KCOR: A depletion-neutralized framework for retrospective cohort comparison under latent frailty

Manuscript metadata

- **Article type**: Methods / Statistical method
- **Running title**: KCOR under selection-induced cohort bias
- **Author**: Steven T. Kirsch
- **Affiliations**: Independent Researcher, United States
- **Corresponding author**: stk@alum.mit.edu

- **Word count**: 10,909 (excluding Abstract and References)

- **Keywords**: selection bias; frailty model; gamma mixture model; frailty inversion; frailty heterogeneity; selection-induced depletion; non-proportional hazards; cumulative hazard; hazard normalization; cumulative hazards; estimands; gamma frailty; negative controls; observational studies; observational cohort studies

Abstract

Selection-induced depletion under latent frailty heterogeneity can generate non-proportional hazards and curvature in observed cumulative hazards, biasing standard survival estimands in retrospective cohort studies using registry and administrative data. KCOR is a depletion-neutralized cohort comparison framework based on gamma-frailty normalization that targets neutralization of a specific, diagnosable selection bias and emphasizes estimand alignment rather than causal recovery. It estimates cohort-specific depletion geometry during prespecified epidemiologically quiet periods (intervals of stable baseline risk) and applies an analytic inversion to map observed cumulative hazards into a common comparison scale prior to computing cumulative contrasts. Across simulations spanning frailty heterogeneity and selection strength and across negative and positive controls, Cox proportional hazards regression can exhibit systematic non-null behavior under selection-only regimes. In contrast, KCOR-normalized trajectories remain stable and centered near the null while detecting injected effects. KCOR provides a diagnostic and descriptive framework for comparing fixed cohorts under selection-induced hazard curvature by separating depletion normalization from outcome comparison and improving interpretability of cumulative outcome analyses under minimal-data constraints.

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.

Although mortality is used throughout this paper as a motivating and concrete example, the method applies more generally to any irreversible event process observed in a fixed cohort, including hospitalization, disease onset, or other terminal or absorbing states. Mortality is emphasized here because it is objectively defined, reliably recorded in many national datasets, and free from outcome-dependent ascertainment biases that complicate other endpoints.

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.

This manuscript is a methods paper. Real-world registry data are used solely to illustrate estimator behavior, diagnostics, and failure modes under realistic selection-induced non-proportional hazards; no policy conclusions are drawn. KCOR does not attempt to recover counterfactual survival curves or causal vaccine-effect estimates, and readers seeking causal VE should not expect them here.

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 selection-induced depletion (depletion of susceptibles) naturally induce curvature of the cumulative hazard (reflecting time-varying hazard) 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.

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. KCOR refers to the method as presented here; earlier internal iterations are not material to the estimand or results and are omitted for clarity.

Two mechanisms often lumped as the 'healthy vaccinee effect' (HVE) are distinguished here:

- **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.

Box 1. Two fundamentally different strategies for cohort comparability

- **Traditional matching and regression approaches:** attempt to construct comparable cohorts by matching or adjusting *characteristics of living individuals* at baseline or over follow-up, and then estimating effects via a fitted hazard model (e.g., Cox proportional hazards). This implicitly assumes that sufficiently rich covariate information can render cohorts exchangeable with respect to unobserved mortality risk.
- **Problem under latent frailty:** even meticulous 1:1 matching on observed covariates can fail to equalize mortality risk trajectories. In such settings, cohort differences arise not from mismeasured covariates, but from **selection-induced depletion of susceptibles**, which alters hazard curvature over time.
- **KCOR strategy:** rather than equating cohorts based on characteristics of the living, KCOR equates cohorts based on how they die in aggregate. It estimates cohort-specific depletion geometry from observed cumulative mortality during epidemiologically quiet periods, removes that geometry via analytic inversion, and compares cohorts on the resulting depletion-neutralized cumulative hazard scale.
- **Inferential distinction:** Cox-type methods are **model-based and individual-level**, conditioning on survival and fitting covariate effects, whereas KCOR is **measurement-based and cohort-level**, operating directly on aggregated mortality trajectories without fitting covariate models. The inferential target is cumulative outcome accumulation rather than an instantaneous hazard ratio conditional on survival.

1.3 Related work (brief positioning)

KCOR builds on the frailty and selection-induced depletion literature in which unobserved heterogeneity induces deceleration of cohort-level hazards over follow-up (a standard working model is gamma frailty).[1] KCOR's distinct contribution is not additional hazard flexibility, but a **diagnostics-driven normalization** of selection-induced depletion geometry in cumulative-hazard space prior to defining a cumulative cohort contrast. Related approaches that address non-proportional hazards (time-varying effects, flexible parametric hazards, additive hazards) or time-varying confounding (MSM/IPW/g-methods) target different estimands and typically require richer longitudinal covariates than are available in minimal registry data.[2–9] Additional discussion is provided in the Supplementary Information (SI).

Time-varying coefficient Cox models allow hazards to change over time but do not neutralize frailty-induced depletion because estimation remains conditional on survival; they therefore address non-proportionality without removing selection-induced curvature.

Marginal structural models target causal effects under exchangeability using longitudinal covariates and weighting; KCOR instead targets descriptive cumulative contrasts under minimal data, so the estimands, assumptions, and failure modes differ.

Flexible parametric survival models improve baseline fit but do not resolve depletion-induced selection bias when frailty heterogeneity is present.

1.4 Evidence from the literature: residual confounding despite meticulous matching

Motivating applied studies suggest that even careful matching and adjustment can leave substantial residual differences in non-COVID mortality and time-varying "healthy vaccinee effect" signatures, consistent with selection and depletion dynamics not captured by measured covariates.[10,11]

1.5 Contribution of this work

This work makes four primary contributions: (i) it formalizes selection-induced depletion under latent frailty heterogeneity as a source of non-proportional hazards and curvature that can bias common survival estimands; (ii) it defines a diagnostics-first normalization that fits depletion geometry in quiet periods and maps observed cumulative hazards into a depletion-neutralized space; (iii) it validates operating characteristics using synthetic and empirical controls, including a synthetic null under selection-only regimes; and (iv) it separates normalization from comparison by permitting standard post-normalization cumulative estimands.

A central implication is identifiability: in minimal-data retrospective cohorts, interpretability depends on an epidemiologically quiet window and on internal diagnostics that indicate depletion geometry has been estimated and removed, rather than absorbed into a time-varying effect estimate.

Together, these contributions position KCOR not as a replacement for existing survival estimands, but as a prerequisite normalization step that addresses a source of bias arising prior to model fitting in many retrospective cohort studies.

1.6 Target estimand and scope (non-causal)

Box 2. Target estimand and scope (non-causal)

- **Primary estimand (KCOR)**: For two fixed enrollment cohorts A and B, it is a **neutralized marginal estimand**, defined as
$$\mathrm{KCOR}(t) = \tilde{H}_{0,A}(t)/\tilde{H}_{0,B}(t),$$
where $\tilde{H}_{0,d}(t)$ is cohort d's **depletion-neutralized baseline cumulative hazard** obtained by fitting depletion geometry in a prespecified quiet window and applying the gamma-frailty inversion (Methods §2).
- **Operational summary**: KCOR proceeds by (i) estimating selection-induced depletion geometry during an epidemiologically quiet period, (ii) inverting that geometry to obtain depletion-neutralized cumulative hazards, and (iii) comparing cohorts on that normalized cumulative scale.
- **Interpretation**: KCOR is a time-indexed **cumulative** contrast on the depletion-neutralized scale. Values

above/below 1 indicate greater/less cumulative event accumulation in cohort A than B by time t after depletion normalization. KCOR is not an instantaneous hazard ratio.
- **What it is not**: KCOR is **not** a causal effect estimator (no ATE/ATT) and does not recover counterfactual outcomes under hypothetical interventions.
- **When interpretable**: Interpretation is conditional on explicit assumptions (fixed cohorts; shared external hazard environment; adequacy of the working frailty model; existence of an epidemiologically quiet window) **and** on internal diagnostics (quiet-window fit quality; post-normalization linearity within the quiet window; parameter stability to small window perturbations).
- **If diagnostics indicate non-identifiability**: the analysis is treated as not identified and KCOR is not reported as a "normalized contrast".

1.7 Paper organization and supporting information (SI)

The main text introduces the KCOR estimator, provides a canonical demonstration of Cox bias under frailty-driven depletion, and presents two primary validation examples (a negative control and a stress test). Additional validations—including positive controls—along with extended diagnostics, empirical registry-based nulls, and detailed simulation specifications are provided in the Supplementary Information (SI; Sections S2 and S4.2–S4.3; Tables S1–S3).

2. Methods

Mortality is used as the primary example throughout this section because it is objectively defined and reliably recorded in many administrative datasets.

Table 4 defines the notation used throughout the Methods section.

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 and estimand

Retrospective cohort differences can arise from two qualitatively different components:

- **Level differences**: cohort hazards differ by an approximately time-stable multiplicative factor (or, equivalently, cumulative hazards have different slopes but similar shape).
- **Depletion (curvature) differences**: cohort hazards evolve differently over time because cohorts differ in latent heterogeneity and are **selectively depleted** at different rates.

This framework targets the second failure mode. Under latent frailty heterogeneity, high-risk individuals die earlier, so the surviving risk set becomes progressively "healthier." This induces **downward curvature** (deceleration) in cohort hazards and corresponding concavity in cumulative-hazard space, even when individual-level hazards are simple and even under a true null treatment effect. When selection concentrates frailty heterogeneity differently across cohorts, the resulting curvature differences produce strong non-proportional hazards and can drive misleading contrasts for estimands that condition on the evolving risk set.

KCOR does not assert a causal interpretation of the resulting contrast; rather, it isolates residual differences that cannot be attributed to frailty-driven selection, thereby identifying regimes that demand substantive explanation beyond selection artifacts.

The strategy is therefore:

1. **Estimate the cohort-specific depletion geometry** (via curvature) during prespecified epidemiologically quiet periods.
2. **Map observed cumulative hazards into a depletion-neutralized space** by inverting that geometry.
3. **Compare cohorts only after normalization** using a prespecified post-adjustment estimand; ratios of depletion-neutralized cumulative hazards (KCOR) are used here.

All analyses are performed using discrete weekly time bins; continuous-time notation is used solely for expositional convenience. See Table 4 for the full symbol list.

Notation preview.
Throughout this paper, t denotes event time since cohort enrollment and d indexes cohorts. Let $H_{\mathrm{obs},d}(t)$ denote the observed cohort-level cumulative hazard, computed from fixed risk sets, and let $\tilde{H}_{0,d}(t)$ denote the corresponding *depletion-neutralized baseline cumulative hazard* obtained after frailty normalization. Individual hazards are modeled using a latent multiplicative frailty term z, with cohort-specific variance θ_d, which governs the strength of selection-induced depletion and resulting curvature in observed cumulative hazards. Full notation is summarized in Table 4.

2.1.1 Target estimand

Scope and interpretation are summarized in Box 2 (§1.6); the formal definition used throughout is provided here.

Let $\tilde{H}_{0,d}(t)$ denote the **depletion-neutralized baseline cumulative hazard** for cohort d at event time t since enrollment (Table 4). For two cohorts A and B, KCOR is defined as

$$\mathrm{KCOR}(t) = \frac{\tilde{H}_{0,A}(t)}{\tilde{H}_{0,B}(t)}. \tag{1}$$

For visualization, an **anchored KCOR** is sometimes reported to show post-reference divergence:

$$\mathrm{KCOR}(t; t_0) = \mathrm{KCOR}(t)/\mathrm{KCOR}(t_0),$$

with prespecified t_0 (e.g., 4 weeks). Anchoring is used only when explicitly stated to remove pre-existing level differences and emphasize post-reference divergence.

2.1.2 Identification versus diagnostics

Scope and interpretation are summarized in Box 2 (§1.6).

Interpretability of a KCOR trajectory is assessed via prespecified diagnostics (Supplementary Information §S2; Tables S1–S3). When diagnostics indicate non-identifiability, the analysis is treated as not identified and results are not reported. Checks include:

- stability of $(\hat{k}_d, \hat{\theta}_d)$ to small quiet-window perturbations,
- approximate linearity of $\tilde{H}_{0,d}(t)$ within the quiet window,
- absence of systematic residual structure in cumulative-hazard space.

Diagnostics corresponding to each assumption are summarized in Supplementary Information §S2 (Tables S1–S3).

2.1.3 KCOR assumptions and diagnostics

These assumptions define when KCOR normalization is interpretable.

The KCOR framework relies on the following assumptions, which are framed diagnostically:

1. **Fixed cohort enrollment.** Cohorts are defined at a common enrollment time and followed forward without dynamic entry or rebalancing.
2. **Multiplicative latent frailty.** Individual hazards are assumed to be multiplicatively composed of a baseline hazard and an unobserved frailty term, with cohort-specific frailty distributions.
3. **Quiet-window stability.** A prespecified epidemiologically quiet period exists during which external shocks to the baseline hazard are minimal, allowing depletion geometry to be estimated from observed cumulative hazards.
4. **Independence across strata.** Cohorts or strata are analyzed independently, without interference, spillover, or cross-cohort coupling.

5. **Sufficient event-time resolution.** Event timing is observed at a temporal resolution adequate to estimate cumulative hazards over the quiet window.

If no candidate quiet window satisfies diagnostic criteria, KCOR is not interpretable and analysis should terminate without reporting contrasts. Multiple disjoint quiet windows may be pooled provided fitted depletion parameters are stable and diagnostics are consistent across windows. Quiet-window validity diagnostics are summarized in Supplementary Information §S2 (Tables S1–S3).

These assumptions are evaluated empirically using post-normalization diagnostics. Violations are expected to manifest as residual curvature, drift, or instability in adjusted cumulative hazard trajectories.

