect. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Appendix B.2).[12]

Figure 7: Empirical negative control with approximately 20-year age difference between cohorts. Even under extreme composition differences, KCOR(t) 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 9: Example end-of-window KCOR(t) 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 8 and Table 10
confirm this behavior.

Table 10: 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 (default constant baseline over the fit window; alternatives can be assessed as
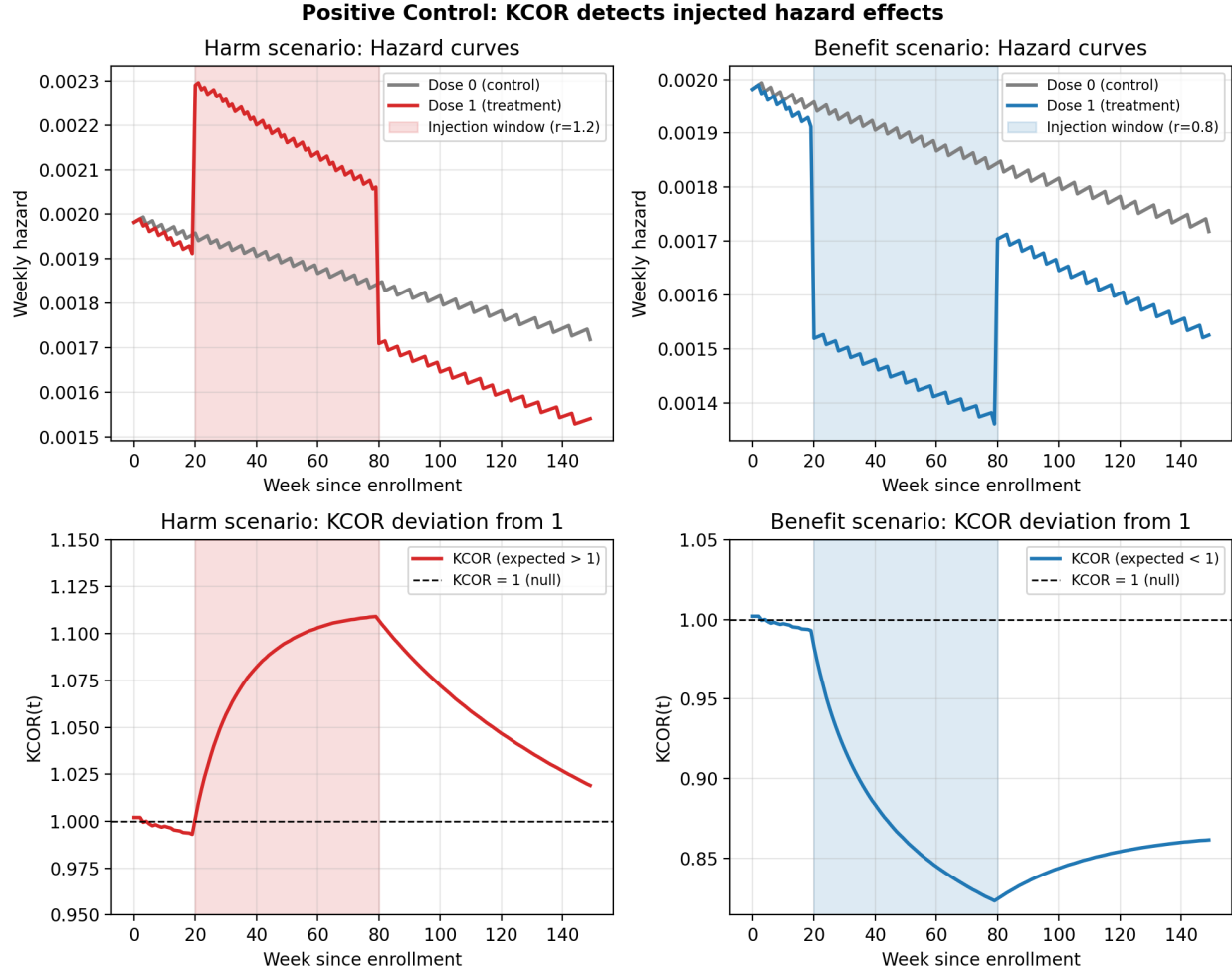 sensitivity)

