KCOR: Depletion-Neutralized Cohort Comparison via Gamma-Frailty Normalization

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**: 12,100 (excluding Abstract, References, and Appendices)
- **Keywords**: selection bias; frailty model; gamma mixture model; frailty inversion; frailty heterogeneity; depletion of susceptibles; non-proportional hazards; cumulative hazard; hazard normalization; cumulative hazards; estimands; gamma frailty; negative controls; observational studies; observational cohort studies

Abstract

Background
Retrospective cohort analyses frequently involve heterogeneous populations subject to selection-induced depletion under latent frailty. This process produces non-proportional hazards and curvature in observed cumulative hazards that can bias standard survival estimands when applied directly to registry and administrative data.

Methods
We introduce KCOR, a depletion-neutralized cohort comparison framework based on gamma-frailty normalization. KCOR estimates cohort-specific depletion geometry during prespecified epidemiologically quiet periods and applies an analytic inversion to map observed cumulative hazards into a common, depletion-neutralized scale prior to comparison. The method requires only minimal event-time information and does not rely on proportional hazards assumptions or rich covariate adjustment.

Results
Through extensive simulation studies spanning a wide range of frailty heterogeneity and selection strength, as well as empirical negative and positive controls, we show that commonly used methods—including 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 across these settings.

Conclusions
KCOR provides a diagnostic and descriptive framework for comparing fixed cohorts under selection-induced hazard curvature. By separating depletion normalization from outcome comparison, the method restores a common comparison scale prior to model fitting and improves the interpretability of cumulative outcome analyses in heterogeneous real-world data.

Key contributions

- Introduces a principled depletion-neutralization mapping for heterogeneous cohorts under latent frailty.
- Demonstrates systematic non-null behavior of standard survival methods under selection-only regimes.
- Provides a practical diagnostic framework requiring minimal registry data.

Key messages

• Selection-induced depletion under latent frailty heterogeneity produces non-proportional hazards and curvature in cumulative hazard trajectories that can bias direct application of standard survival estimands in many retrospective cohort studies.

• KCOR provides a diagnostic and normalization framework that removes selection-induced depletion curvature from cumulative hazards using minimal registry data, restoring a common comparison scale across cohorts.

• KCOR separates normalization from comparison: once hazards are depletion-neutralized, cohorts may be compared using standard post-adjustment estimands (e.g., ratios, differences, slopes, or restricted mean survival time), with the choice driven by interpretability rather than identifiability constraints.

• Simulation studies and empirical controls show that under selection-only regimes, KCOR-normalized trajectories remain stable and centered near the null, while commonly used estimands such as Cox regression can exhibit systematic non-null behavior when applied to unadjusted data.

Methods Summary

KCOR is a cumulative-hazard normalization framework for retrospective cohort comparisons under selection-induced non-proportional hazards. The method requires minimal inputs (enrollment date, intervention date, outcome time) and proceeds as follows:

- **Input data**: For each cohort, aggregate weekly counts of individuals alive and dead, indexed by enrollment date, birth year (or age group), and intervention status (dose/cohort label). Minimal required fields are dates of birth, intervention, and event (or censoring).
- **Compute observed hazards**: From weekly alive/dead counts, compute weekly mortality risk and accumulate to form observed cumulative hazards $H_{\text{obs},d}(t)$ for each cohort d.
- **Identify quiet window**: Prespecify an epidemiologically quiet period (ISO weeks) during which external shocks and cohort-differential perturbations are minimal. This window is used exclusively for frailty parameter estimation.
- **Fit frailty curvature parameter**: During the quiet window, fit gamma-frailty parameters $(\hat{k}_d, \hat{\theta}_d)$ independently for each cohort using constrained nonlinear least squares in cumulative-hazard space (Equation (10)). The frailty variance parameter θ_d captures selection-induced depletion curvature.
- **Compute adjusted cumulative hazard**: Apply the gamma-frailty inversion (Equation (11)) to transform observed cumulative hazards into depletion-neutralized baseline cumulative hazards $\tilde{H}_{0,d}(t)$ for the full follow-up period.

After normalization, cohort comparisons are performed on the depletion-neutralized scale using a prespecified estimand; in this manuscript, we report cumulative hazard ratios for concreteness.

- **Compute KCOR trajectory**: Form the ratio $\text{KCOR}(t) = \tilde{H}_{0,A}(t)/\tilde{H}_{0,B}(t)$ between cohorts. Interpret flatness (drift < 5% per year) during quiet windows as successful depletion normalization; deviations during effect windows indicate treatment-related cohort differences when temporal separability holds.
- **Uncertainty quantification (optional)**: For Monte Carlo confidence intervals, resample cohort data with replacement (stratified by cohort and stratum), re-estimate frailty parameters, recompute KCOR trajectories, and form percentile-based intervals (2.5th and 97.5th percentiles) at each time point. Recommended: 1000–2000 bootstrap replicates for stable percentile intervals.
- **Diagnostics**: Assess quiet-window fit quality (RMSE in cumulative-hazard space), post-normalization linearity (R^2 from linear fit), and parameter stability under boundary perturbations. If diagnostics fail, treat KCOR as not identified rather than reporting potentially misleading contrasts.

Observed cumulative hazards are computed by summing weekly integrated hazard increments $\Delta H_d(t) = -\log\left(1 - D_{d,t}/Y_{d,t}\right)$, where $D_{d,t}$ is the number of deaths during week t and $Y_{d,t}$ is the number at risk at the start of that week.

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 demonstrate estimator behavior, diagnostics, and failure modes under realistic selection-induced non-proportional hazards; no causal or policy conclusions are drawn.

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

Selection does not merely shift mortality **levels**; it can alter mortality **curvature**—the time-evolution of cohort hazards. Frailty heterogeneity and depletion of susceptibles naturally induce curvature 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.