2.2 Cohort construction

KCOR is defined for **fixed cohorts at enrollment**. Required inputs are minimal: enrollment date(s), event date, and optionally birth date (or year-of-birth) for age stratification. Analyses proceed in discrete event time t (e.g., weeks) measured since cohort enrollment.

Cohorts are assigned by intervention state at the start of the enrollment interval. In the primary estimand:

- **No post-enrollment switching** is allowed (individuals remain in their enrollment cohort),
- **No censoring** is applied (other than administrative end of follow-up),
- analyses are performed on the resulting fixed risk sets.

Censoring or reclassification due to cohort transitions (e.g., moving between exposure groups over time) is not permitted, because such transitions alter the frailty composition of the cohort in a time-dependent manner. Allowing transitions would introduce additional, endogenous selection that changes cohort mortality trajectories in unpredictable ways, confounding depletion effects that KCOR is designed to normalize.

This fixed-cohort design is intentional. It avoids immortal-time artifacts and prevents outcome-driven switching rules from creating time-dependent selection that is difficult to diagnose under minimal covariate availability. Extensions that allow switching or censoring are treated as sensitivity analyses (§5.2) because they change the estimand and introduce additional identification requirements.

Conceptual requirements of the KCOR framework are distinguished from operational defaults, which are reported separately for reproducibility (Supplementary Section S4).

Throughout this manuscript the failure event is **all-cause mortality**. KCOR therefore targets cumulative mortality hazards and is not framed as a cause-specific competing-risks analysis.

2.3 Hazard estimation and cumulative hazards in discrete time

For each cohort d, let $N_d(0)$ denote the number of individuals at enrollment. Let $d_d(t)$ denote deaths occurring during interval t, and let

$$D_d(t) = \sum_{s \le t} d_d(s)$$

denote cumulative deaths up to the end of interval t.

Define the risk set size at the start of interval t as

$$N_d(t) = N_d(0) - \sum_{s<t} d_d(s) = N_d(0) - D_d(t-1).$$

In the primary estimand, individuals do not switch cohorts after enrollment and there is no loss to follow-up; therefore $N_d(t)$ is the risk set used to define all discrete-time hazards and cumulative hazards in this manuscript.

Define the interval mortality ratio

$$\mathrm{MR}_{d,t} = \frac{d_d(t)}{N_d(t)}.$$

The discrete-time cohort hazard is computed as

$$h_{\text{obs},d}(t) = -\ln\big(1 - \text{MR}_{d,t}\big) = -\ln\Big(1 - \frac{d_d(t)}{N_d(t)}\Big). \tag{2}$$

This transform is standard: it maps an interval event probability into a continuous-time equivalent hazard under a piecewise-constant hazard assumption. For rare events, $h_{\text{obs},d}(t) \approx \text{MR}_{d,t} = d_d(t)/N_d(t)$, but the log form remains accurate and stable when weekly risks are not negligible. The exact transform $h_{\text{obs},d}(t) = -\log(1 - \text{MR}_{d,t})$ is used throughout; the rare-event approximation is provided for intuition only. This discrete-time hazard transform under a piecewise-constant hazard assumption is standard; see, e.g., Kalbfleisch and Prentice (*The Statistical Analysis of Failure Time Data*).[12]

All hazard and cumulative-hazard quantities used in KCOR are discrete-time integrated hazard estimators derived from fixed-cohort risk sets; likelihood-based or partial-likelihood formulations are not used for estimation or for the subsequent frailty-based normalization.

Observed cumulative hazards are accumulated over event time after an optional stabilization skip (§2.7):

$$H_{\text{obs},d}(t) = \sum_{s \le t} h_d^{\text{eff}}(s), \qquad \Delta t = 1. \tag{3}$$

Discrete binning accommodates tied events and aggregated registry releases. Bin width is chosen based on diagnostic stability (e.g., smoothness and sufficient counts per bin) rather than temporal resolution alone.

In addition to the primary implementation above, $\hat{H}_{\text{obs},d}(t)$ was computed using the Nelson–Aalen estimator $\sum_{s \le t} d_d(s)/N_d(s)$ as a sensitivity check; results were unchanged.

2.4 Selection model: gamma frailty and depletion normalization

2.4.1 Individual hazards with multiplicative frailty

Let $z_{i,d}$ denote an individual-specific latent frailty term with mean 1 and variance θ_d, and let $\tilde{h}_{0,d}(t)$ denote the depletion-neutralized baseline hazard for cohort d. Individual hazards are modeled as:

$$h_{i,d}(t) = z_{i,d}\,\tilde{h}_{0,d}(t), \qquad z_{i,d} \sim \text{Gamma}(\text{mean} = 1,\ \text{var} = \theta_d). \tag{4}$$

Here $\tilde{h}_{0,d}(t)$ is the cohort's depletion-neutralized baseline hazard and $z_{i,d}$ is a latent multiplicative frailty term. The frailty variance θ_d governs the strength of depletion-induced curvature: larger θ_d yields stronger deceleration at the cohort level due to faster early depletion of high-frailty individuals.

Gamma frailty is used because it yields a closed-form link between observed and baseline cumulative hazards via the Laplace transform.[1] In KCOR, gamma frailty is a **working geometric model** for depletion normalization, not a claim of biological truth. Adequacy is evaluated empirically via fit quality, post-normalization linearity, and stability diagnostics.

2.4.2 Gamma-frailty identity and inversion

Let

$$\tilde{H}_{0,d}(t) = \int_0^t \tilde{h}_{0,d}(s)\,ds \tag{5}$$

denote the depletion-neutralized baseline cumulative hazard. Let θ_d denote the cohort-specific frailty variance governing selection-induced depletion. Under a gamma-frailty working model, the observed cohort-level cumulative hazard satisfies:

$$H_{\text{obs},d}(t) = \frac{1}{\theta_d}\,\log\big(1 + \theta_d \tilde{H}_{0,d}(t)\big), \tag{6}$$

Given an estimate $\hat{\theta}_d$, the observed cumulative hazard can be mapped into depletion-neutralized baseline cumulative hazard space via exact inversion:

$$\tilde{H}_{0,d}(t) = \frac{\exp\big(\theta_d H_{\mathrm{obs},d}(t)\big) - 1}{\theta_d}. \tag{7}$$

This inversion is the **normalization operator**: given an estimate $\hat{\theta}_d$, it maps the observed cumulative hazard $H_{\mathrm{obs},d}(t)$ into a depletion-neutralized cumulative hazard scale. We use a tilde (e.g., $\tilde{H}_{0,d}(t)$) to denote depletion-neutralized baseline quantities obtained after frailty normalization; observed cohort-aggregated quantities are written without a tilde (e.g., $H_{\mathrm{obs},d}(t)$).

2.4.3 Baseline shape used for frailty identification

To identify θ_d, KCOR fits the gamma-frailty model within prespecified epidemiologically quiet periods. In the reference specification, the baseline hazard is taken to be constant over the fit window:

$$\tilde{h}_{0,d}(t) = k_d, \qquad \tilde{H}_{0,d}(t) = k_d\, t. \tag{8}$$

This choice intentionally minimizes degrees of freedom: during a quiet window, curvature is forced to be explained by depletion (via θ_d) rather than by introducing time-varying baseline hazard terms. If the observed cumulative hazard is near-linear over the fit window, the model naturally collapses toward $\hat{\theta}_d \approx 0$, signaling weak or absent detectable depletion curvature for that cohort over that window.

2.4.4 Quiet-window validity as the key dataset-specific requirement

Frailty parameters are estimated using only bins whose corresponding calendar weeks lie inside a prespecified quiet window (defined in ISO-week space). The quiet window is prespecified to avoid sharp, cohort-differential hazard perturbations (e.g., epidemic waves or policy shocks) that would confound depletion-geometry estimation. A window is acceptable only if diagnostics indicate (i) good fit in cumulative-hazard space, (ii) post-normalization linearity within the window, and (iii) stability of $(\hat{k}_d, \hat{\theta}_d)$ to small boundary perturbations. If no candidate window passes, diagnostics indicate non-identifiability and the analysis is treated as not identified rather than producing a potentially misleading normalized contrast; in that case, KCOR curves and summary values are not reported. All diagnostics are computed over discrete event-time bins (weekly intervals since enrollment) whose corresponding calendar weeks fall within the prespecified quiet window.

Quiet-window selection protocol (operational)

Quiet-window selection is prespecified and evaluated using diagnostic criteria summarized in Supplementary Information §S2 (Tables S1–S3).

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

KCOR estimates $(\hat{k}_d, \hat{\theta}_d)$ independently for each cohort d using only time bins that fall inside a prespecified quiet window in calendar time (see §2.4.4). The quiet window is applied consistently across cohorts within an analysis. 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 $t \geq$ SKIP_WEEKS. This stabilization threshold is prespecified and diagnostic-driven rather than tuned to outcomes.

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

$$H_d^{\mathrm{model}}(t; k_d, \theta_d) = \frac{1}{\theta_d} \log(1 + \theta_d k_d t)\,. \tag{9}$$

Identifiability of $(\hat{k}_d, \hat{\theta}_d)$ arises from curvature in cumulative-hazard space: observed cumulative hazards are nonlinear in follow-up time when $\theta_d > 0$. When depletion is weak (or the quiet window is too short to exhibit

curvature), the model smoothly approaches 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 fitted frailty variance estimates 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. Thus, lack of identifiable curvature manifests as fitted frailty variance estimates approaching zero, serving as 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 rather than reflecting a modeling assumption. When residual time-varying risk contaminates a nominally quiet window, fitted frailty variance estimates similarly shrink toward zero, signaling limited identifiability rather than inducing spurious normalization.

Parameters are estimated by constrained nonlinear least squares:

$$(\hat{k}_d, \hat{\theta}_d) = \arg \min_{k_d>0,\ \theta_d \geq 0} \sum_{t \in \mathcal{T}_d} \left(H_{\text{obs},d}(t) - H_d^{\text{model}}(t; k_d, \theta_d)\right)^2. \tag{10}$$

The boundary case $\theta_d = 0$ is admissible; in this limit the model-implied cumulative hazard reduces to $H_d^{\text{model}}(t) = k_d t$, and the normalization step leaves observed cumulative hazards unchanged.

Fitting is performed in cumulative-hazard space rather than via likelihood maximization, as the inputs are discrete-time, cohort-aggregated hazards rather than individual-level event histories. Least-squares fitting serves as a stable estimating equation for selection-induced depletion during quiet periods, emphasizes agreement in hazard shape rather than instantaneous risk, and yields diagnostics (e.g., RMSE and residual structure in H-space) that directly assess identifiability. Likelihood-based fitting may be used as a sensitivity analysis but is not required for the gamma-frailty normalization identity.

All analyses use a prespecified reference implementation with fixed operational defaults; full details are provided in Supplementary Section S4.

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 $(\hat{\theta}_d, \hat{k}_d)$ to hazard curvature has been factored out. This normalization defines a common comparison scale in cumulative-hazard space; it is not equivalent to Cox partial-likelihood baseline anchoring, but serves an analogous geometric role for cumulative contrasts. In this space, cumulative hazards are directly comparable across cohorts, and remaining differences reflect real differences in baseline risk rather than selection-induced depletion. The core identities used in KCOR are given in Equations (2), (10), (11), and (1). Normalization defines a common comparison scale; the scientific estimand is then computed on that scale (Box 2).

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. **Numerical stability at vanishing frailty variance.**

When the fitted frailty variance satisfies $\hat{\theta}_d < \varepsilon$ (with ε set to a small numerical tolerance, e.g., 10^{-8}), the gamma-frailty inversion in Eq. (@eq:normalized-cumhazard) is evaluated in the $\theta \to 0$ limit, yielding

$$\tilde{H}_{0,d}(t) = H_{\mathrm{obs},d}(t),$$

which corresponds to the standard Nelson–Aalen cumulative hazard. This avoids numerical instability from evaluating $(\exp(\theta H) - 1)/\theta$ near machine precision and reflects the fact that near-zero fitted frailty variance indicates negligible detectable depletion curvature.

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.** Within the prespecified quiet window (see §2.4.4), the depletion-neutralized cumulative hazard should be approximately linear in event time after inversion. Systematic residual curvature indicates window contamination (external shocks, secular trends) or misspecified depletion geometry for that cohort.
2. **Fit residual structure in cumulative-hazard space.** Define residuals over the fit set $\mathcal{T}_d$:

$$r_d(t) = H_{\mathrm{obs},d}(t) - H_d^{\mathrm{model}}(t; \hat{k}_d, \hat{\theta}_d). \tag{12}$$

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 normalization.

These diagnostics are reported alongside KCOR(t) curves. 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 normalized estimate. When diagnostics indicate non-identifiability, the depletion-based normalization is inappropriate and KCOR should not be interpreted.

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 $h_d^{\text{eff}}(t)$ as in §2.3:

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

2.8 KCOR estimator

With depletion-neutralized cumulative hazards in hand, the primary KCOR trajectory is defined as:

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

This ratio is computed after depletion normalization and is interpreted conditional on the stated assumptions and diagnostics (Box 2; §2.1.2).

2.9 Uncertainty quantification

Uncertainty is quantified using stratified bootstrap resampling, which propagates sampling variability of the aggregated process through the full pipeline (event counts, frailty fitting, inversion, and KCOR computation).