Figure 8: 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.
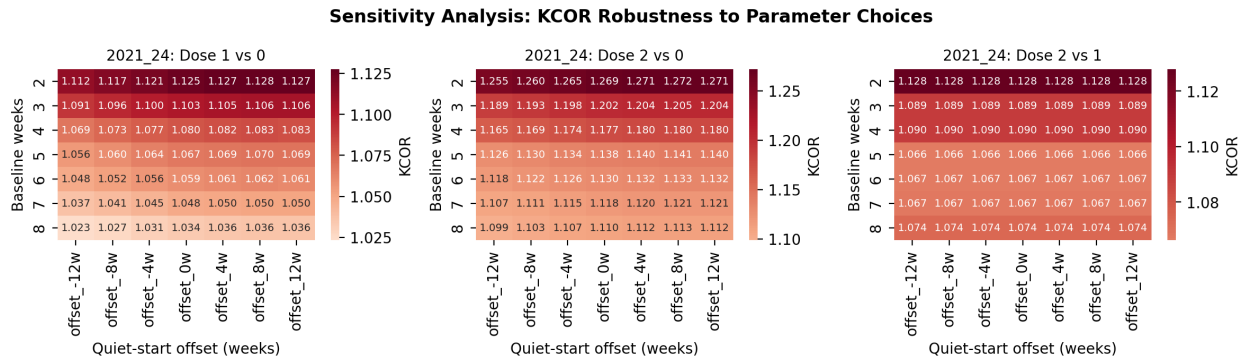

Figure 9: Sensitivity analysis summary showing KCOR(t) values across parameter grid. Heatmaps display KCOR(t) estimates for different combinations of baseline weeks (rows) and quiet-window start offsets (columns). Color scale centered at 1.0 shows stability of estimates across parameter choices, with values remaining close to 1.0 across the grid.

Figure 9 summarizes KCOR stability across the parameter grid.

Across all tested parameter ranges, KCOR(t) 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; unvaccinated 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 10 and 11 below.

Joint frailty and treatment-effect simulation (S7).
Figure 12 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(t) trajectories no longer stabilize during the nominal quiet period, and fit diagnostics degrade (Figure 13), including increased cumulative-hazard fit error and reduced post-normalization linearity. In these cases, KCOR(t) 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 times because both enrollments are subject to the same short-horizon deferral mechanisms. Therefore,
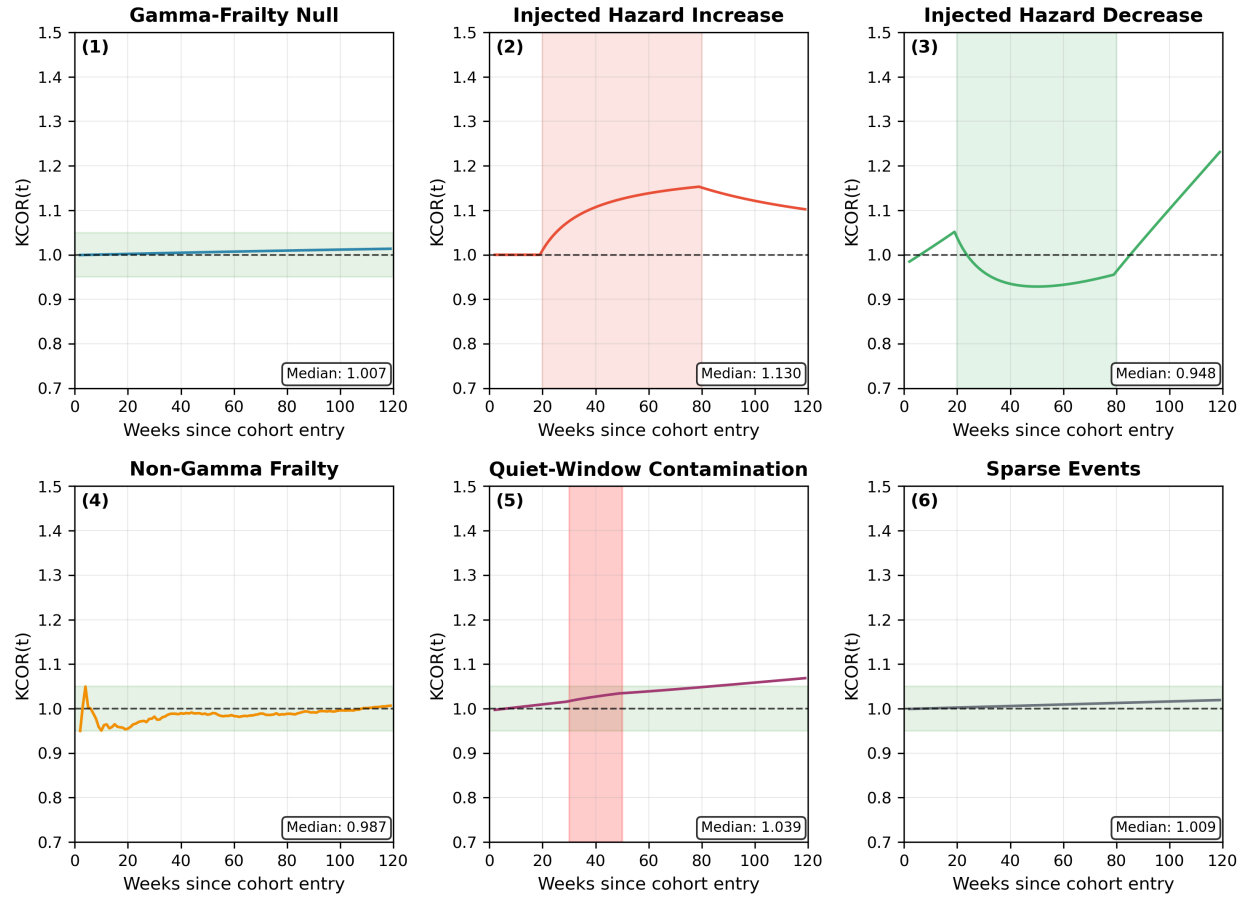
KCOR Simulation Grid: Operating Characteristics