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

This violates core assumptions of many standard tools:

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

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

The methodological problem addressed here is general. The COVID-19 period provides a natural empirical regime characterized by strong selection heterogeneity and non-proportional hazards, serving as a useful illustration for the proposed framework. However, KCOR is not specific to COVID, vaccination, or infectious disease. The estimator applies to any retrospective cohort comparison in which selection induces differential depletion dynamics that violate proportional hazards assumptions. KCOR refers to the method as presented here; earlier internal iterations are not material to the estimand or results and are omitted for clarity.

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

- **Static HVE:** baseline differences in latent frailty distributions at cohort entry (e.g., vaccinated cohorts are healthier on average). In the KCOR framework, this manifests as differing depletion curvature (different θ_d) and is the primary target of frailty normalization.
- **Dynamic HVE:** short-horizon, time-local selection processes around enrollment that create transient hazard suppression immediately after enrollment (e.g., deferral of vaccination during acute illness, administrative timing, or short-term behavioral/health-seeking changes). Dynamic HVE is operationally addressed by prespecifying a skip/stabilization window (§2.7) and can be evaluated empirically by comparing early-period signatures across related cohorts in multi-dose settings.

1.3 Related work (brief positioning)

KCOR builds on the frailty and depletion-of-susceptibles 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 Supporting Information (SI).

1.4 Evidence from the literature: residual confounding despite meticulous matching

Motivating applied studies show 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]. This motivates estimators that explicitly address selection-induced hazard curvature rather than treating time variation as signal.

1.5 Contribution of this work

This work makes four primary contributions.

First, we identify and formalize a common failure mode in retrospective cohort survival analysis: selection-induced depletion under latent frailty heterogeneity generates non-proportional hazards and curvature in cumulative hazard trajectories that can bias standard survival estimands when applied directly to observed data.

Second, we introduce KCOR as a diagnostic and normalization framework that estimates cohort-specific depletion geometry during epidemiologically quiet periods and maps observed cumulative hazards into a depletion-neutralized space via gamma-frailty inversion. This normalization step restores approximate stationarity of hazards and comparability across cohorts using only minimal registry data.

Third, we demonstrate through simulation and empirical negative controls that commonly used estimands—including Cox proportional hazards regression and related survival-based summaries—can exhibit systematic non-null behavior under selection-only regimes, while KCOR-normalized trajectories remain stable and centered near the null.

Fourth, we clarify that KCOR does not privilege a single comparison estimand. Rather, it separates normalization from comparison: once hazards are depletion-neutralized, cohorts may be compared using a range of standard post-adjustment estimands (e.g., ratios, differences, slopes, or restricted mean survival time), with the choice driven by interpretability and scientific context. In this manuscript, we focus on ratios of adjusted cumulative hazards as a stable and interpretable summary aligned with the normalization geometry.

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 1. Target estimand and scope (non-causal).

- **Primary estimand (KCOR)**: For two fixed enrollment cohorts A and B, we define

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

  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).
- **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 fail**: treat the analysis as not identified and do not report KCOR as a "corrected effect".

1.7 Paper organization and supporting information (SI)

The main text presents the KCOR estimator, a single canonical demonstration of Cox bias under frailty-driven depletion, and two main validation examples (negative control and stress test). Additional validations (including positive controls), extended diagnostics, and detailed simulation/control specifications are provided in the Supporting Information (SI) document.

2. Methods

While mortality is used as the primary example throughout this section, KCOR applies to any irreversible event process. The methodological framework is event-agnostic; mortality serves as a concrete illustration 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.

KCOR 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's 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; in this work, we use ratios of depletion-neutralized cumulative hazards (KCOR).

2.1.1 Target estimand

KCOR separates normalization from comparison. The normalization step produces depletion-neutralized baseline cumulative hazards that render cohorts comparable; the subsequent comparison may be carried out using a variety of cumulative or summary estimands. In this manuscript, we define and report the cumulative hazard ratio as the primary estimand.

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

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

Unlike hazard ratios, KCOR compares cumulative hazard accumulation over time and does not rely on proportional hazards assumptions or conditioning on survival at time t.

Interpretation (unanchored KCOR). KCOR(t) is the ratio of depletion-normalized cumulative baseline hazards accumulated by two cohorts from enrollment to time t. KCOR(t) > 1 indicates that, after accounting for selection-induced depletion via frailty normalization, cohort A has accumulated greater cumulative hazard than cohort B over $[0, t]$. Because KCOR(t) reflects cumulative hazard levels rather than instantaneous rates, it incorporates both baseline hazard differences and any pre-existing cohort differences present at enrollment. Unanchored KCOR is level-dependent and retains baseline offsets; it is not centered at 1 even under parallel hazards.

For visualization we sometimes report an **anchored KCOR** that shows post-reference divergence:

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

with prespecified t_0 (e.g., 4 weeks).

$\mathrm{KCOR}(t)$ is a **cumulative outcome contrast** after removal of curvature attributed to selection-induced depletion under the working frailty model. The estimand is defined regardless of whether it has a causal interpretation.

2.1.2 Identification versus diagnostics

KCOR is presented here as a **normalization-and-comparison framework**, not as a general causal estimator under unmeasured confounding. A causal interpretation of $\mathrm{KCOR}(t)$ requires additional substantive conditions (e.g., that the quiet-window curvature is dominated by selection-induced depletion rather than cohort-differential external shocks). Because these conditions are inherently dataset- and design-dependent, KCOR emphasizes **diagnostic enforcement**: when assumptions required for interpretable normalization are not supported, the method should signal this through degraded fit, residual structure, or instability to window perturbations rather than silently producing a "corrected" contrast.

Operationally, interpretability of a KCOR trajectory is assessed via prespecified checks (Supporting Information, SI), including:

- 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 Table D.1 and discussed in detail in the Supporting Information (SI).

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 diagnostic rather than causal in nature:

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.

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 (Appendix E).

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

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

We compute the discrete-time cohort hazard as

$$h_{\mathrm{obs},d}(t) = -\ln\bigl(1 - \mathrm{MR}_{d,t}\bigr) = -\ln\Bigl(1 - \frac{d_d(t)}{N_d(t)}\Bigr). \tag{2}$$

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, $h_{\mathrm{obs},d}(t) \approx \mathrm{MR}_{d,t} = d_d(t)/N_d(t)$, but the log form remains accurate and stable when weekly risks are not negligible.