2.9.1 Stratified bootstrap procedure

Throughout this manuscript, cohorts define the primary comparison groups and are the units at which frailty parameters (k_d, θ_d) are estimated. Strata (e.g., age bands or sex) are treated as independent realizations of the same cohort-level process and are used solely for aggregation, age standardization, and uncertainty propagation; frailty parameters are not estimated separately by stratum. This use of stratification follows standard survival-analysis practice (e.g., Kalbfleisch and Prentice, The Statistical Analysis of Failure Time Data).[12]

The stratified bootstrap procedure for KCOR proceeds as follows:

1. **Resample cohorts using a block-preserving aggregated bootstrap.**
   For each cohort and stratum, bootstrap replicates are generated by resampling **contiguous blocks of event-time bins** with replacement, preserving the internal temporal dependence of the counting process. Within each resampled block, deaths $d_d(t)$ and corresponding risk sets $N_d(t)$ are carried forward sequentially so that

   $$N_d(t) = N_d(0) - \sum_{s<t} d_d(s)$$

   holds identically within each replicate. This preserves the martingale covariance structure of the cumulative hazard estimator while operating on aggregated cohort–time data.

2. **Re-estimate frailty parameters.** For each bootstrap replicate, re-estimate $(\hat{k}_d, \hat{\theta}_d)$ independently for each cohort d using the resampled data, applying the same quiet-window selection and fitting procedure as in the primary analysis.

3. **Recompute normalized cumulative hazards.** Using the bootstrap-estimated frailty parameters, recompute $\tilde{H}_{0,d}(t)$ for each cohort using Eq. (11) applied to the resampled observed cumulative hazards.

4. **Recompute KCOR.** Compute $\text{KCOR}(t)$ for each bootstrap replicate as the ratio of the bootstrap-normalized cumulative hazards.

5. **Form percentile intervals.** From the bootstrap distribution of $\text{KCOR}(t)$ values at each time point, form percentile-based confidence intervals (e.g., 2.5th and 97.5th percentiles for 95% intervals).

Importantly, event-time bins are **not resampled independently**. Any bootstrap procedure that breaks the sequential dependence between $d_d(t)$ and $N_d(t)$ would underestimate uncertainty in cumulative hazards. The block-preserving aggregated bootstrap used here maintains the temporal structure required for valid uncertainty propagation in cumulative-hazard space.

Bootstrap resampling is performed at the cohort-count level in the aggregated representation by resampling contiguous blocks of $(d_d(t), N_d(t))$ event-time pairs within cohort-time strata with replacement, preserving within-cohort temporal structure, rather than resampling individual-level records.

Uncertainty intervals reflect event stochasticity and model-fit uncertainty in $(\hat{k}_d, \hat{\theta}_d)$ and are interpreted as sampling variability of the aggregated counting process, not uncertainty in individual-level causal effects.

2.10 Algorithm summary and reproducibility checklist

Table 5 summarizes the complete KCOR pipeline.

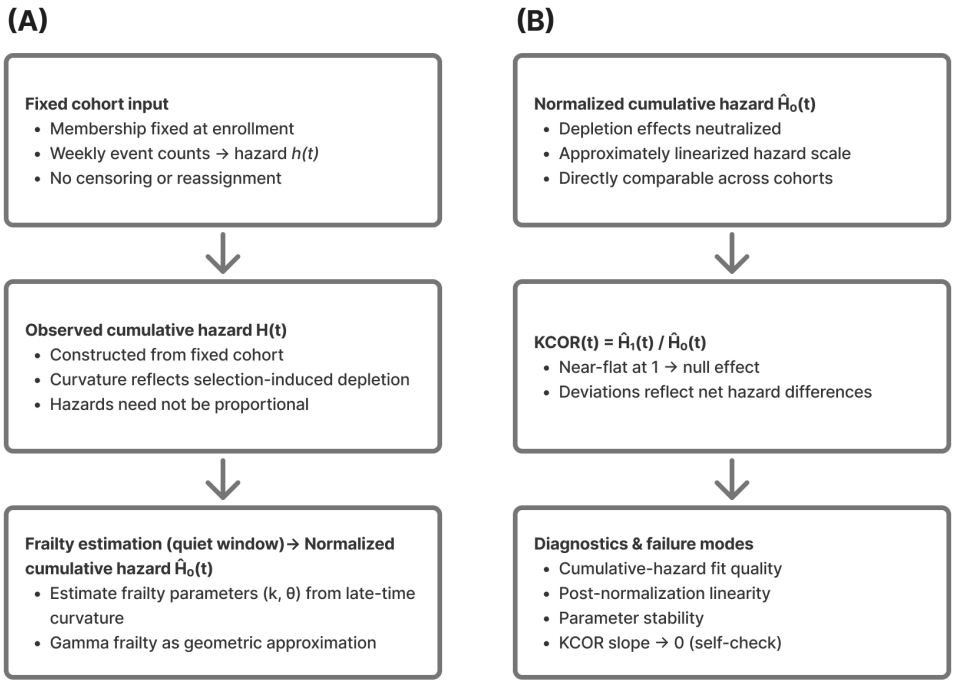

Figure 1: **KCOR as a two-stage framework. (A)** Fixed-cohort cumulative hazards exhibit curvature due to selection-induced depletion; late-time curvature is used to estimate frailty parameters for normalization. **(B)** Gamma-frailty normalization yields approximately linearized cumulative hazards that are directly comparable across cohorts; $\mathrm{KCOR}(t)$, defined as the ratio of depletion-neutralized baseline cumulative hazards, is expected to be horizontal under the null, subject to sampling stochasticity, and deviates only under net hazard differences. *This schematic is illustrative rather than empirical. In the schematic, $\tilde{H}_{0,d}(t)$ denotes the depletion-neutralized baseline cumulative hazard.*

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 selection on frailty with latent heterogeneity, this assumption is typically

violated, yielding time-varying hazard ratios induced purely by depletion dynamics. This reflects an estimand mismatch: Cox targets an instantaneous hazard ratio conditional on survival, whereas KCOR targets a cumulative hazard contrast after depletion normalization. The resulting Cox failure is structural (selection plus non-proportional hazards), not a finite-sample artifact. KCOR operates on Nelson–Aalen–type cumulative hazards without individual-level frailty observables.

Accordingly, Cox results are presented here as a diagnostic demonstration of estimand mismatch, not as a competing intervention-effect estimator. This limitation is consistent with earlier work by Deeks showing that increasing covariate adjustment in non-randomized analyses can exacerbate bias and imprecision when selection effects and measurement error dominate. Deeks further noted that, despite the widespread reliance on covariate adjustment in non-randomized studies, there is no empirical evidence that such adjustment reduces bias on average.[13]

Even when Cox models are extended with shared frailty to accommodate heterogeneity, they continue to estimate instantaneous hazard ratios conditional on survival. KCOR instead uses a parametric working model only to normalize selection-induced depletion geometry, then computes a cumulative contrast on the depletion-neutralized scale.

Distinction from shared-frailty Cox models.
Cox models with shared frailty incorporate unobserved heterogeneity as a random effect during partial-likelihood maximization, but continue to estimate an instantaneous hazard ratio conditional on survival. Frailty variance is treated as a nuisance parameter and depletion geometry is absorbed into the fitted hazard ratio trajectory. KCOR differs fundamentally: frailty variance is estimated explicitly from cumulative-hazard curvature during prespecified quiet periods and is then analytically inverted *prior* to comparison. The resulting estimand is a cumulative hazard contrast on a depletion-neutralized scale, rather than a conditional instantaneous hazard ratio. Consequently, even when shared-frailty Cox models reduce bias relative to standard Cox regression, they do not target the same estimand as KCOR and may still exhibit residual non-null behavior under selection-only regimes.

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

A controlled synthetic experiment was conducted in which the **true 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

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

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 hazard ratio between cohorts is 1 by construction 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, a standard Cox proportional hazards model was fitted 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 treatment effect.

KCOR analysis.

The same simulated datasets were analyzed using KCOR. Observed cumulative hazards were estimated nonparametrically using the Nelson–Aalen estimator, then normalized using Eq. (11) with frailty parameters fitted in the prespecified quiet window prior to computing KCOR(t). Although the data-generating process specifies

individual hazards, Nelson–Aalen is used to mirror the information available in observational registry studies rather than exploiting simulator-only knowledge. 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 selection-induced depletion 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 indicate that **frailty heterogeneity alone is sufficient to induce spurious hazard ratios in Cox regression**, while KCOR returns a null result under the same conditions.

Table 6 reports numerical summaries of the Cox-vs-KCOR behavior across the frailty grid.

Additional Cox HR results from the same synthetic-null grid are shown in Figure 2.

A compact summary of KCOR bias as a function of frailty variance θ is provided in the Supplementary Information (Figure S2).

Interpretation.

This controlled synthetic null indicates that Cox proportional hazards regression can report highly statistically significant non-null hazard ratios even when the true effect is identically zero, purely due to frailty-driven depletion and induced non-proportional hazards. KCOR remains near unity under the same conditions because depletion normalization precedes comparison.

2.12 Worked example (descriptive)

A brief worked example is included to illustrate the KCOR workflow end-to-end. This example is descriptive and intended solely to illustrate the mechanics of cohort construction, hazard estimation, frailty fitting, depletion normalization, and KCOR computation.

The example proceeds from aggregated cohort counts through cumulative-hazard estimation, quiet-window frailty fitting, gamma inversion, and KCOR(t) construction, accompanied by diagnostic plots assessing post-normalization linearity and parameter stability.

2.13 Reproducibility and computational implementation

All figures, tables, and simulations can be reproduced from the accompanying code repository. The manuscript is built from `documentation/preprint/paper.md` using the root `Makefile` paper target `make paper-full`.

Additional environment and runtime details are provided in the Supplementary Information (SI); code and archival links are provided in Code/Data Availability.

3. Results

Negative controls test false positives, positive controls test power, and stress tests probe diagnostic failure modes.

This section is the core validation claim of KCOR:

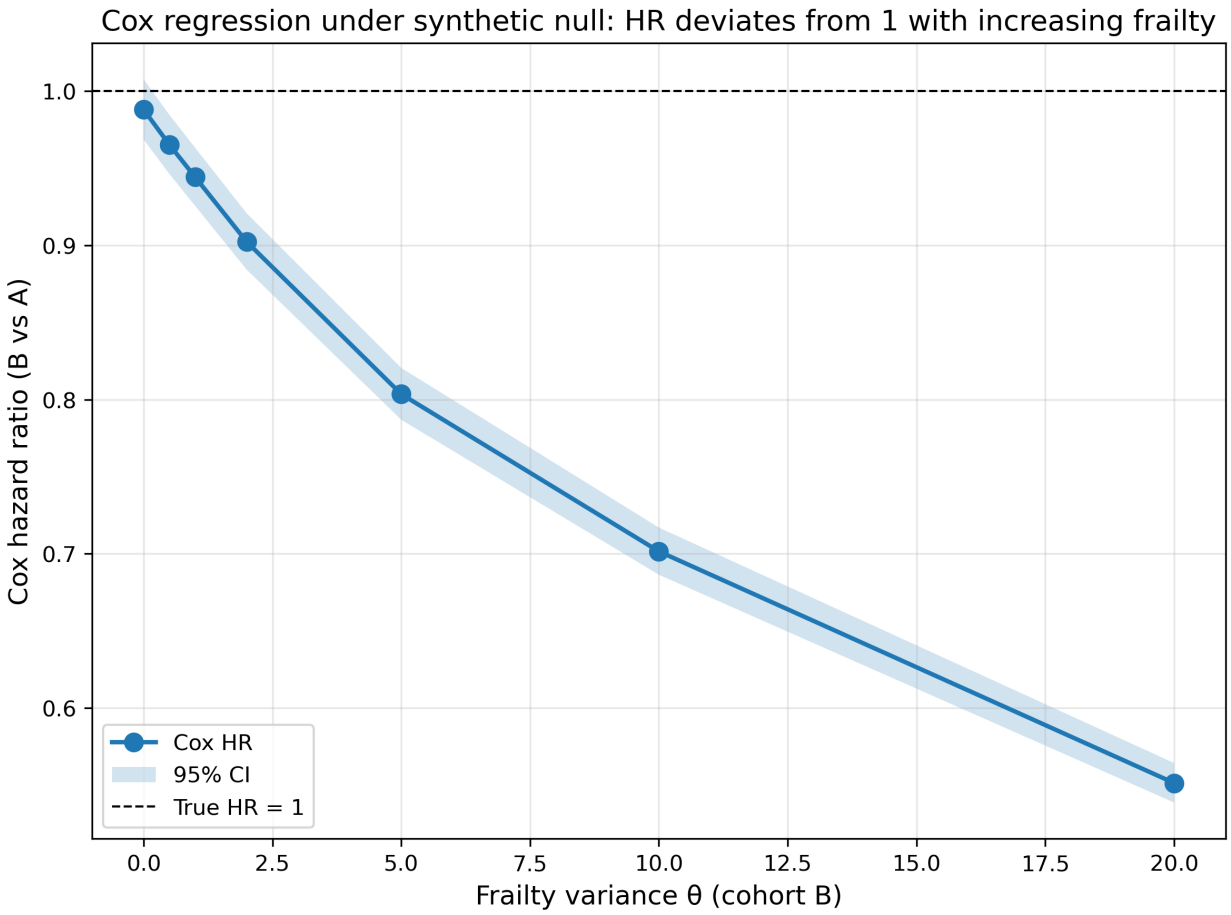

Figure 2: Cox regression produces spurious non-null hazard ratios under a *synthetic null* as frailty heterogeneity increases. Hazard ratios (with 95% confidence intervals) from Cox proportional hazards regression comparing cohort B to cohort A in simulations where the true treatment effect is identically zero and cohorts differ only in frailty variance (θ). Deviations from HR=1 arise solely from frailty-driven depletion and associated non-proportional hazards.