Figure 10: 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(t) 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 11: 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.
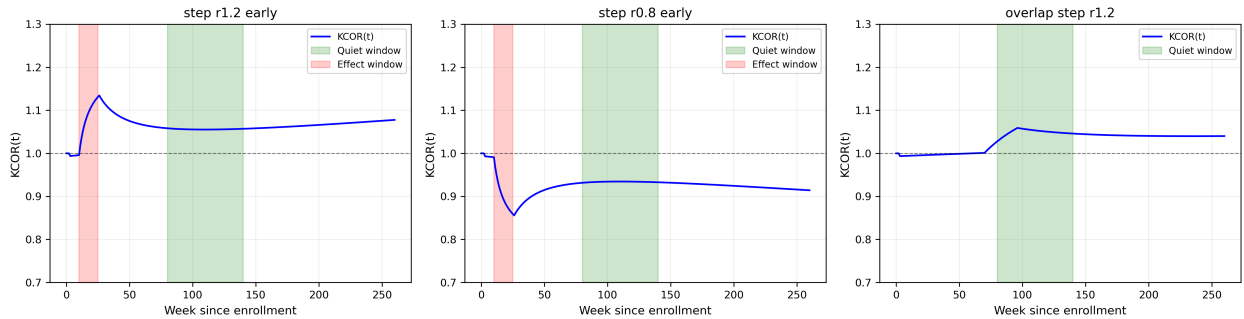