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

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

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

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, we computed $\hat{H}_{\mathrm{obs},d}(t)$ 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

Within cohort d, individual i is modeled as having hazard

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

Here $h_{0,d}(t)$ 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

$$H_{0,d}(t) = \int_0^t h_{0,d}(s)\, ds \tag{5}$$

denote the baseline cumulative hazard. Integrating over gamma frailty yields the gamma-frailty identity

$$H_{\mathrm{obs},d}(t) = \frac{1}{\theta_d} \log\bigl(1 + \theta_d H_{0,d}(t)\bigr), \tag{6}$$

which can be inverted exactly as

$$H_{0,d}(t) = \frac{\exp\bigl(\theta_d H_{\mathrm{obs},d}(t)\bigr) - 1}{\theta_d}. \tag{7}$$

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.

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:

$$h_{0,d}(t) = k_d, \qquad H_{0,d}(t) = k_d\, t. \tag{8}$$

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). 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, KCOR is treated as not identified for that analysis rather than producing a potentially misleading normalized contrast. 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 protocol. The quiet window is selected prior to KCOR estimation using the following operational criteria:

1. Calendar-time hazard curves exhibit approximate linearity with no sustained trend breaks.
2. Periods containing epidemic waves, reporting artifacts, or policy shocks are excluded.
3. The window spans a minimum duration sufficient for stable slope estimation.
4. Sensitivity is assessed by perturbing the window boundaries (± several weeks).

Failure signals (do not report KCOR as identified). Treat the analysis as not identified if any cohort shows: (i) poor fit in cumulative-hazard space over the quiet window; (ii) persistent post-normalization nonlinearity within the quiet window; or (iii) instability of $(\hat{k}_d, \hat{\theta}_d)$ under small boundary perturbations (e.g., ±4 weeks).

Practical example. In COVID-19 mortality analyses, a quiet window may be defined as an inter-wave period between major variant surges, verified by approximately linear all-cause cumulative hazards in the general population and the absence of cohort-differential policy or reporting shocks. The specific calendar bounds are not assumed to be unique or correct a priori; instead, robustness to small perturbations of the window boundaries (e.g., ± several weeks) is treated as a core diagnostic. If fitted depletion parameters or post-normalization linearity are unstable under such perturbations, the quiet-window assumption is deemed violated and KCOR is treated as not identified for that analysis.

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 (ISO week space). The quiet window is prespecified and applied consistently across cohorts within an analysis; robustness to alternate quiet-window bounds is assessed in sensitivity analyses. Quiet periods are identified diagnostically via stability of observed cumulative hazards and absence of external shocks, rather than by a fixed universal numeric threshold. The epidemiological quiet period is not assumed to be free of all processes, but to lack sharp, cohort-differential hazard perturbations (e.g., epidemic waves or policy shocks) capable of inducing systematic curvature asymmetry unrelated to selection dynamics. 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.

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

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

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

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

Parameters are estimated by constrained nonlinear least squares:

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

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

All analyses use a prespecified reference implementation with fixed operational defaults; full details are provided in Appendix E.

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:

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

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 1).

2.6.1 Computational considerations

KCOR operates on aggregated event counts in discrete time and cumulative-hazard space. Computational complexity scales linearly with the number of time bins and strata rather than the number of individuals, making the method feasible for very large population registries. In practice, KCOR analyses on national-scale datasets (millions of

individuals) are memory-bound rather than CPU-bound and can be implemented efficiently using standard vectorized numerical libraries. No iterative optimization over individual-level records is required.

2.6.2 Internal diagnostics and 'self-check' behavior

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

1. **Post-normalization linearity in quiet periods.** During a prespecified quiet window, the working model assumes that curvature in observed cumulative hazard is primarily driven by depletion under heterogeneity. After inversion, the depletion-neutralized cumulative hazard should be approximately linear in event time over the same quiet window. Systematic residual curvature (e.g., sustained concavity/convexity) indicates that the quiet-window assumption is violated (external shocks, secular trends) or that the depletion geometry is misspecified for that cohort.
2. **Fit residual structure in cumulative-hazard space.** Define residuals over the fit set $\mathcal{T}_d$:

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

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 correction.

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 'corrected' estimate. Failure of these diagnostics indicates that the depletion-based normalization is inappropriate, in which case 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:

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

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

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

2.8 KCOR estimator

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

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

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

2.9 Uncertainty quantification

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

2.9.1 Stratified bootstrap procedure

The stratified bootstrap procedure for KCOR proceeds as follows:

1. **Resample individuals (or counts).** Within each cohort and stratum (e.g., age group), resample individuals with replacement, preserving the original cohort and stratum structure. Alternatively, for aggregated data, resample event counts and risk-set sizes within each time bin and stratum.
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 via the gamma-frailty inversion 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).

Uncertainty intervals reflect event stochasticity and model-fit uncertainty in $(\hat{k}_d, \hat{\theta}_d)$ and are interpreted conditional on the observed risk sets and modeling assumptions.

2.10 Algorithm summary and reproducibility checklist

Table 5 summarizes the complete KCOR pipeline.

2.11 Relationship to Cox proportional hazards

Cox proportional hazards models estimate an instantaneous hazard ratio under the assumption that hazards differ by a time-invariant multiplicative factor. Under selective uptake with latent frailty heterogeneity, this assumption is typically violated, yielding time-varying hazard ratios induced purely by depletion dynamics. This reflects an estimand mismatch: Cox targets a different quantity under depletion than KCOR's cumulative hazard estimand. Cox is behaving correctly for its estimand, but that estimand may not align with the scientific question when selection-induced depletion is present. Accordingly, Cox results are presented here as a diagnostic demonstration of estimand mismatch, not as a competing causal estimator.