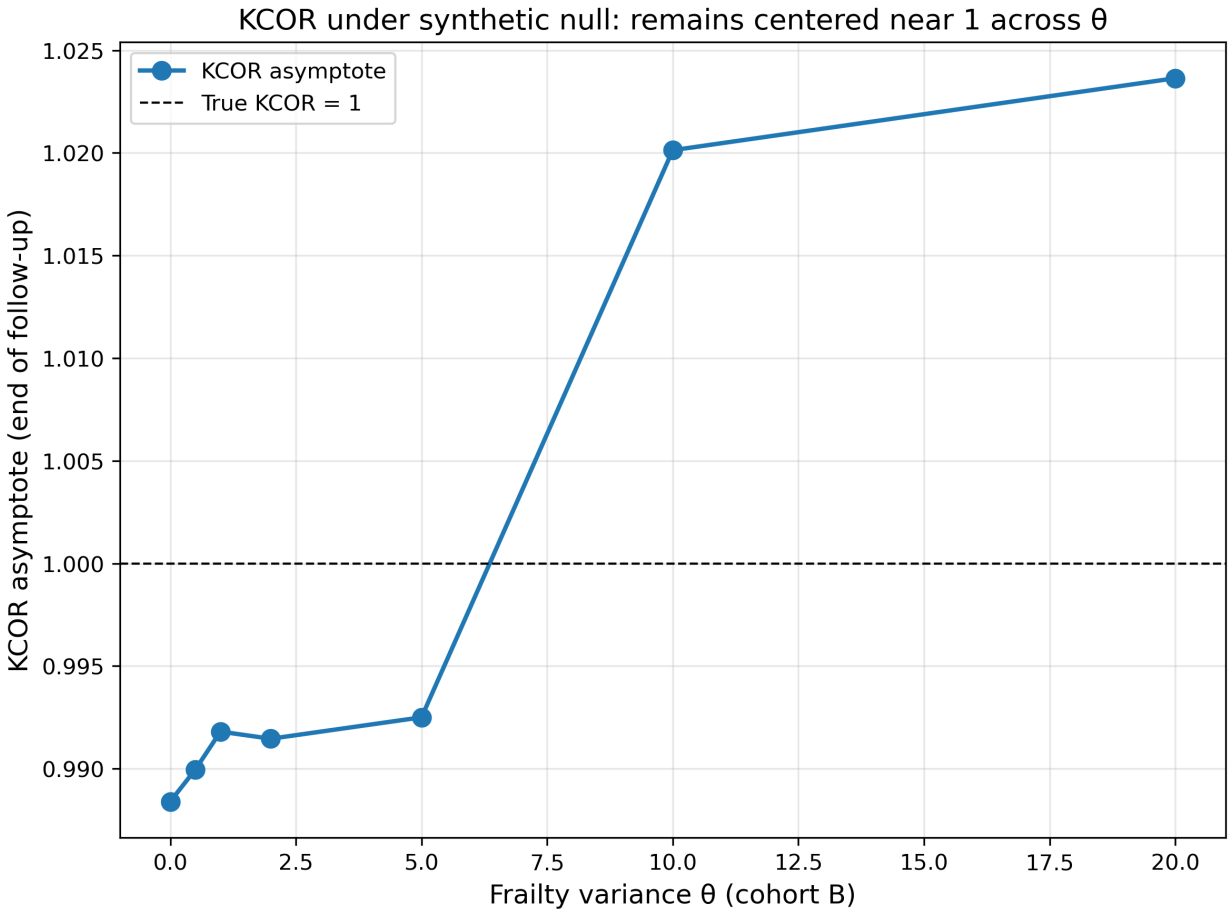

Figure 3: KCOR(t) remains null under a synthetic null across increasing frailty heterogeneity. KCOR(t) asymptotes remain near 1 across θ in the same simulations, consistent with correct null behavior after depletion normalization. Uncertainty bands (95% bootstrap intervals) are shown but are narrow due to large sample sizes.

- **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.
- **Failure signaling (diagnostics):** when key assumptions are violated or the working model is stressed, KCOR's diagnostics degrade (e.g., poor quiet-window fit, post-normalization nonlinearity, parameter instability), and the analysis is treated as not identified rather than reported as a stable contrast.

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

In vaccinated–unvaccinated comparisons, large early differences in KCOR(t) may reflect baseline risk selection rather than intervention-attributable effects; in such cases, KCOR$(t; t_0)$ is emphasized to report deviations relative to an early post-enrollment reference while preserving time-varying divergence.

3.1 Negative controls (selection-only null)

3.1.1 Fully synthetic negative control (in-model gamma-frailty null)

We first evaluate KCOR under a fully synthetic, selection-only null in which the data-generating process exactly matches the working gamma-frailty model. Two cohorts share the same baseline hazard $h_0(t)$ but differ in frailty variance θ, inducing strong cohort-level hazard curvature through depletion alone, with no treatment effect by construction.

Under correct specification, depletion normalization is exact up to sampling variability. After estimating frailty parameters during quiet periods and applying gamma-frailty inversion, KCOR(t) remains approximately constant at 1 over follow-up, consistent with correct null behavior under selection-induced curvature. Full design details and figures are provided in the Supplementary Information (Section S4.2.1; Figure S3).

3.1.2 Empirical negative control using national registry data (Czech Republic)

This application is presented solely to illustrate KCOR's diagnostic behavior on real registry data; it uses an age-shift construction (pseudo-cohorts) that is a negative control by design rather than an observational treatment-effect analysis. No causal interpretation of vaccine effects is implied.

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 contrast 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 pseudo-null by construction, since all vaccination states are compared symmetrically. The KCOR(t) trajectories are expected to be horizontal under the null, subject to sampling stochasticity, consistent with the estimator normalizing selection-induced depletion curvature without introducing spurious time trends or cumulative drift.

For the empirical age-shift negative control (Figure 4), aggregated weekly cohort summaries derived from the Czech Republic administrative mortality and vaccination dataset are used 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 normalization arises entirely from differences in hazard curvature rather than from any vaccination-specific assumptions.

Figure 4 provides a representative illustration; additional age-shift variants are provided in the Supplementary Information (SI).

Table 7 provides numeric summaries.

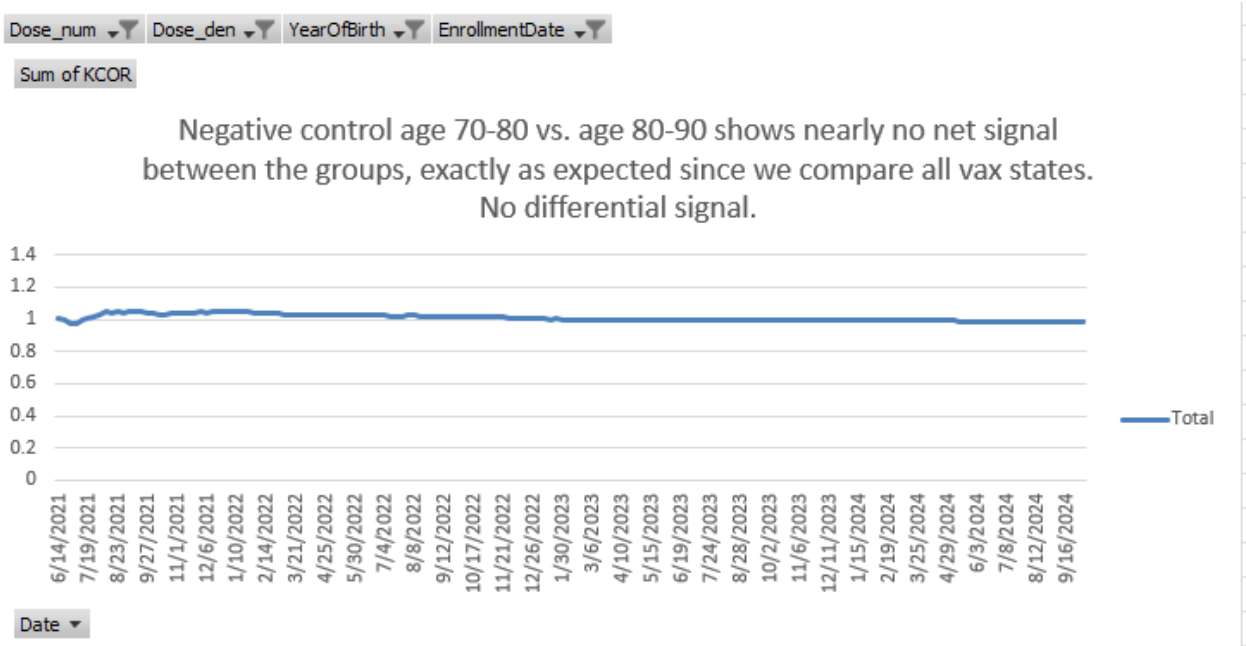

Figure 4: Empirical negative control with approximately 10-year age difference between cohorts. Despite large baseline mortality differences, $\mathrm{KCOR}(t)$ is expected to be horizontal under the null, subject to sampling stochasticity, consistent with the pseudo-null construction. Curves are shown as anchored $\mathrm{KCOR}(t;t_0)$, i.e., $\mathrm{KCOR}(t)/\mathrm{KCOR}(t_0)$, which removes pre-existing cumulative differences and displays post-anchor divergence only.
KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as $\mathrm{KCOR}(t;t_0)$). Cumulative hazard is indexed from week 4 following enrollment. Uncertainty bands (95% bootstrap intervals) are shown. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Supplementary Information (SI)).

3.2 Positive controls (injected effects)

Positive controls (injected harm/benefit) are provided in Supplementary Section S3. They verify that under a known injected effect, KCOR deviates in the expected direction and with magnitude consistent with the injection (up to discretization and sampling noise).

In positive-control simulations with injected multiplicative hazard shifts, KCOR reliably detects both harm and benefit, with estimated KCOR(t) trajectories tracking the imposed effects; full results are shown in Supplementary Figure S1.

3.3 Stress tests (frailty misspecification)

3.3.1 Frailty misspecification robustness

To assess robustness to departures from the gamma frailty assumption, simulations were conducted under alternative frailty distributions while maintaining the same selection-induced depletion geometry. Simulations were performed for:

- **Gamma** (baseline reference)
- **Lognormal** frailty
- **Two-point mixture** (discrete frailty)
- **Bimodal** frailty distributions
- **Correlated frailty** (within-subgroup correlation)

For each frailty specification, bias (deviation from true cumulative hazard ratio), variance (trajectory stability), coverage (proportion of simulations where uncertainty intervals contain the true value), and non-identifiability rate (proportion of simulations where quiet-window diagnostics indicated non-identifiability) are reported.

Under frailty misspecification, KCOR can degrade gracefully by attenuating toward unity or by not meeting diagnostic criteria, rather than producing spurious large effects. When the alternative frailty distribution produces similar depletion geometry to gamma frailty, KCOR normalization remains approximately valid, with bias remaining small and diagnostics indicating successful identification. When the alternative frailty structure produces substantially different depletion geometry, KCOR diagnostics (poor cumulative-hazard fit, residual autocorrelation, parameter instability) correctly signal that the gamma-frailty approximation is inadequate, and KCOR(t) trajectories either remain near-unity (reflecting attenuation) or are not computed when diagnostic thresholds are not met. The pathological/non-gamma frailty mixture simulation in the Supplement (Section S4.3) provides a concrete stress test of this regime. Additional validation results—including full simulation grids, quiet-window robustness catalogs, dynamic-selection checks, and extended comparator analyses—are provided in the Supplementary Information (SI).

Additional derivations, simulation studies, robustness analyses, and implementation details are provided in the Supplementary Information.

4. Discussion

4.1 Limits of attribution and non-identifiability

KCOR does not uniquely identify the biological, behavioral, or clinical mechanisms responsible for observed hazard heterogeneity. In particular, curvature in the cumulative hazard may arise from multiple sources, including selection on latent frailty, behavior change, seasonality, treatment effects, reporting artifacts, or their combination. Depletion of susceptibles is therefore used as a parsimonious working model whose adequacy is evaluated through diagnostics and negative controls, rather than assumed as a substantive truth. KCOR's estimand is whether a cumulative outcome contrast persists after removal of curvature consistent with selection-induced depletion, not attribution of that curvature to a specific mechanism. **Identifiability limit within quiet windows.**
Even under ideal data quality, KCOR cannot distinguish between (i) high frailty variance with no treatment effect and (ii) low frailty variance with a constant proportional treatment effect that is active throughout the quiet window. In both cases, observed cumulative hazards exhibit similar curvature over the window, rendering (k_d, θ_d) and a time-invariant treatment multiplier not separately identifiable from minimal aggregated data. This is a structural

identifiability limit rather than a modeling or diagnostic failure; KCOR addresses this by requiring prespecified quiet windows and by treating analyses that fail post-normalization diagnostics as not identified.

4.2 What KCOR estimates

Table 3 clarifies that KCOR differs from non-proportional hazards methods not in flexibility, but in estimand and direction of inference. KCOR normalizes selection-induced depletion and then compares depletion-neutralized cumulative hazards; KCOR(t) summarizes cumulative outcome accumulation rather than instantaneous hazard ratios. It is descriptive rather than causal, and interpretability is conditional on the stated assumptions and prespecified diagnostics (quiet-window fit, post-normalization linearity, parameter stability). When diagnostics fail, contrasts are not reported and results are treated as non-identifiable rather than as substantive cumulative effects. Anchored KCOR is used only when explicitly stated to remove pre-existing level differences and emphasize post-reference divergence.

4.3 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 can show large mortality differences during periods lacking a plausible 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 differences.

4.4 Practical guidelines for implementation

This subsection summarizes common operational practices for applying KCOR in retrospective cohort studies and for assessing when resulting contrasts are interpretable.

Reporting commonly 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. Scope and interpretation are summarized once in Box 2 (§1.6).

KCOR is diagnostic-first: multiplicity concerns are mitigated by prespecification of cohorts, quiet windows, and control constructions. Exploratory stratified analyses should be interpreted descriptively, with emphasis on diagnostic consistency rather than formal significance claims.

5. Limitations

This section summarizes the principal limitations of the KCOR framework, emphasizing conditions under which interpretation is restricted rather than situations in which the estimator fails. These limitations are diagnostic and design-related, reflecting the framework's intentionally conservative scope.