Figure 12: S7 simulation results: KCOR(t) trajectories demonstrating temporal separability. Left panel shows harm scenario (r=1.2) with effect window (weeks 10-25) and quiet window (weeks 80-140) non-overlapping. Middle panel shows benefit scenario (r=0.8). Right panel shows overlap variant where effect 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.
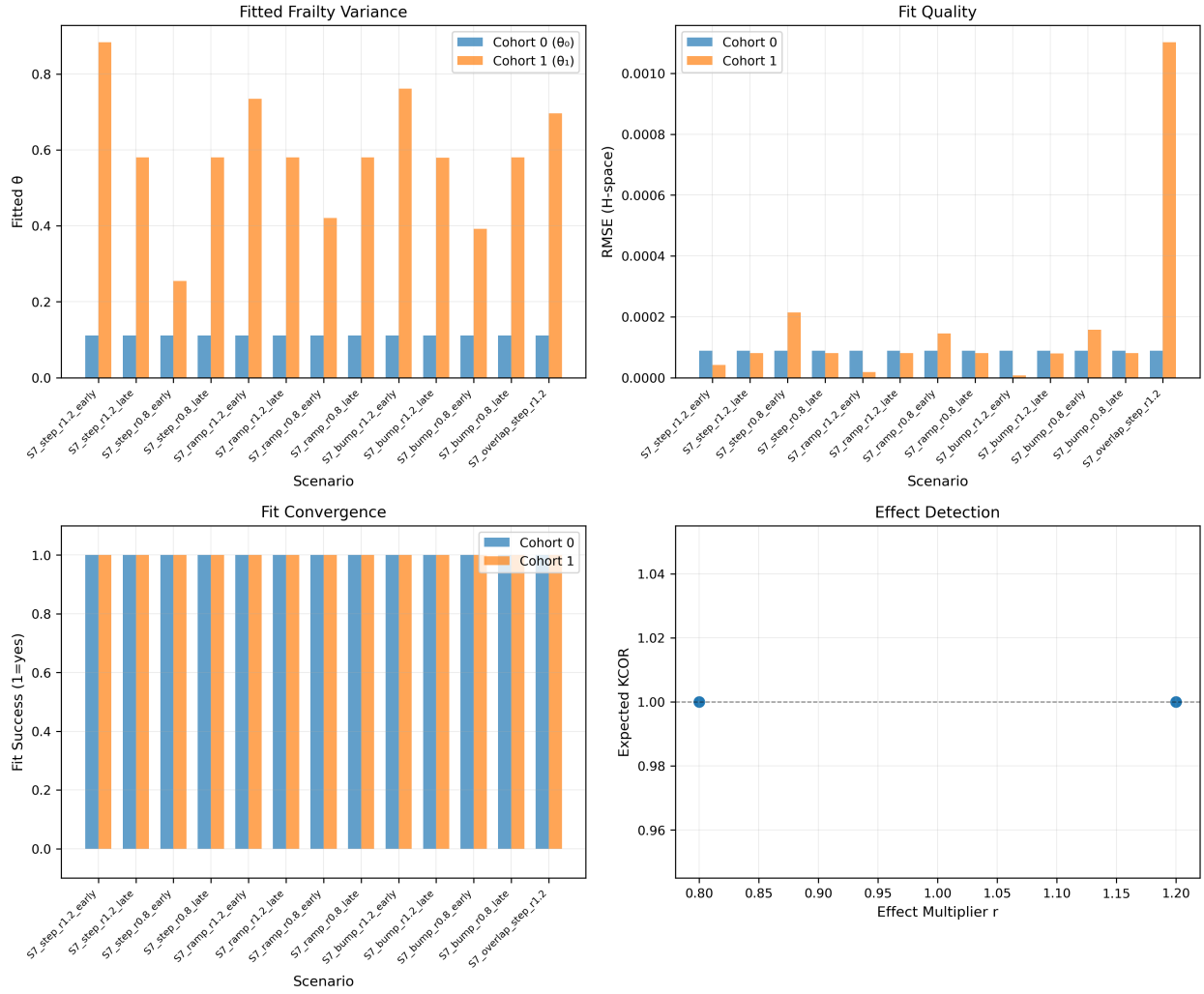
Figure 13: S7 simulation diagnostics: Fitted frailty variance parameters (θ_0, θ_1), fit quality (RMSE), and convergence status across S7 scenarios. The overlap variant shows degraded fit quality, correctly signaling violation of temporal separability assumptions.

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(t) 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

Scope statement (to avoid over-interpretation). KCOR is a normalization-and-comparison framework designed to remove selection-induced depletion curvature in cumulative-hazard space and then report a cumulative cohort contrast (KCOR). It does not claim to identify a treatment-specific causal effect under arbitrary baseline and time-varying confounding, and it should be interpreted in conjunction with prespecified controls, diagnostics, and sensitivity analyses.

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. Stabilization of KCOR(t) following frailty normalization is not an assumption but a falsification test; failure to flatten indicates residual curvature or loss of identifiability, not a substantive cumulative effect. 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. Whereas Cox regression infers treatment effects through fitted model coefficients, KCOR uses modeling only to normalize selection effects, allowing causal contrasts to be read directly from the adjusted

data via cumulative hazard ratios. For infectious-disease vaccines, any plausible mortality effect is inherently time-local, occurring only during periods of pathogen circulation; the proportional hazards assumption therefore fails structurally, as a constant multiplicative effect over the entire follow-up is biologically implausible.

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. A controlled synthetic null experiment (Section 2.11.1) shows that Cox regression can return statistically significant non-null hazard ratios solely from frailty-induced depletion—even when the true treatment effect is identically zero—whereas KCOR remains centered near unity with negligible post-normalization slope. 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 $\text{KCOR}(t)$.