Cox regression estimates a weighted average hazard ratio under non-proportional hazards; KCOR targets a cumulative hazard estimand. Even when Cox models are extended with shared frailty to accommodate heterogeneity,

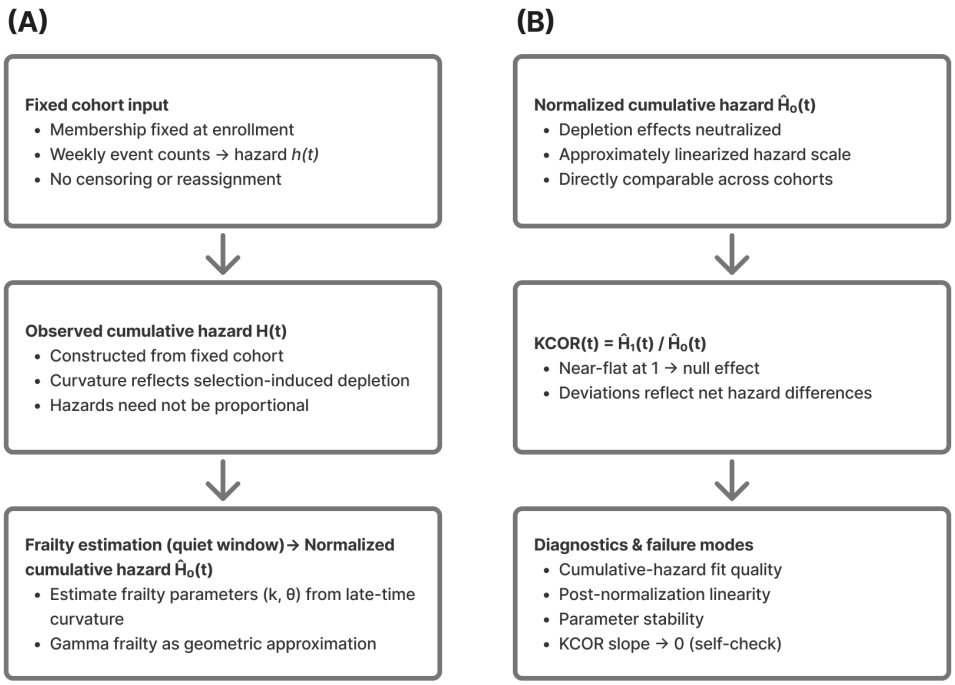

Figure 1: **KCOR as a two-stage framework. (A)** Fixed-cohort cumulative hazards exhibit curvature due to selection-induced depletion; late-time curvature is used to estimate frailty parameters for normalization. **(B)** Gamma-frailty normalization yields approximately linearized cumulative hazards that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

they continue to estimate instantaneous hazard ratios conditional on survival, whereas KCOR estimates cumulative contrasts after explicit depletion normalization.

Conceptually, Cox regression estimates an instantaneous hazard ratio by fitting a hazard model to observed data, whereas KCOR uses a parametric working model only to normalize selection-induced depletion geometry and then computes a cumulative contrast on the depletion-neutralized scale.

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

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

Data-generating process.

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

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

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 true causal hazard ratio between cohorts is exactly 1 for all t.

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

Cox analysis.

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

KCOR analysis.

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

Expected behavior under the null.

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

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

Summary of findings.

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

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

Table 6 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.