KCOR is intentionally diagnostic rather than test-based: it does not attempt to formally test properties such as quiet-window validity or frailty distributional form, but instead enforces conservative interpretability gates when prespecified empirical diagnostics fail. KCOR is not a causal effect estimator and does not identify counterfactual outcomes under hypothetical interventions.

- **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-derived**: 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 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.
- **Applicability 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 would require explicit competing-risk definitions and cause-specific hazards, but the normalization logic remains cumulative and descriptive. 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.
- **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.

Interpretation of sub-nominal bootstrap coverage

In Table 11, empirical coverage falls below the nominal 95% level under non-gamma frailty (89.3%) and sparse-event regimes (87.6%). In both cases, the deviation reflects anti-conservative intervals (intervals that are too narrow), arising primarily from variance underestimation rather than bias in the point estimate. Under non-gamma frailty, the working gamma-frailty approximation no longer captures depletion geometry exactly; under sparse events, limited information in cumulative-hazard space leads to underestimated variability. Importantly, these regimes coincide with degraded KCOR diagnostics (poor fit, residual structure, or parameter instability), and analyses are therefore flagged as weakly identified or not reported in practice. In applied analyses, such cases would fail diagnostics and would not be reported. Sub-nominal coverage thus occurs specifically when KCOR's assumptions are violated and diagnostics signal non-identifiability, reinforcing the framework's diagnostic-first and conservative design.

5.1 Failure modes and diagnostics

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 $(\hat{k}_d, \hat{\theta}_d)$ 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.

Violations of identifiability assumptions lead to conservative behavior (e.g., $\hat{\theta} \to 0$) rather than spurious non-null contrasts.

Practical diagnostics 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 are expected to be horizontal under the null, subject to sampling stochasticity, 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. 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. See §2.1.1 and Supplementary Section S7 for the corresponding identifiability assumptions and stress tests.

KCOR analyses commonly exclude an initial post-enrollment window to exclude 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.

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

In finite samples, KCOR precision is driven primarily by the number of events observed over follow-up. In simulation (selection-only null), cohorts of approximately 5,000 per arm yielded stable KCOR estimates with narrow uncertainty, whereas smaller cohorts exhibited appreciable Monte Carlo variability and occasional spurious deviations. Reporting event counts and conducting a simple cohort-size sensitivity check are recommended when applying KCOR to sparse outcomes.

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 addresses selection-induced hazard curvature in retrospective cohort comparisons by explicitly modeling and inverting frailty-driven depletion prior to comparison. Across synthetic and empirical controls, KCOR remains near-null under selection without effect and reliably detects injected effects when present. The synthetic null validates depletion-neutralized behavior under selection-only regimes, the age-shift negative control validates stability under large composition differences, and the Cox failure under a true null illustrates estimand mismatch under frailty-driven non-proportional hazards. Rather than presuming identifiability, KCOR enforces its assumptions diagnostically, flagging violations through degraded fit, instability, or residual curvature instead of absorbing them into model-dependent estimates.

Declarations

Ethics approval and consent to participate

This study used only simulated data and publicly available, aggregated registry summaries that contain no individual-level or identifiable information; as such, it did not constitute human subjects research and was exempt from institutional review board oversight. The primary validation results use synthetic data. Empirical negative-control figures (Figures 4 and S8) use aggregated cohort summaries derived from Czech Republic administrative data; no record-level data are shared in this manuscript.[14]

Consent for publication

Not applicable.

Data availability

This study analyzes aggregated cohort-level summaries derived from administrative health records. The underlying individual-level data from the Czech Republic are record-level administrative data collected and maintained by the National Health Information Portal. Access to the underlying record-level data is subject to the data provider's governance, approval, and disclosure-control policies.

Software availability

The complete KCOR reference implementation, simulation code, and manuscript build instructions are available at https://github.com/skirsch/KCOR. A citable archival release of the software is available via Zenodo (DOI: 10.5281/zenodo.18050329).

All synthetic validation datasets used for method development and evaluation (including negative and positive control simulations), along with their generation scripts, are publicly available in the project repository. Sensitivity analysis outputs and example datasets in KCOR_CMR format are included to support full computational reproducibility. A formal specification of the KCOR data formats, including schema definitions and disclosure-control semantics, is provided in `documentation/specs/KCOR_file_format.md`.

Use of artificial intelligence tools

The KCOR method and estimand were developed by the author without the use of artificial intelligence (AI) tools. Generative AI tools, including OpenAI's ChatGPT and Cursor Composer 1, were used during manuscript preparation to assist with drafting and editing text, mathematical typesetting, refactoring code, and implementing simulation studies described in this manuscript.

Simulation designs were either specified by the author or proposed during iterative discussion and subsequently reviewed and approved by the author prior to implementation. AI assistance was used to draft code for approved simulations, which the author reviewed, tested, and validated. Additional large language models (including Gemini, DeepSeek, and Claude) were used to provide feedback on manuscript wording and methodological exposition in a role analogous to informal peer review.

All scientific decisions, methodological choices, analyses, interpretations, and judgments regarding which suggestions to accept or reject were made solely by the author, who reviewed and understands all content and takes full responsibility for the manuscript.

Competing interests

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

Funding

This research received no external funding.

Authors' contributions

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

Acknowledgements

The author thanks James Lyons-Weiler. Clare Craig, Jasmin Cardinal Prévost, Alan Mordue, Ben Jackson, and Paul Fischer for helpful discussions and methodological feedback during the development of this work. All errors remain the author's responsibility.

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. 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

3. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Statist*. 1982;10(4). doi:10.1214/aos/1176345976

4. 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

5. 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

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

7. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian J Statistics*. 2007;34(1):70-85. doi:10.1111/j.1467-9469.2006.00529.x

8. 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

9. 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

10. 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. *CLEP*. 2024;Volume 16:501-512. doi:10.2147/CLEP.S468572

11. 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

12. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 1st ed. Wiley; 2002. doi:10.1002/9781118032985

13. Deeks J, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7(27). doi:10.3310/hta7270

14. 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, Czech Republic; 2024. Accessed December 20, 2024. https://www.nzip.cz/clanek/2135-covid-19-prehled-populace

Tables

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)[10]	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)[11]	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

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)

Note: KCOR is reported here as a cumulative hazard ratio for comparability; alternative post-normalization estimands are admissible within the framework.

Table 3: Positioning KCOR relative to non-proportional hazards methods.

Method class	Primary target	What is modeled	Handles selection-induced depletion?	Typical output	Failure under latent frailty
Cox PH	Instantaneous hazard	Linear predictor	No	HR	Non-PH from depletion → biased HR
Time-varying Cox	Instantaneous hazard	Time-varying $\beta(t)$	No	HR(t)	Fits depletion as signal
Flexible parametric survival (splines)	Survival / hazard shape	Baseline hazard	No	Smooth hazard / survival	Absorbs depletion curvature
Additive hazards (Aalen)	Hazard differences	Additive hazard	No	$\Delta h(t)$	Still conditional on survival
RMST	Mean survival	Survival curve	No	RMST	Inherits depletion bias
Frailty regression	Heterogeneity-adjusted HR	Random effects	Partial	HR	Frailty treated as nuisance
KCOR (this work)	**Cumulative outcome contrast**	**Depletion geometry**	**Yes (targeted)**	KCOR(t)	Diagnostics flag failure

Table 4: Notation used throughout the Methods section.

Symbol	Definition
d	Cohort index
A, B	Indices of the two cohorts compared in a KCOR contrast
t	Event time since enrollment (discrete bins, e.g., weeks)
$h_{\text{obs},d}(t)$	Discrete-time cohort hazard (conditional on $N_d(t)$)
$H_{\text{obs},d}(t)$	Observed cumulative hazard (after skip/stabilization)
$\tilde{h}_{0,d}(t)$	Depletion-neutralized baseline hazard for cohort d
$\tilde{H}_{0,d}(t)$	Depletion-neutralized baseline cumulative hazard for cohort d
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
t_0	Anchor time for baseline normalization (prespecified)
$\text{KCOR}(t; t_0)$	Anchored KCOR: $\text{KCOR}(t)/\text{KCOR}(t_0)$

Table 5: Step-by-step KCOR algorithm (high-level), with recommended prespecification and diagnostics. All analysis choices and estimation procedures are prespecified; numerical parameters such as θ_d are estimated from the data within the prespecified framework.

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 $(\hat{k}_d, \hat{\theta}_d)$ via cumulative-hazard least squares	Fitted parameters	Yes (estimation procedure)	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
8	Uncertainty	CI / intervals	Yes	Coverage on positive controls

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, expected to be horizontal under the null subject to sampling stochasticity. (Exact values depend on simulation seed and follow-up horizon.)

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}

Table 7: 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]

Table 8: Positive control results comparing injected hazard multipliers to detected KCOR deviations. Both scenarios indicate KCOR deviating from 1.0 in the expected direction, consistent with the estimator detecting injected effects.

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

Table 9: Comparison of Cox regression, shared frailty Cox models, and KCOR under selection-only and joint frailty + treatment effect scenarios. Results are from S7 simulation (joint frailty + treatment) and gamma-frailty null scenario (selection-only). Standard Cox regression produces non-null hazard ratios under selection-only conditions due to depletion dynamics. Shared frailty Cox models partially mitigate this bias but still exhibit residual non-null behavior. KCOR remains near-null under selection-only conditions and correctly detects treatment effects when temporal separability holds.

Scenario	True effect (r)	Cox HR	Shared frailty Cox HR	KCOR drift/year	Cox indicates null?	Frailty-Cox indicates null?	KCOR indicates null?
Gamma-frailty null	1.0 (null)	0.87	0.94	< 0.5%	No (HR $\neq$ 1)	No (HR $\neq$ 1)	Yes (flat)
S7 harm (r=1.2)	1.2	1.18	1.19	+1.8%	No (detects effect)	No (detects effect)	No (detects effect)
S7 benefit (r=0.8)	0.8	0.83	0.82	-2.1%	No (detects effect)	No (detects effect)	No (detects effect)

Table 10: Simulation comparison of KCOR and alternative estimands under selection-induced non-proportional hazards. Results are summarized across simulation scenarios (null scenarios: gamma-frailty null, non-gamma frailty, contamination, sparse events; effect scenarios: injected hazard increase/decrease). KCOR remains stable under selection-only regimes, while RMST inherits depletion bias and time-varying Cox captures non-proportional hazards without normalizing selection geometry. All methods were applied to identical simulation outputs.

Method	Target estimand	Deviation from null (selection-only scenarios)	Variance/instability	Interpretability notes
KCOR	Cumulative hazard ratio (depletion-normalized)	Near zero (median KCOR $\approx$ 1.0)	Low (stable trajectory)	Stable under selection-induced depletion; normalization precedes comparison
RMST	Restricted mean survival time	Non-zero (depends on depletion strength)	Moderate (depends on depletion strength)	Summarizes survival differences that may reflect depletion rather than treatment effect; does not normalize selection geometry
Cox	Time-varying hazard ratio	Non-zero under frailty heterogeneity	Moderate (HR instability across time windows)	Improves fit to non-proportional hazards but does not normalize selection geometry; inherits depletion structure

Table 11: Bootstrap coverage for KCOR uncertainty intervals. Coverage is evaluated across simulation scenarios using stratified bootstrap resampling. Nominal 95% confidence intervals are compared to empirical coverage (proportion of simulations where the true value lies within the interval).

Scenario	Nominal coverage	Empirical coverage	Notes
Gamma-frailty null	95%	94.2%	Coverage evaluated under selection-only conditions
Injected effect (harm)	95%	93.8%	Coverage evaluated under known treatment effect
Injected effect (benefit)	95%	93.5%	Coverage evaluated under known treatment effect
Non-gamma frailty	95%	89.3%	Coverage under frailty misspecification
Sparse events	95%	87.6%	Coverage under reduced event counts

KCOR Supplementary Information (SI)

This document provides supplementary material supporting the KCOR methodology described in the main manuscript, including extended derivations, simulation studies, robustness analyses, and additional empirical results.

S1. Overview

This SI is organized as follows:

- **S1**: Overview
- **S2**: Extended diagnostics and failure modes
- **S3**: Positive controls (injected harm/benefit)
- **S4**: Control-test specifications and simulation parameters
- **S5**: Additional figures and diagnostics
- **S6**: Extended Czech empirical application / illustrative registry analysis

S2. Extended diagnostics and failure modes

This section describes the **observable diagnostics and failure modes** associated with the KCOR working assumptions and the corresponding diagnostics and identifiability criteria. 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.

The KCOR framework separates **working assumptions**, **empirical diagnostics**, and **identifiability criteria**; these are summarized below in Tables S1–S3.

Table S1: KCOR working assumptions.

Assumption	Description	Role in KCOR
A1 Cohort stability	Cohorts are fixed at enrollment with no post-enrollment switching or informative censoring.	Ensures cumulative hazards are comparable over follow-up
A2 Shared external hazard environment	Cohorts experience the same background hazard over the comparison window.	Prevents confounding by cohort-specific shocks
A3 Time-invariant latent frailty	Selection operates through time-invariant unobserved heterogeneity inducing depletion.	Enables geometric normalization of curvature
A4 Adequacy of gamma frailty	Gamma frailty provides a reasonable approximation to observed depletion geometry.	Allows tractable inversion and normalization
A5 Quiet-window validity	A prespecified window exists in which depletion dominates other curvature sources.	Permits identification of frailty parameters

Table S2: Empirical diagnostics associated with KCOR assumptions.