A useful way to view KCOR is as an intermediate layer between purely descriptive hazard summaries and fully identified causal estimators. KCOR is descriptive in that it summarizes cohort differences in a cumulative-hazard scale under explicit normalization of depletion geometry; it is inferential in that it provides falsifiable diagnostics and control-test behavior that constrain when the normalized contrast is interpretable. This positioning is intentional: under minimal-data constraints, insisting on point-identification of a causal effect can be less reliable than explicitly normalizing a dominant bias geometry and transparently reporting when identifiability is not supported.

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. Having established the behavior of KCOR and the failure modes of standard estimators under controlled conditions, we apply KCOR to complete national registry data from the Czech Republic in a companion analysis. 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 (default constant baseline over the fit window) 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.

4.4 Extension to other outcomes

Although this paper focuses on all-cause mortality, KCOR is applicable to other irreversible outcomes provided that event timing and risk sets are well defined. Application to cause-specific mortality requires careful consideration of competing risks and interpretation of cumulative hazards within cause-restricted populations. Extension to non-fatal outcomes such as hospitalization is conceptually straightforward but may require additional attention to outcome definitions, censoring mechanisms, and recurrent events. These considerations affect interpretation rather than the core KCOR framework.

5. Limitations

- **Model dependence**: Normalization relies on the adequacy of the gamma-frailty model and the baseline-shape assumption during the quiet window.
- **Relation to existing non-PH methods**: KCOR is complementary to time-varying Cox, flexible parametric, additive hazards, and MSM approaches; these methods address different estimands and identification strategies, whereas KCOR targets depletion-geometry normalization under minimal-data constraints (see §1.3.1).
- **θ estimation is data-driven**: KCOR does not impose $\theta = 0$ for any cohort. The frequent observation that fitted frailty variance estimates collapse toward zero 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 fitted frailty variance 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 parameters 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, observed cumulative hazards become noisy and (k_d, θ_d) can be weakly identified, often manifesting as unstable fitted frailty variance estimates 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 fitted parameters under small perturbations of the quiet-window bounds.
- **Sensitivity analyses** over plausible quiet windows and skip-weeks values.
- **Prespecified negative controls**: KCOR(t) 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(t) 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(t) 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(t) 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. KCOR relies on five explicit assumptions, of which only one requires substantive dataset-specific validation, and enforces these assumptions diagnostically rather than presuming them, allowing violations to be detected rather than absorbed into model-dependent estimates. Because standard methods for retrospective vaccine evaluation fail under non-proportional hazards and selection-induced depletion, KCOR provides a novel and practical alternative that operates on minimal individual-level information—dates of birth, intervention, and death—while remaining applicable to national registry data where richer covariates are unavailable or unreliable.

Declarations

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 6 and 7) use aggregated cohort summaries derived from Czech Republic administrative data; no record-level data are shared in this manuscript.[12]

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 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

Competing interests

The author is a board member of the Vaccine Safety Research Foundation.

Funding

Authors' contributions

Steven T. Kirsch conceived the method, wrote the code, performed the analysis, and wrote the manuscript.

Acknowledgements

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. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526. doi:10.1093/biomet/81.3.515

5. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*. 1982;10(4). doi:10.1214/aos/1176345976

6. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*. 2002;21(15):2175-2197. doi:10.1002/sim.1203

7. Aalen OO. A linear regression model for the analysis of life times. *Statistics in Medicine*. 1989;8(8):907-925. doi:10.1002/sim.4780080803

8. Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika*. 1994;81(1):61-71. doi:10.1093/biomet/81.1.61

9. Van Houwelingen HC. Dynamic Prediction by Landmarking in Event History Analysis. *Scandinavian Journal of Statistics*. 2007;34(1):70-85. doi:10.1111/j.1467-9469.2006.00529.x

10. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology: *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011

11. Cole SR, Hernan MA. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*. 2008;168(6):656-664. doi:10.1093/aje/kwn164

12. 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.

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 d where individual i has hazard $h_{i,d}(t) = z_{i,d} \cdot h_{0,d}(t)$, with frailty $z_{i,d}$ drawn from a distribution with mean 1 and variance $\theta_d > 0$. Let $S_{i,d}(t) = \exp(-z_{i,d}H_{0,d}(t))$ be the individual survival function, where $H_{0,d}(t) = \int_0^t h_{0,d}(s)\,ds$.