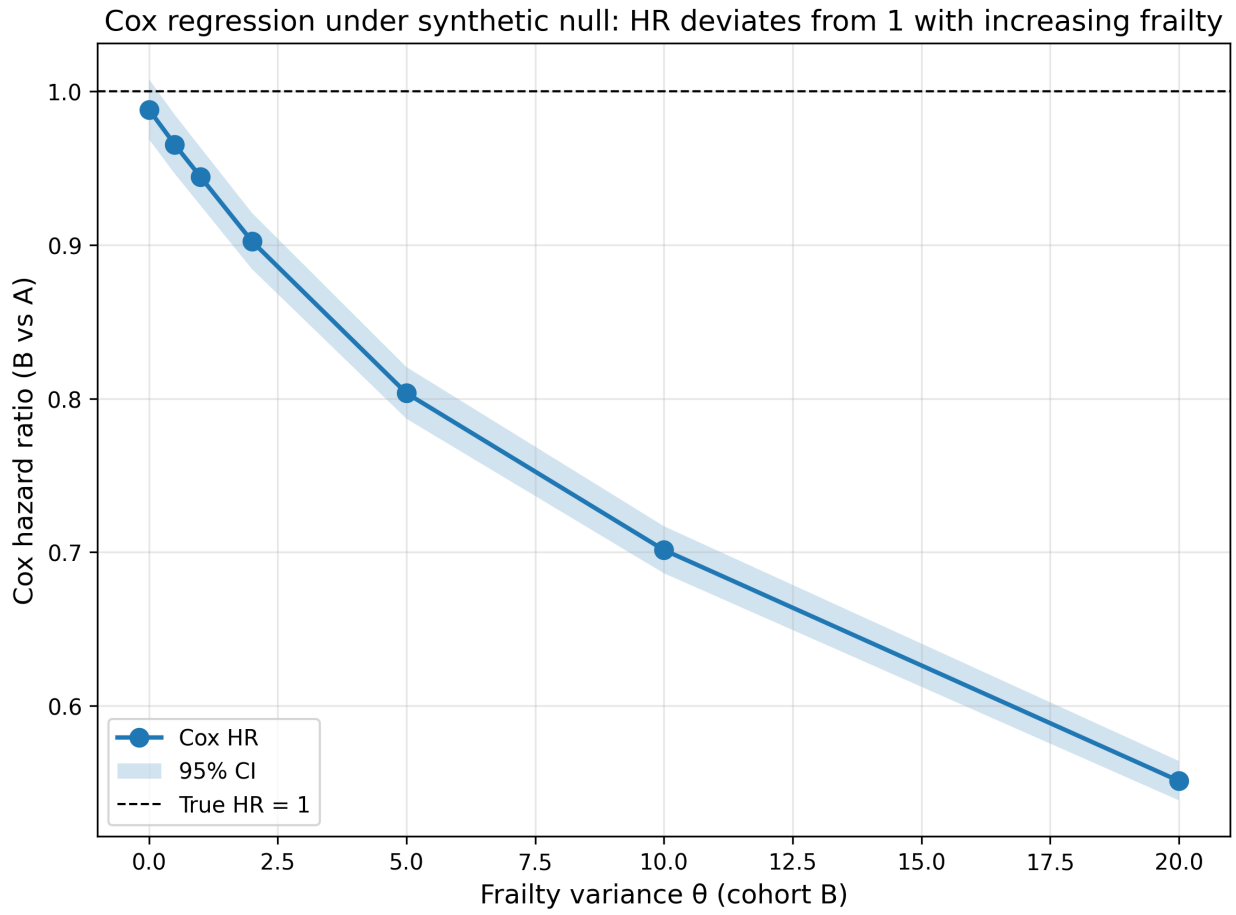

Figure 2: Cox regression produces spurious non-null hazard ratios under a *synthetic null* as frailty heterogeneity increases. Hazard ratios (with 95% confidence intervals) from Cox proportional hazards regression comparing cohort B to cohort A in simulations where the true treatment effect is identically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

Interpretation.

This controlled synthetic null shows 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)

We include a brief worked example to illustrate the KCOR workflow end-to-end. This example is descriptive and intended solely to demonstrate the mechanics of cohort construction, hazard estimation, frailty fitting, depletion normalization, and KCOR computation, without supporting causal interpretation.

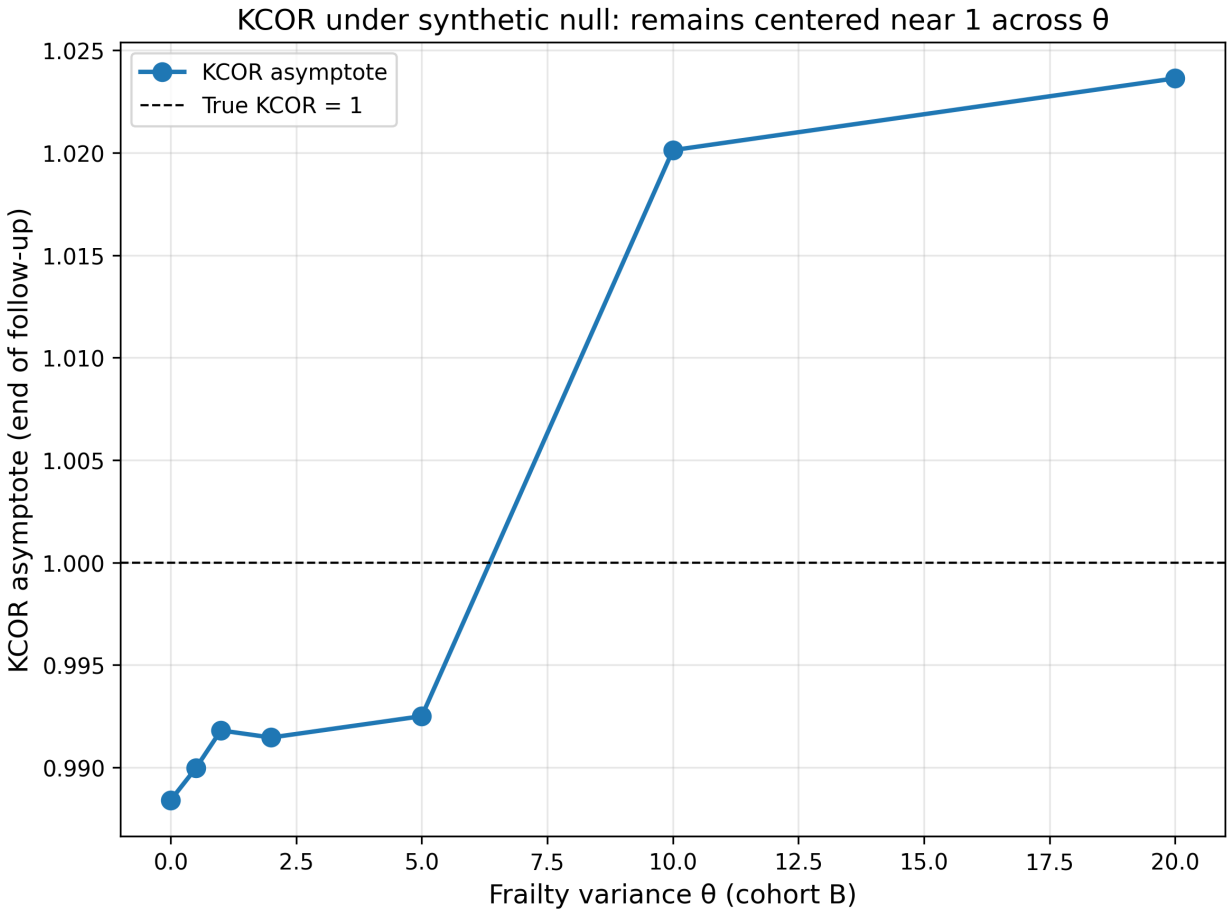

Figure 3: KCOR(t) remains null under a synthetic null across increasing frailty heterogeneity. KCOR(t) asymptotes remain near 1 across θ in the same simulations, consistent with correct null behavior after depletion normalization. Uncertainty bands (95% bootstrap intervals) are shown but are narrow due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

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

All figures, tables, and simulations can be reproduced from the accompanying code repository. Commands and execution order are provided in the repository README.

2.14 Computational implementation and reproducibility

Analyses were implemented in Python and are fully reproducible from the public repository and archived release (see Code/Data Availability). Additional environment and runtime details are provided in the Supporting Information (SI).

3. Validation and control tests

This section is the core validation claim of KCOR:

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

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

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

3.1 Negative controls: null under selection-induced curvature

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 and does not support causal inference.

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

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

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

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

Figure **??** provides a representative illustration; additional age-shift variants are provided in the Online Supplement.