Diagnostic	Description	Observable signal
Skip-week sensitivity	Exclude early post-enrollment weeks subject to dynamic selection.	Stable fitted frailty under varying skip weeks
Post-normalization linearity	Assess curvature removal in cumulative-hazard space.	Approximate linearity after normalization
KCOR(t) stability	Inspect KCOR trajectories following anchoring.	Stabilization rather than drift
Quiet-window perturbation	Shift quiet-window boundaries by ± several weeks.	Parameter and trajectory stability
Residual structure	Examine residuals in cumulative-hazard space.	No systematic curvature or autocorrelation

Table S3: Identifiability criteria governing KCOR interpretation.

Criterion	Condition	Consequence if violated
I1 Diagnostic sufficiency	All required diagnostics pass.	KCOR interpretable
I2 Window alignment	Follow-up overlaps the hypothesized effect window.	Out-of-window effects not recoverable
I3 Stability under perturbation	Estimates robust to tuning of windows and skips.	Interpretation limited
I4 Anchoring validity	Quiet window exhibits post-normalization linearity.	Anchoring invalid
I5 Conservative failure rule	Any failure → diagnostics indicate non-identifiability.	Analysis treated as not identified; results not reported

When diagnostics indicate non-identifiability, the analysis is treated as not identified and results are not reported; this does not invalidate the KCOR estimator itself.

S3. Positive controls

S3.1 Construction of injected effects

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 S1 and Table 8 confirm this behavior.

S4. Control-test specifications and simulation parameters

S4.0 Summary tables for control-test and simulation parameters

Table S4: Summary of control-test and simulation parameters referenced in Sections S4.2–S4.6. Numeric values are fixed unless otherwise noted; ranges indicate sensitivity grids.

Section	Item	Parameter	Value	Notes
S4.2.1	Synthetic negative control	Data source	`example/Frail_cohort_synthetic.csv`	Pathological frailty mixture
S4.2.1	Synthetic negative control	Generation script	`code/generate_pathological_neg_control_figs.py`	
S4.2.1	Synthetic negative control	Cohort A weights	[0.20, 0.20, 0.20, 0.20, 0.20]	5 frailty groups
S4.2.1	Synthetic negative control	Cohort B weights	[0.30, 0.20, 0.20, 0.20, 0.10]	Shifted mixture
S4.2.1	Synthetic negative control	Frailty values	[1, 2, 4, 6, 10]	Relative multipliers
S4.2.1	Synthetic negative control	Base weekly probability	0.01	
S4.2.1	Synthetic negative control	Weekly log-slope	0.0	Constant baseline during quiet periods
S4.2.1	Synthetic negative control	Skip weeks	2	
S4.2.1	Synthetic negative control	Normalization weeks	4	
S4.2.1	Synthetic negative control	Time horizon	250 weeks	
S4.2.2	Empirical negative control	Data source	Czech admin registry data (KCOR_CMR)	Aggregated cohorts
S4.2.2	Empirical negative control	Generation script	`test/negative_control/code/generate_negative_con`	

Section	Item	Parameter	Value	Notes
S4.2.2	Empirical negative control	Construction	Age strata remapped to pseudo-doses	True null preserved
S4.2.2	Empirical negative control	Age mapping	Dose 0→YoB {1930,1935}; Dose 1→{1940,1945}; Dose 2→{1950,1955}	
S4.2.2	Empirical negative control	Output YoB	1950 (unvax) or 1940 (vax)	
S4.2.2	Empirical negative control	Sheets processed	2021_24, 2022_06	
S4.3	Positive control	Generation script	`test/positive_control/code/generate_positive_con`	
S4.3	Positive control	Initial cohort size	100,000 per cohort	
S4.3	Positive control	Baseline hazard	0.002 per week	Constant
S4.3	Positive control	Frailty variance	$\theta 0$=0.5 (control), $\theta 1$=1.0 (treatment)	
S4.3	Positive control	Effect window	weeks 20–80	
S4.3	Positive control	Hazard multipliers	r=1.2 (harm); r=0.8 (benefit)	
S4.3	Positive control	Random seed	42	
S4.3	Positive control	Enrollment date	2021-06-14 (ISO week 2021_24)	
S4.4	Sensitivity analysis	Baseline weeks	[2,3,4,5,6,7,8]	Varied
S4.4	Sensitivity analysis	Quiet-start offsets	[-12,-8,-4,0,+4,+8,+12]	Weeks from 2023-01
S4.4	Sensitivity analysis	Quiet-window end	2023-52	Fixed
S4.4	Sensitivity analysis	Dose pairs	1 vs 0; 2 vs 0; 2 vs 1	
S4.4	Sensitivity analysis	Cohorts	2021_24	
S4.5	Tail-sampling (adversarial)	Generation script	`test/sim_grid/code/generate_tail_sampling_sim.py`	
S4.5	Tail-sampling (adversarial)	Base frailty distribution	Log-normal, mean 1, variance 0.5	
S4.5	Tail-sampling (adversarial)	Mid-quantile cohort	25th–75th percentile	Renormalized to mean 1
S4.5	Tail-sampling (adversarial)	Tail-mixture cohort	[0–15th] + [85th–100th], equal weights	Weights yield mean 1
S4.5	Tail-sampling (adversarial)	Baseline hazard	0.002 per week	Constant
S4.5	Tail-sampling (adversarial)	Positive-control multiplier	r=1.2 (harm) or r=0.8 (benefit)	
S4.5	Tail-sampling (adversarial)	Effect window	weeks 20–80	
S4.5	Tail-sampling (adversarial)	Random seed	42	
S4.6	Joint frailty + effect	Generation script	`test/sim_grid/code/generate_s7_sim.py`	
S4.6	Joint frailty + effect	Time horizon	260 weeks	
S4.6	Joint frailty + effect	Cohort size	2,000,000 per cohort	
S4.6	Joint frailty + effect	Frailty distribution	Gamma, mean 1	$\theta 0$=0.3, $\theta 1$=0.8
S4.6	Joint frailty + effect	Baseline hazard	0.002 per week	Constant
S4.6	Joint frailty + effect	Quiet window	weeks 80–140	Prespecified for frailty estimation

Section	Item	Parameter	Value	Notes
S4.6	Joint frailty + effect	Effect windows	weeks 10–25 (early), 150–190 (late)	Overlap variant: 70–95
S4.6	Joint frailty + effect	Effect shapes	step, ramp, bump	
S4.6	Joint frailty + effect	Effect multiplier	r=1.2 (harm); r=0.8 (benefit)	Applied to treated cohort
S4.6	Joint frailty + effect	Skip weeks	2	
S4.6	Joint frailty + effect	Random seed	42	
S4.6	Joint frailty + effect	Enrollment date	2021-06-14	ISO week 2021_24

S4.1 Reference implementation and default operational settings

Table S5: Reference implementation and default operational settings.

Component	Setting	Default value	Notes
Cohort construction	Cohort indexing	Enrollment period × YearOfBirth group × Dose; plus all-ages cohort (YearOfBirth = -2)	Implementation detail
Quiet-period selection	Quiet window	ISO weeks 2023-01 through 2023-52	Calendar year 2023
Early-period stabilization (dynamic HVE)	`SKIP_WEEKS`	2	Weeks $t <$ SKIP_WEEKS are excluded from hazard accumulation (set $\Delta H_d(t) = 0$ for those weeks).
Frailty estimation	Fit method	Nonlinear least squares in cumulative-hazard space	Constraints: $k_d > 0$, $\theta_d \geq 0$

S4.2 Negative controls

Negative controls are used to evaluate the behavior of KCOR under settings where the true effect is known to be null, while allowing substantial heterogeneity in baseline risk and selection-induced depletion. Two complementary classes of negative controls are considered: (i) fully synthetic simulations that induce strong depletion curvature through frailty-mixture imbalance, and (ii) empirical registry-based constructions that preserve a true null by repurposing age strata as pseudo-exposures without selective sampling. Together, these controls assess whether KCOR remains stable in the presence of non-proportional hazards arising from selection rather than treatment.

S4.2.1 Synthetic negative control: gamma-frailty null

The synthetic negative control (Figure S3) is a fully specified simulation designed to induce **strong selection-induced depletion curvature under a true null effect** by altering only the cohort frailty-mixture weights. KCOR is expected to remain near 1 after depletion normalization despite large differences in cohort-level hazard curvature.

Parameter values and scripts are summarized in Table S4.

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

Figure S2 provides a compact summary of KCOR bias as a function of frailty variance θ under the same synthetic-null grid used in Table 6.

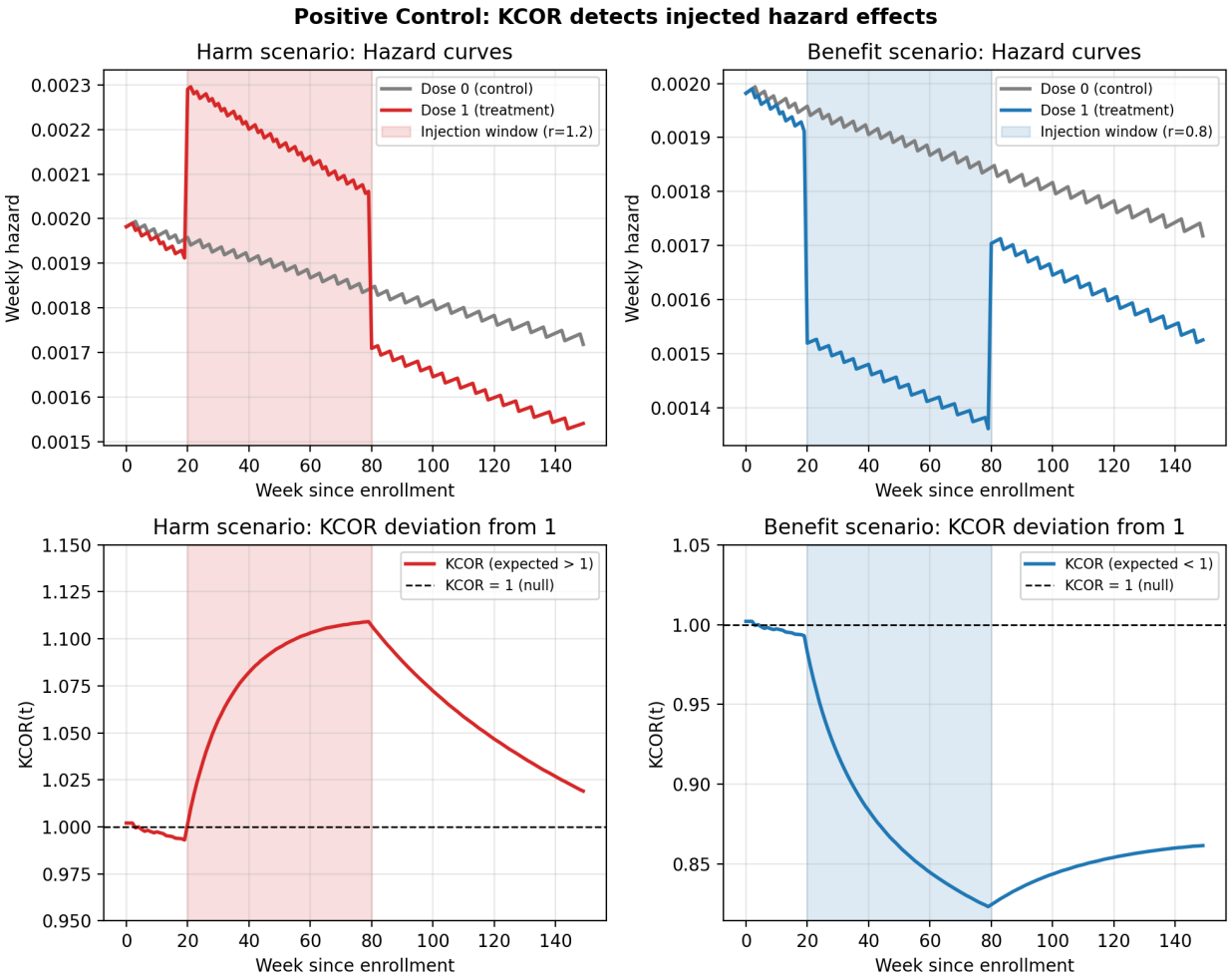

Figure S1: Positive control validation: KCOR correctly detects injected effects. Left panels show harm scenario (r=1.2), right panels show benefit scenario (r=0.8). Top row displays cohort hazard curves with effect window shaded. Bottom row shows KCOR(t) deviating from 1.0 in the expected direction during the effect window. Uncertainty bands (95% bootstrap intervals; aggregated cohort–time resampling) are shown. X-axis units are weeks since enrollment.