The cohort survival function is the expectation over frailty:

$$S_d^{\text{cohort}}(t) = E_z[S_{i,d}(t)] = E_z\big[\exp(-zH_{0,d}(t))\big] = \mathscr{L}_z\big(H_{0,d}(t)\big),$$

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

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

Even when $h_{0,d}(t) = k_d$ is constant (so $H_{0,d}(t) = k_d t$), the cohort hazard $h_d^{\text{cohort}}(t)$ 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 $z \sim \text{Gamma}(\alpha = 1/\theta_d, \beta = 1/\theta_d)$ with mean 1 and variance θ_d, the Laplace transform is:

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

The cohort survival function becomes:

$$S_d^{\text{cohort}}(t) = \left(1 + \theta_d H_{0,d}(t)\right)^{-1/\theta_d}.$$

The observed cumulative hazard is defined as:

$$H_{\text{obs},d}(t) = -\log S_d^{\text{cohort}}(t).$$

Substituting the gamma Laplace transform yields the canonical gamma-frailty identity:

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

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

A.3 Inversion formula

Solving for $H_{0,d}(t)$ from the gamma-frailty identity gives the canonical inversion:

$$H_{0,d}(t) = \frac{\exp\left(\theta_d H_{\text{obs},d}(t)\right) - 1}{\theta_d}.$$

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

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_d.

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 KCOR(t) can be approximated via the delta method. Define:

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

If the variance of the depletion-neutralized cumulative hazard is available (e.g., from bootstrap or analytic propagation through the inversion), then:

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

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 5) 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 6 and 7) 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 8 and Table 10) 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 9) 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(t) 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(t) remains approximately flat and near unity during the quiet window,
2. KCOR(t) 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**: baseline hazard level and frailty variance. Very small frailty variance (< 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:

```
KCOR_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 should be examined for. Define residuals:

$$r_d(t) = H_{obs,d}(t) - H_d^{model}(t; \hat{k}_d, \hat{\theta}_d).$$

- **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 fitted parameters 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(t) trajectories remain qualitatively similar.
- **Baseline-shape alternatives**: Compare the default constant baseline over the fit window to mild linear trends and verify normalization is not sensitive to this choice.

The sensitivity analysis (§3.3 and Figure 9) 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.

Appendix D — Diagnostics and Failure Modes for KCOR Assumptions

This appendix describes the **observable diagnostics and failure modes** associated with each of the five KCOR assumptions (A1–A5). No additional assumptions are introduced here. KCOR is designed to **fail transparently rather than silently**: when an assumption is violated, the resulting lack of identifiability or model stress manifests through explicit diagnostic signals rather than spurious estimates.

D.1 Diagnostics for Assumption A1 (Fixed cohorts at enrollment)

Assumption A1 requires that cohorts be fixed at enrollment, with no post-enrollment switching or censoring in the primary estimand.

Diagnostic signals of violation.

- Inconsistencies in cohort risk sets (e.g., unexplained increases in at-risk counts).
- Early-time hazard suppression or inflation inconsistent with selection or depletion geometry.
- Dependence of results on as-treated reclassification or censoring rules.

Interpretation. KCOR is not defined for datasets with post-enrollment switching or informative censoring in the primary estimand. Such violations are design-level failures rather than modeling failures and indicate that KCOR should not be applied without redefining cohorts.

D.2 Diagnostics for Assumption A2 (Shared external hazard environment)

Assumption A2 requires that all cohorts experience the same calendar-time external mortality environment.

Diagnostic signals of violation.

- Calendar-time hazard spikes or drops that appear in only one cohort.
- Misalignment of major mortality shocks (e.g., epidemic waves) across cohorts.
- Cohort-specific reporting artifacts or administrative discontinuities.

Interpretation. External shocks are permitted under KCOR provided they act symmetrically across cohorts. Cohort-specific shocks violate comparability and are visible directly in calendar-time hazard overlays. When detected, such violations limit interpretation of KCOR contrasts over affected periods.

D.3 Diagnostics for Assumption A3 (Selection via time-invariant latent frailty)

Assumption A3 posits that selection at enrollment operates primarily through differences in a time-invariant latent frailty distribution that induces depletion of susceptibles.

Diagnostic signals of violation.

- Strongly structured residuals in cumulative-hazard space inconsistent with depletion.
- Instability of fitted frailty parameters not attributable to window placement.
- Early-time transients that do not decay and are inconsistent across related cohorts.

Interpretation. Frailty in KCOR is a geometric construct capturing unobserved heterogeneity, not a causal mechanism. If dominant time-varying individual risk unrelated to depletion is present, curvature attributed to frailty becomes unstable. Such cases are revealed by residual structure and parameter instability rather than masked by the model.

D.4 Diagnostics for Assumption A4 (Adequacy of gamma frailty approximation)

Assumption A4 requires that gamma frailty provides an adequate approximation to the depletion geometry observed in cumulative-hazard space over the estimation window.

Diagnostic signals of violation.

- Poor fit of the gamma-frailty cumulative-hazard model during the quiet window.
- Systematic residual curvature after frailty normalization.
- Strong sensitivity of results to minor model or window perturbations.

Additional internal diagnostics for Assumption A4 include the magnitude, coherence, and stability of the fitted frailty variance parameter (θ). Values of θ approaching zero are expected when cumulative hazards are approximately linear, while larger values correspond to visible depletion-induced curvature. Implausible θ estimates—such as large values in the absence of curvature, sign instability, or extreme sensitivity to small changes in the estimation window—indicate model stress or misspecification rather than substantive cohort effects.

Interpretation. Gamma frailty is used as a mathematically tractable approximation, not as a claim of biological truth. When depletion geometry deviates substantially from the gamma form, KCOR normalization fails visibly through poor fit and residual curvature. Such behavior indicates model inadequacy rather than supporting alternative interpretation.

D.5 Diagnostics for Assumption A5 (Quiet-window validity)

Assumption A5 requires the existence of a prespecified quiet window in which selection-induced depletion dominates other sources of curvature, permitting identification of frailty parameters.

Diagnostic signals of violation.

- Failure of KCOR(t) trajectories to stabilize or asymptote following frailty normalization.
- Persistent nonzero slope in KCOR(t), indicating residual curvature after normalization.
- Instability of fitted frailty parameters (θ) under small perturbations of quiet-window boundaries.
- Failure of depletion-neutralized cumulative hazards to become approximately linear during the quiet window.
- Degraded cumulative-hazard fit error concentrated within the nominal quiet period.

Interpretation. Quiet-window validity is the primary dataset-specific requirement for KCOR applicability. When this assumption fails—e.g., due to overlap with strong treatment effects or external shocks—KCOR does not amplify spurious signals. Instead, normalization becomes unstable and KCOR(t) trajectories attenuate toward unity or fail to stabilize, explicitly signaling loss of identifiability.

Under a valid quiet window, frailty-normalized cumulative hazards are expected to be approximately linear and KCOR(t) trajectories to stabilize rather than drift. Persistent KCOR slope or θ instability indicates contamination of the quiet window by external shocks or time-varying effects and signals loss of identifiability rather than evidence of cohort differences.

D.6 Diagnostic coherence across assumptions

Several diagnostics operate across assumptions A4 and A5, including stabilization of KCOR(t) trajectories and coherence of fitted θ parameters with observed cumulative-hazard curvature. These diagnostics are not assumptions of the KCOR framework; rather, they are observable consequences of successful frailty normalization. When these behaviors fail to emerge, KCOR explicitly signals reduced interpretability through residual curvature, parameter instability, or attenuation toward unity.

D.7 Identifiability under sparse data

KCOR does not require large sample sizes by assumption; however, reliable estimation of frailty parameters and depletion-neutralized cumulative hazards requires sufficient event information within the identification window. When event counts are very small, frailty estimates may become unstable, resulting in noisy normalization, non-linear baseline cumulative hazards, or drifting KCOR(t) trajectories.

Importantly, such failures are diagnosable: sparse-data regimes are characterized by instability of estimated frailty parameters under small perturbations of the quiet window, loss of post-normalization linearity, and non-stabilizing KCOR(t). In these cases, KCOR signals loss of identifiability rather than producing spurious effects. Applicability should therefore be assessed via diagnostic stability rather than nominal sample size thresholds.

D.8 Summary: Diagnostic enforcement rather than assumption inflation

KCOR relies on exactly five assumptions (A1–A5), stated exhaustively in §2.1.1. This appendix demonstrates that each assumption has **explicit, observable diagnostics** and **well-defined failure modes**. When assumptions are violated, KCOR signals reduced interpretability through instability, poor fit, or residual structure rather than producing misleading cumulative contrasts. This diagnostic enforcement is a core design feature of the KCOR framework.