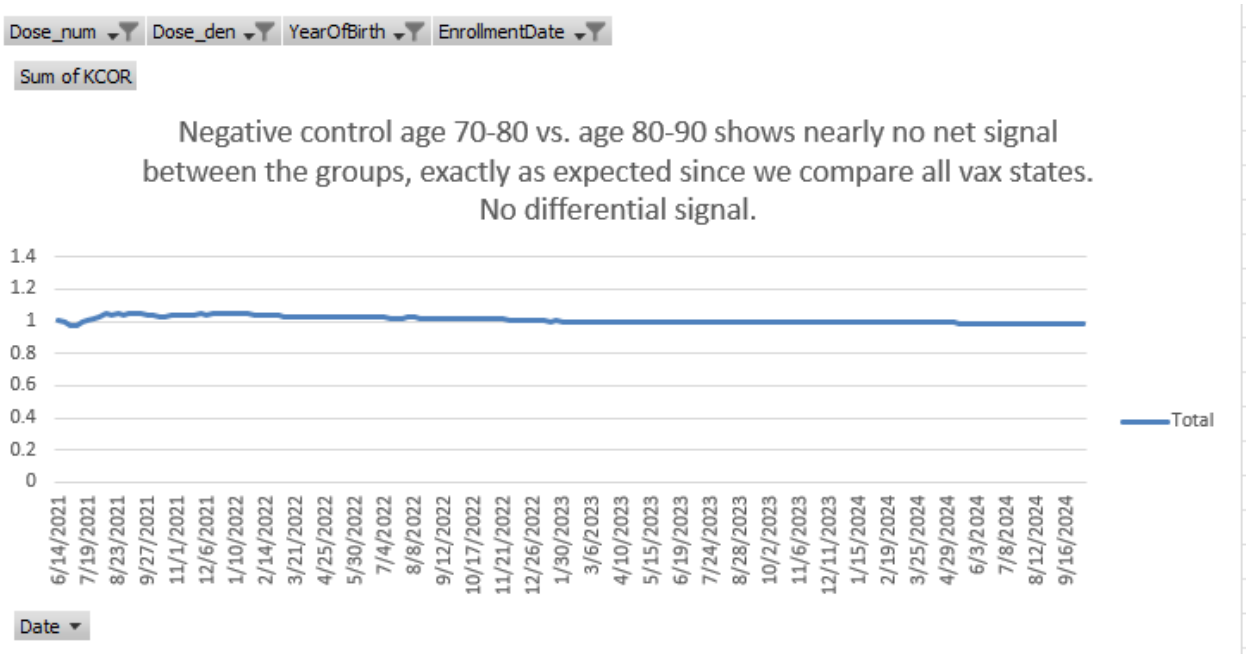

Table 7 provides numeric summaries (Online Supplement).

3.2 Positive controls: detect injected harm/benefit

Positive controls (injected harm/benefit) are provided in the Supporting Information (SI), 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).

3.3 Stress test: robustness to frailty misspecification

3.3.1 Frailty misspecification robustness

To assess robustness to departures from the gamma frailty assumption, we conducted simulations 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, we report bias (deviation from true cumulative hazard ratio), variance (trajectory stability), coverage (proportion of simulations where uncertainty intervals contain the true value), and diagnostic failure rate (proportion of simulations where quiet-window diagnostics indicated non-identifiability).

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. Additional validation results—including full simulation grids, quiet-window robustness catalogs, dynamic-selection checks, and extended comparator analyses—are provided in the Supporting Information (SI).

4. Discussion

Scope statement (non-causal positioning). As summarized in Box 1 (§1.6), KCOR is a normalization-and-diagnostic framework rather than a causal identification strategy. It produces a depletion-neutralized **cumulative** contrast whose interpretability is conditional on explicit assumptions and on internal diagnostics; when diagnostics fail, KCOR is intended to signal non-identifiability rather than silently produce a "corrected" estimate.

What KCOR does not provide

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

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

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

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 causal 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.

4.1 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 operates at a specific but critical layer of the retrospective inference stack: it both neutralizes selection-induced depletion dynamics and defines how the resulting depletion-neutralized baseline cumulative hazards must be compared. Once cohorts are mapped into depletion-neutralized baseline cumulative hazard space, $\mathrm{KCOR}(t)$ answers whether one cohort accumulated higher or lower cumulative event risk than another by time t, conditional on the stated assumptions and diagnostics (Box 1). For intuition, $\mathrm{KCOR}(t) = 1.2$ indicates that, after depletion normalization, cohort A has accumulated approximately 20% greater cumulative hazard than cohort B by time t. Stabilization of $\mathrm{KCOR}(t)$ in quiet windows is a falsification check: failure to flatten indicates residual curvature or loss of identifiability, not a substantive cumulative effect.

Many commonly used survival estimands—such as hazard ratios, cumulative hazard differences, or restricted mean survival time—are not intrinsically invalid. Their failure in retrospective cohort studies arises when they are applied to unadjusted data exhibiting selection-induced depletion. KCOR does not replace these estimands; instead, it provides a normalization step that restores comparability. After depletion normalization, such estimands may be meaningfully computed, with the choice driven by interpretability rather than by identifiability constraints imposed by selection bias.

The frailty term is used as a geometric working model for selection-induced depletion; it is not interpreted mechanistically.