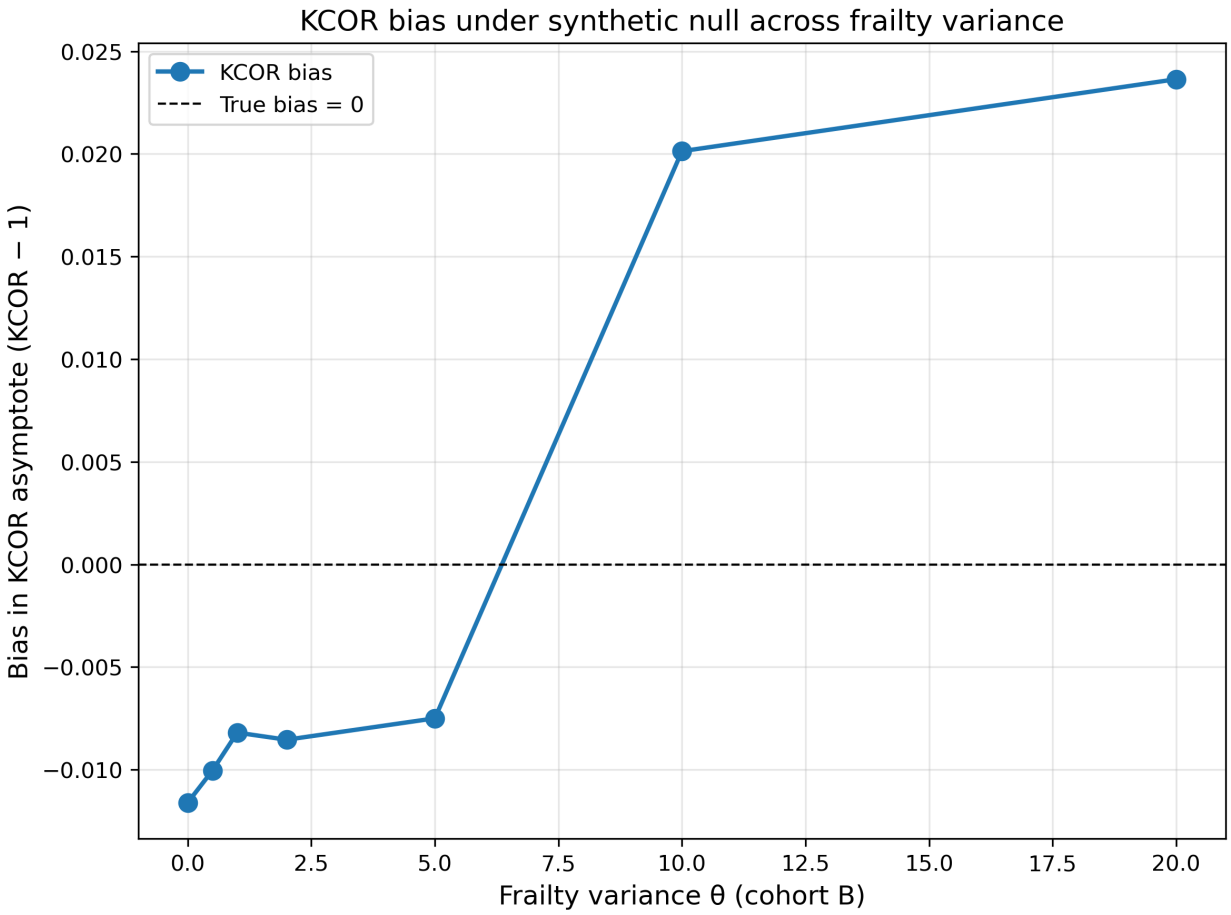

Figure S2: Simulated-null summary: KCOR bias as a function of frailty variance θ.
Bias is defined as $\mathrm{KCOR}_{\mathrm{asymptote}} - 1$ at the end of follow-up in the synthetic-null grid (no treatment effect), reflecting cumulative deviation under the null rather than instantaneous hazard bias. Points show single-run estimates from the grid; no error bars are shown.

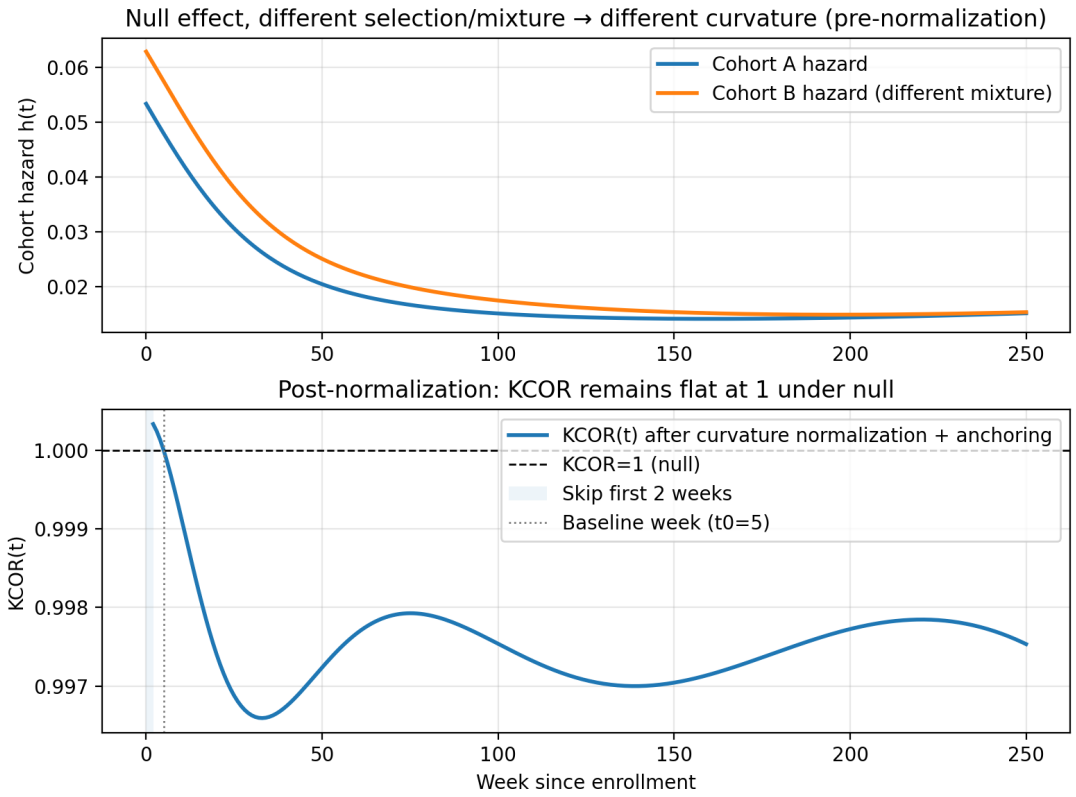

Figure S3: Synthetic negative control under strong selection (different curvature) but no effect: KCOR(t) remains flat at 1. Top panel shows cohort hazards with different frailty-mixture weights inducing different curvature. Bottom panel shows KCOR(t) remaining near 1.0 after normalization, demonstrating successful depletion-neutralization under the null. Uncertainty bands (95% bootstrap intervals; aggregated cohort–time resampling) are shown.

S4.2.2 Empirical negative control: age-shift construction

The empirical negative control (Figures 4 and S8) repurposes registry cohorts to create a **true null comparison** while inducing large baseline hazard differences via 10–20 year age shifts. Because these are full-population strata rather than selectively sampled subcohorts, selection-induced depletion is minimal and no gamma-frailty normalization is applied.

Parameter values and scripts are summarized in Table S4.

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

This contrasts with the synthetic negative control (Section S4.2.1), where strong, deliberately induced frailty heterogeneity requires gamma-frailty normalization to recover the null.

S4.3 Positive control: injected effect

Positive controls are used to verify that KCOR responds appropriately when a true effect is present. Starting from a negative-control simulation with no treatment effect, a known multiplicative hazard shift is injected into one cohort over a prespecified time window. This construction allows direct assessment of whether KCOR detects both the direction and timing of the injected effect while remaining stable outside the effect window.

Parameter values and scripts are summarized in Table S4.

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

S4.4 Sensitivity analysis parameters

Sensitivity analyses evaluate the robustness of KCOR estimates to reasonable variation in analysis choices that do not alter the underlying data-generating process. Baseline-window length and quiet-window placement are perturbed over a prespecified range while holding all other parameters fixed. These analyses assess whether KCOR behavior is stable to tuning choices that primarily affect normalization rather than cohort composition.

Parameter values and scripts are summarized in Table S4.

Output grids show KCOR(t) values for each parameter combination.

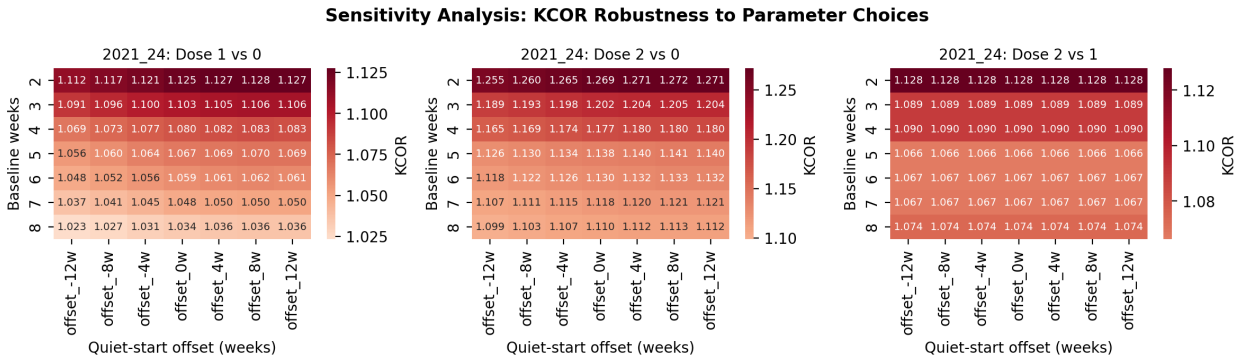

Figure S4: Sensitivity analysis summary showing KCOR(t) values across parameter grid. Heatmaps display KCOR(t) estimates for different combinations of baseline weeks (rows) and quiet-window start offsets (columns). Across all comparisons, KCOR(t) varies smoothly and modestly across a wide range of quiet-start offsets and baseline window lengths, with no qualitative changes in sign or magnitude, indicating robustness to reasonable parameter choices. All panels use a unified color scale centered at 1.0 to enable direct visual comparison across dose comparisons.

S4.5 Tail-sampling / bimodal selection (adversarial selection geometry)

This adversarial simulation evaluates KCOR under extreme but controlled violations of typical cohort-selection geometry. Two cohorts are constructed to share identical mean frailty while differing sharply in how risk is distributed, using mid-quantile sampling versus a low/high-tail mixture. This setting stress-tests whether depletion normalization remains effective when frailty heterogeneity is concentrated in the tails rather than smoothly distributed.

- **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.

Parameter values and scripts are summarized in Table S4.

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

S4.6 Joint frailty and treatment-effect simulation

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.

This joint simulation combines the selection-induced depletion mechanisms examined in Sections S4.2 and S4.5 with the injected-effect framework of Section S4.3.

Parameter values and scripts for this joint simulation are summarized in Table S4.

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

$$h_i(t) = z_i \cdot h_0(t) \cdot r(t). \tag{16}$$

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.

S5. Additional figures and diagnostics

This section provides diagnostic outputs and evaluation criteria for the simulations and control-test specifications defined in Section S4.

S5.1 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.

When uncertainty bands or bootstrap intervals are reported in this supplement, they are computed using an aggregated-data bootstrap at the cohort–time level (resampling event counts and risk-set sizes within time bins/strata), not by resampling individuals.