KCOR is a **cumulative** comparison of depletion-neutralized cumulative hazards; it does not estimate instantaneous hazard ratios. It is designed for settings where selection induces non-proportional hazards such that conventional proportional-hazards estimators can be difficult to interpret. A controlled synthetic null experiment (Section 2.11.1) shows that Cox regression can return statistically significant non-null hazard ratios solely from frailty-induced depletion—even when the true treatment effect is identically zero—reflecting an estimand mismatch where Cox targets a different quantity under depletion than KCOR's cumulative estimand. Cox is behaving correctly for its estimand, but that estimand may not align with the scientific question when selection-induced depletion is present. KCOR remains centered near unity with negligible post-normalization slope under the same conditions. We did

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

Under the working assumptions that:

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

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

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

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

Because the normalization targets selection-induced depletion curvature, KCOR results alone do not justify claims about net lives saved or lost by a particular intervention. Such claims require (i) clearly specified causal estimands, (ii) validated control outcomes, (iii) sensitivity analyses for remaining time-varying selection mechanisms and external shocks, and (iv) preferably replication across settings and outcomes. Having established the behavior of KCOR and the failure modes of standard estimators under controlled conditions, we apply KCOR to complete national registry data from the Czech Republic in a companion analysis. Accordingly, this manuscript focuses on method definition, diagnostics, and operating characteristics; applied causal conclusions are deferred to separate intervention-specific analyses. Interpretation should be read in light of the non-identifiability considerations described above.

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

4.2 Relationship to negative control methods

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

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

4.3 Practical Guidelines for Implementation

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

Recommended reporting includes:

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

KCOR should therefore be applied and reported as a complete pipeline—from cohort freezing, through depletion normalization, to cumulative comparison and diagnostics—rather than as a standalone adjustment step. Scope and interpretation are summarized once in Box 1 (§1.6).

5. Limitations

- **Model dependence**: Normalization relies on the adequacy of the gamma-frailty model and the baseline-shape assumption during the quiet window.
- **Relation to existing non-PH methods**: KCOR is complementary to time-varying Cox, flexible parametric, additive hazards, and MSM approaches; these methods address different estimands and identification strategies, whereas KCOR targets depletion-geometry normalization under minimal-data constraints (see §1.3.1).
- θ **estimation is data-driven**: KCOR does not impose $\theta = 0$ for any cohort. The frequent observation that fitted frailty variance estimates collapse toward zero for vaccinated cohorts is a data-driven result of the frailty fit and should not be interpreted as an assumption of homogeneity.
- **Sparse events**: When event counts are small, hazard estimation and parameter fitting can be unstable.
- **Contamination of quiet periods**: External shocks (e.g., epidemic waves) overlapping the quiet window can bias selection-parameter estimation.
- **Causal interpretation**: KCOR supports interpretable cohort comparison under stated assumptions, but it is not a substitute for randomization; causal claims require explicit causal assumptions and careful validation.
- **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 requires careful consideration of competing risks and interpretation of cumulative hazards within cause-restricted populations. Extension to non-fatal outcomes such as hospitalization is conceptually straightforward but may require additional attention to outcome definitions, censoring mechanisms, and recurrent events. These considerations affect interpretation rather than the core KCOR framework.
- **Non-gamma frailty**: The KCOR framework assumes that selection acts approximately multiplicatively through a time-invariant frailty distribution, for which the gamma family provides a convenient and empirically testable approximation. In settings where depletion dynamics are driven by more complex mechanisms—such as time-varying frailty variance, interacting risk factors, or shared frailty correlations within subgroups—the curvature structure exploited by KCOR may be misspecified. In such cases, KCOR diagnostics (e.g., poor curvature fit or unstable fitted frailty variance estimates) serve as indicators of model inadequacy rather than targets for parameter tuning. Extending the framework to accommodate dynamic or correlated frailty structures would require explicit model generalization rather than modification of KCOR normalization steps and is left to future work. Empirically, KCOR's validity depends on curvature removal rather than the specific parametric form; alternative frailty distributions that generate similar depletion geometry would yield equivalent normalization.

5.1 Failure modes and diagnostics (recommended)

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

Common failure modes include:

- **Mis-specified quiet window**: If the quiet window overlaps major external shocks (epidemic waves, policy changes, reporting artifacts), the fitted parameters may absorb non-selection dynamics, biasing normalization.
- **External time-varying hazards masquerading as frailty depletion**: Strong secular trends, seasonality, or outcome-definition changes can introduce curvature that is not well captured by gamma-frailty depletion alone. For example, COVID-19 waves disproportionately increase mortality among frail individuals; if one cohort has higher baseline frailty, such a wave can preferentially deplete that cohort, producing the appearance of a benefit in the lower-frailty cohort that is actually due to differential frailty-specific mortality from the external hazard rather than from the intervention under study.
- **Extremely sparse cohorts**: When events are rare, observed cumulative hazards become noisy and $(\hat{k}_d, \hat{\theta}_d)$ 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.

Practical diagnostics to increase trustworthiness include:

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

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

5.2 Conservativeness and edge-case detection limits

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

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

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

KCOR analyses commonly exclude an initial post-enrollment window to 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. We therefore recommend reporting event counts and conducting a simple cohort-size sensitivity check 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 provides a principled approach to retrospective cohort comparison under selection-induced hazard curvature by estimating and inverting a gamma-frailty mixture model to remove cohort-specific depletion dynamics prior to comparison. Validation via negative and positive controls supports that KCOR remains near-null under selection without effect and detects injected effects when present. Applied analyses on specific datasets are best reported separately from this methods manuscript. KCOR relies on five explicit assumptions, of which only one requires substantive dataset-specific validation, and enforces these assumptions diagnostically rather than presuming them, allowing violations to be detected rather than absorbed into model-dependent estimates. Because standard methods for retrospective vaccine evaluation can exhibit systematic deviation under non-proportional hazards and selection-induced depletion, KCOR provides a practical alternative that operates on minimal individual-level information—dates of birth, intervention, and death—while remaining applicable to national registry data where richer covariates are unavailable or unreliable.

Reproducibility: code, simulations, and manuscript build instructions are available in the project repository (https://github.com/skirsch/KCOR).

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. This is a methods-only manuscript. The primary validation results use synthetic data. Empirical negative-control figures (Figures **??** and **??**) use aggregated cohort summaries derived from Czech Republic administrative data; no record-level data are shared in this manuscript.[12]

Consent for publication

Not applicable.

Data availability

No individual-level data were accessed or analyzed in this study.