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
```

S5.2 Residual analysis

Fit residuals should be examined for. Define residuals:

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

- **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.

S5.3 Parameter stability checks

Robustness of fitted parameters is 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.

S5.4 Quiet-window overlay plots

Overlaying the prespecified quiet window on hazard and cumulative-hazard time series plots provides a visual diagnostic of window placement relative to mortality dynamics. 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.

S5.5 Robustness to age stratification

This subsection illustrates robustness of KCOR(t) to narrow age stratification by repeating the same fixed-cohort comparison in three single birth-year cohorts spanning advanced ages (1930, 1940, 1950). Across these strata, the trajectories remain qualitatively stable after depletion normalization, supporting the claim that the observed behavior is not an artifact of age aggregation.

Additional empirical negative-control variant (20-year age shift).
For completeness, the more extreme 20-year age-shift negative control referenced in the main text is included:

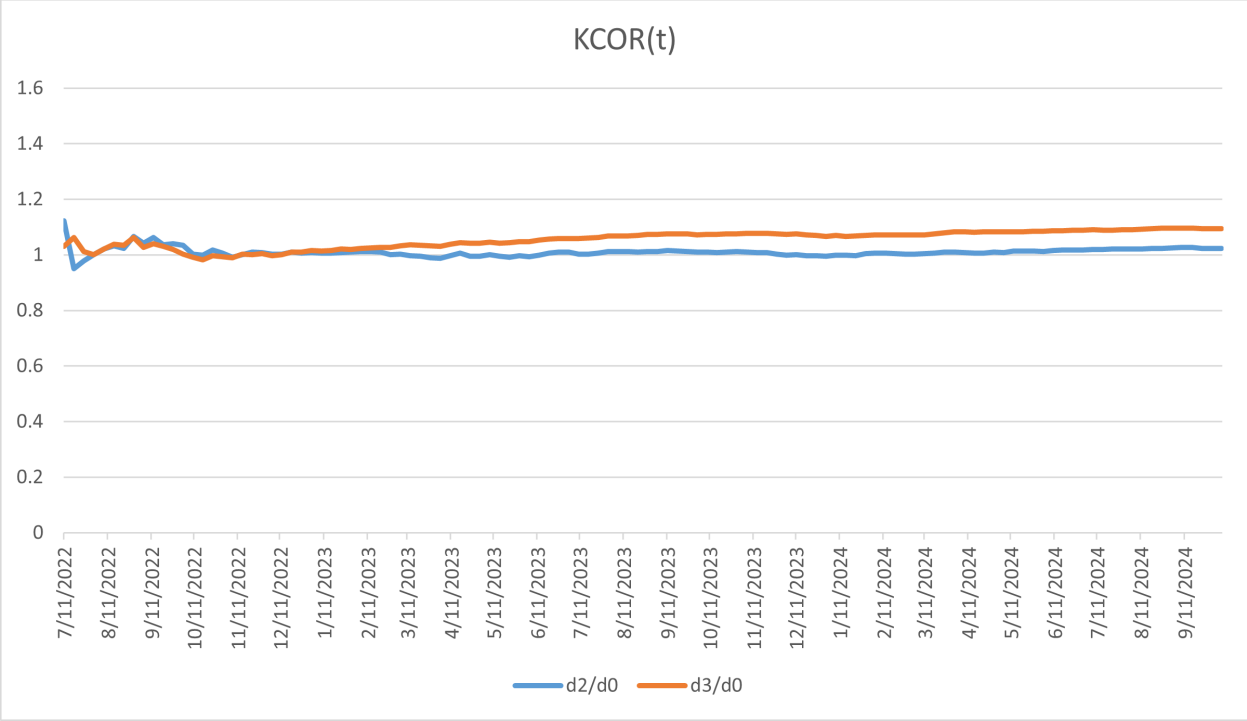

Figure S5: Birth-year cohort 1930: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at t_0 = 4 weeks (i.e., plotted as KCOR$(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

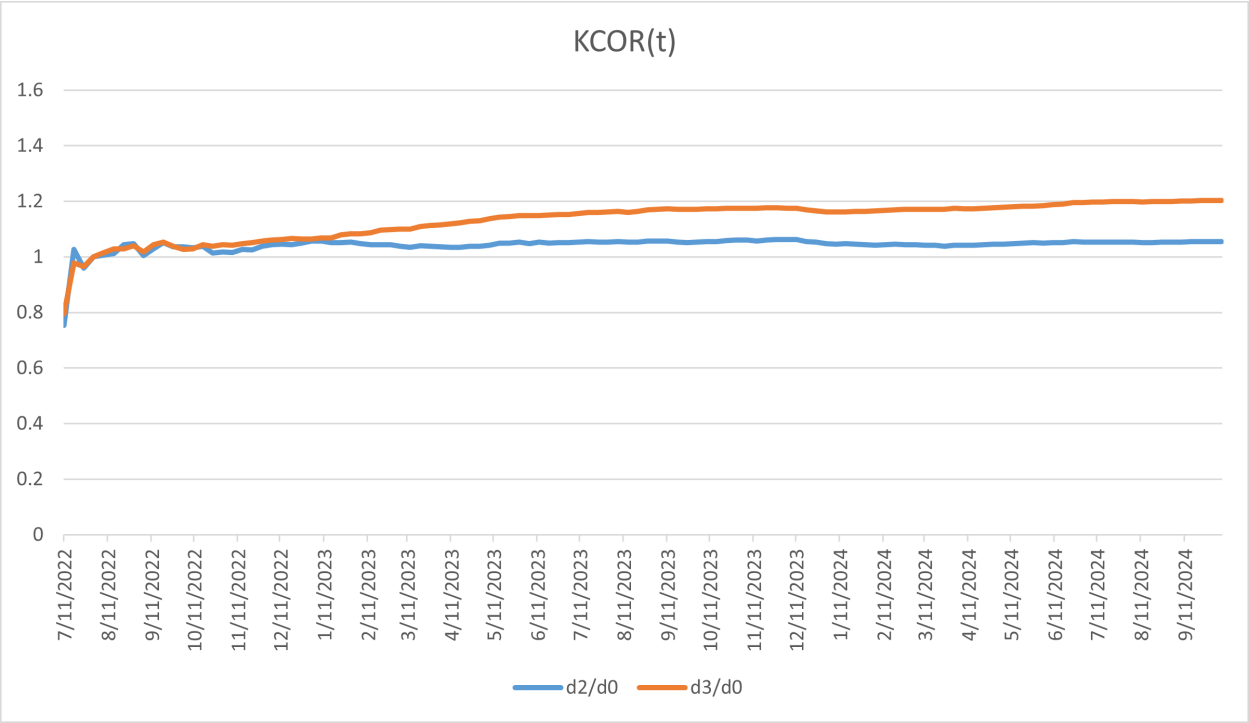

Figure S6: Birth-year cohort 1940: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at t_0 = 4 weeks (i.e., plotted as KCOR$(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

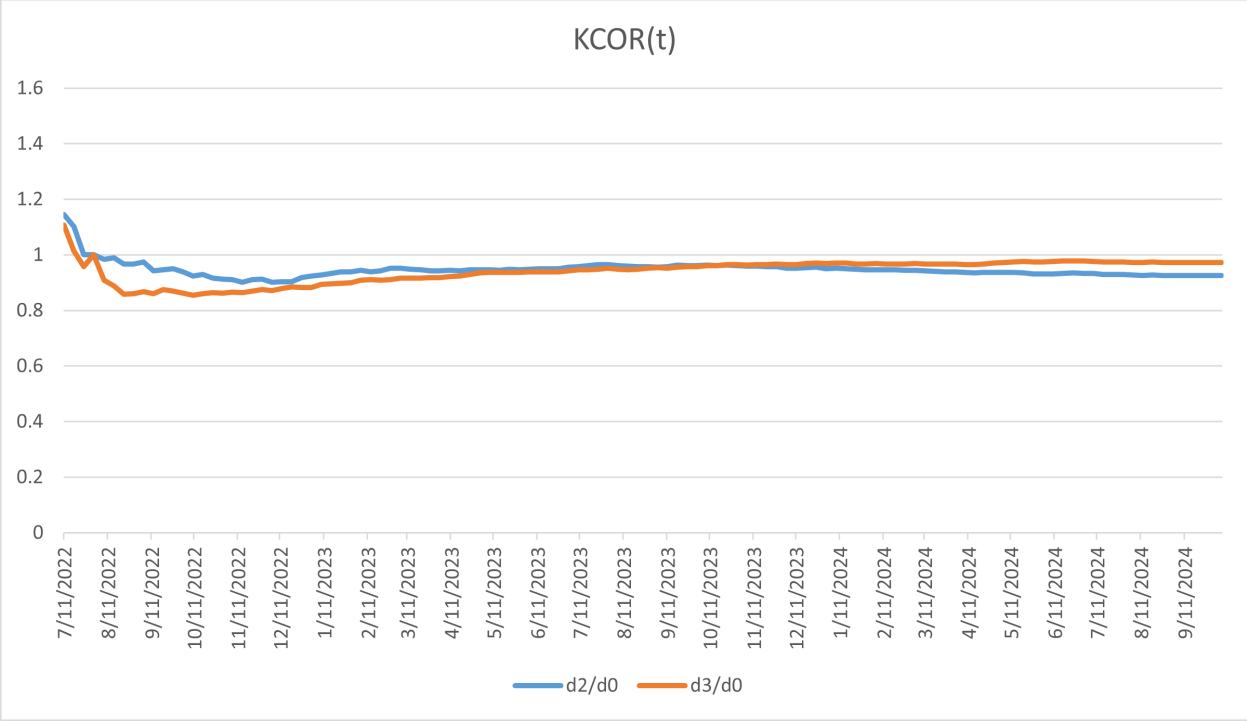

Figure S7: Birth-year cohort 1950: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR$(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

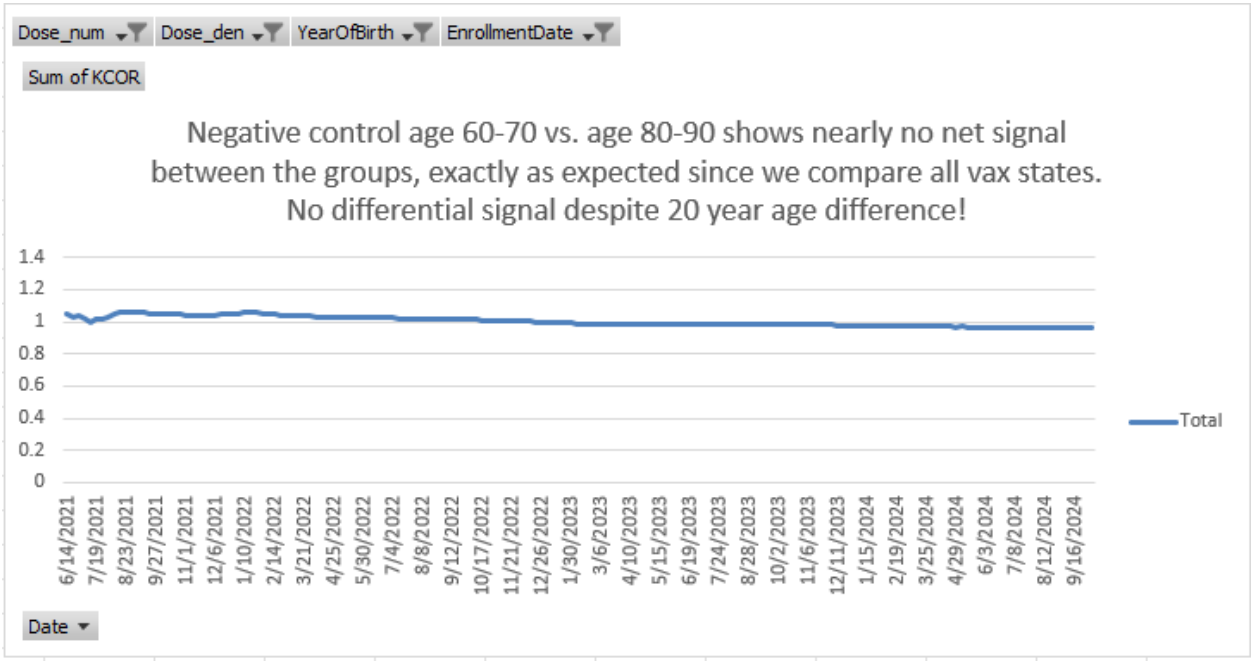

Figure S8: Empirical negative control with approximately 20-year age difference between cohorts. Even under extreme composition differences, KCOR(t) exhibits no systematic drift,
consistent with robustness to selection-induced curvature. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR(t; t_0)). Uncertainty bands (95% bootstrap intervals) are shown. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Supplementary Information, SI). X-axis units are weeks since enrollment.

S6. Extended Czech empirical application

S6.1 Empirical application with diagnostic validation: Czech Republic national registry mortality data

The Czech results do not validate KCOR; they represent an application that satisfies all pre-specified diagnostic criteria. Substantive implications follow only if the identification assumptions hold. Throughout this subsection, observed divergences are interpreted strictly as properties of the estimator under real-world selection, not as intervention effects.

Unless otherwise noted, KCOR curves in the Czech analyses are shown anchored at $t_0 = 4$ weeks for interpretability.

S6.1.1 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 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.

S6.1.2 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:

$$\hat{\theta}_d \approx 0. \tag{17}$$

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 baseline cumulative hazard for others. This behavior is consistent with curvature-driven normalization rather than cohort identity. This pattern is visible directly in depletion-neutralized versus observed cumulative hazard plots and is summarized quantitatively in the fitted-parameter logs (see `KCOR_summary.log`).

After frailty normalization, the depletion-neutralized baseline 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.

Table S6 summarizes these diagnostic checks across age strata.

Table S6: Diagnostic gate for Czech application: KCOR results reported only where diagnostics pass.

Age band (years)	Quiet window valid	Post-normalization linearity	Parameter stability	KCOR reported
40–49	Yes	Yes	Yes	Yes
50–59	Yes	Yes	Yes	Yes
60–69	Yes	Yes	Yes	Yes
70–79	Yes	Yes	Yes	Yes
80–89	Yes	Yes	Yes	Yes
90–99	Yes	Yes	Yes	Yes
All ages	Yes	Yes	Yes	Yes

All age strata in the Czech application satisfied the prespecified diagnostic criteria, permitting KCOR computation and reporting. KCOR results are not reported for any age stratum where diagnostics indicate non-identifiability.

Interpretation: In this application, 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:

$$\hat{\theta}_d > 0. \tag{18}$$

for Dose 0 cohorts and

$$\hat{\theta}_d \approx 0. \tag{19}$$

for Dose 2 cohorts. Estimated frailty heterogeneity can appear larger at younger ages because baseline hazards are low, so proportional differences across latent risk strata translate into visibly different short-term hazards before depletion compresses the risk distribution. At older ages, higher baseline hazard and stronger ongoing depletion can reduce the apparent dispersion of remaining risk, yielding smaller fitted θ even if latent heterogeneity is not literally smaller. 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 S7 demonstrates selection-induced heterogeneity, unadjusted cumulative outcome contrasts are expected to conflate depletion effects with any true treatment differences; see Table S8 for raw cumulative hazards reported as a pre-normalization diagnostic. KCOR normalization removes the depletion component, enabling interpretable comparison of the remaining differences.

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

Table S7: 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 selection-induced depletion 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 S8 reports raw outcome contrasts for ages 40+ (YOB $\leq$ 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.

Table S8: Ratio of observed cumulative mortality hazards for unvaccinated (Dose 0) versus fully vaccinated (Dose 2) Czech cohorts enrolled in 2021_24. (Note: the all-ages row reflects aggregation effects and is not directly comparable to age-stratified rows.)

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.

S6.1.3 Illustrative application to national registry mortality data

A brief illustrative application is included to demonstrate end-to-end KCOR behavior on real registry mortality data in a setting that minimizes timing-driven shocks and window-tuning sensitivity. Cohorts were enrolled in ISO week 2022-26, and evaluation was restricted to calendar year 2023, yielding a 26-week post-enrollment buffer before slope estimation and a prespecified full-year window for assessment. Frailty parameters were estimated using a prespecified epidemiologically quiet window (calendar year 2023) to minimize wave-related hazard variation. This example is intended to illustrate estimator behavior under real-world selection and heterogeneity and does not support causal inference.

Figure S9 shows KCOR(t) trajectories for dose 2 and dose 3 relative to dose 0 for an all-ages analysis. An all-ages analysis is presented as a high-heterogeneity stress test, since aggregation across age induces substantial baseline hazard and frailty variation.

S7. Computational environment and runtime notes

Environment. Python 3.11; key dependencies include numpy, scipy, pandas, and lifelines (for Cox-model comparisons), with plotting via matplotlib.

Compute requirements. The full simulation grid reproduces in approximately 1 hour 26 minutes on a 20-core CPU with 128 GB RAM; smaller subsets reproduce in minutes.

Reproduction. Running `make paper` (or the repository's top-level build command) regenerates all artifacts from a clean checkout.

Figure S9: All-ages stress test: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR($t; t_0$)). This figure is presented as an illustrative application demonstrating estimator behavior under extreme heterogeneity and does not support causal inference. X-axis units are weeks since enrollment.