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

Code availability

- The KCOR reference implementation and complete validation suite are available in the project repository.
- Repository URL: https://github.com/skirsch/KCOR
- Zenodo DOI: 10.5281/zenodo.18050329
- RMST computation and comparison table generation are implemented via functions `compute_rmst_from_cohort()` and `generate_comparison_table()` in `test/sim_grid/code/generate_sim_grid.py`. Both RMST and time-varying Cox comparators run on the same simulation outputs as KCOR evaluation; no additional random seeds or data generation are required.

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. Dr. Clare Craig, 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. *The Annals of Statistics*. 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 Journal of 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. *Clinical Epidemiology*. 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. Šanca O, Jarkovský J, Klimeš D, et al. Vaccination, positivity, hospitalization for COVID-19, deaths, long covid and comorbidities in people in the Czech Republic. *National Health Information Portal*. Published online 2024.

Tables

Main text 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)
$h_{0,d}(t)$	Baseline hazard for cohort d under the depletion-neutralized model
$H_{0,d}(t)$	Baseline cumulative hazard for cohort d under the depletion-neutralized model
$\tilde{H}_{0,d}(t)$	Depletion-neutralized baseline cumulative hazard
θ_d	Frailty variance (selection strength) for cohort d; governs curvature in the observed cumulative hazard
$\hat{\theta}_d$	Estimated frailty variance from quiet-window fitting
k_d	Baseline hazard level for cohort d under the default baseline shape
$\hat{k}_d$	Estimated baseline hazard level from quiet-window fitting
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.

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	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. (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 show KCOR deviating from 1.0 in the expected direction, validating that the estimator can detect true effects.

Scenario	Effect window	Hazard multiplier r	Expected direction	Observed KCOR(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

Appendix C tables

Table C.1: Estimated gamma-frailty variance (fitted frailty variance) by age band and vaccination status for Czech cohorts enrolled in 2021_24.

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

Notes: - The fitted frailty variance quantifies unobserved frailty heterogeneity and depletion of susceptibles within cohorts. Near-zero values indicate effectively linear cumulative hazards over the quiet window and are typical of strongly pre-selected cohorts. - Each entry reports a single fitted gamma-frailty variance for the specified age band and vaccination status within the 2021_24 enrollment cohort. - The "All ages (full population)" row corresponds to an independent fit over the full pooled age range, included as a global diagnostic. - Table C.3 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 C.2: 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

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

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

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

Values reflect raw cumulative outcome differences prior to KCOR normalization and are not interpreted causally due to cohort non-exchangeability. Cumulative hazards were integrated from cohort enrollment through the end of available follow-up for the 2021_24 enrollment window (through week 2024-16), identically for Dose 0 and Dose 2 cohorts.

Appendix D tables

Table D.1: KCOR assumptions and corresponding diagnostics.

Assumption	What must hold	Diagnostic signal	Interpretation	Action if violated
A1. Fixed cohort at enrollment	Cohort membership does not change over follow-up	Step changes or discontinuities inconsistent with depletion	Endogenous selection or reclassification	Redefine cohort at enrollment; disallow transitions
A2. Shared external hazard environment	Cohorts experience the same background hazard within the comparison window	Divergent slopes during prespecified quiet periods	Unshared exogenous shocks or policy/measurement effects	Restrict calendar window, stratify, or use alternative controls
A3. Time-invariant latent frailty	Individual frailty is time-invariant over follow-up	Systematic residual curvature after normalization	Time-varying susceptibility or competing selection processes	Shorten follow-up window; reinterpret as time-varying selection
A4. Adequacy of gamma frailty	Gamma family adequately approximates frailty mixing	Residual curvature or poor fit diagnostics after inversion	Frailty distribution misspecification	Treat as diagnostic; avoid over-interpretation
A5. Quiet-window validity	No intervention effect during frailty-estimation window	Slope breaks or non-parallel trends within quiet window	Contaminated quiet window	Redefine quiet window; rerun diagnostics

Appendix E tables

Table E.1: 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: $k_d > 0$, $\theta_d \geq 0$

Appendix A: Mathematical derivations

A.1 Frailty mixing induces hazard curvature

Consider a cohort d in which individual i has hazard

$$h_{i,d}(t) = z_{i,d}\, h_{0,d}(t). \tag{A.1}$$

where the frailty $z_{i,d}$ is drawn from a distribution with mean 1 and variance $\theta_d > 0$. Let

$$S_{i,d}(t) = \exp\bigl(-z_{i,d} H_{0,d}(t)\bigr)\,. \tag{A.2}$$

denote the individual survival function, where

$$H_{0,d}(t) = \int_0^t h_{0,d}(s)\, ds. \tag{A.3}$$

is the baseline cumulative hazard.

The cohort survival function, obtained by integrating over the frailty distribution, is

$$S_d(t) = \mathbb{E}[S_{i,d}(t)] = \mathbb{E}[\exp(-z\, H_{0,d}(t))] = \mathscr{L}_z(H_{0,d}(t)). \tag{A.4}$$

where $\mathscr{L}_z(\cdot)$ denotes the Laplace transform of the frailty distribution. The corresponding cohort-level hazard is

$$h_d(t) = -\frac{d}{dt} \log S_d(t). \tag{A.5}$$

Even when $h_{0,d}(t) = k_d$ is constant (so that $H_{0,d}(t) = k_d t$), the cohort hazard $h_d(t)$ is generally time-varying. High-frailty individuals experience events earlier, progressively depleting the higher-risk portion of the cohort and shifting the surviving population, conditional on survival, toward lower frailty over time. This selection-induced depletion is the mechanism by which frailty heterogeneity induces **curvature** in cohort-level hazards.

A.2 Gamma-frailty identity derivation

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

$$\mathscr{L}_z(s) = \left(1 + \theta_d s\right)^{-1/\theta_d}. \tag{A.6}$$

The cohort survival function becomes:

$$S_d^{\text{cohort}}(t) = \left(1 + \theta_d H_{0,d}(t)\right)^{-1/\theta_d}. \tag{A.7}$$

The observed cumulative hazard is defined as:

$$H_{\text{obs},d}(t) = -\log S_d^{\text{cohort}}(t). \tag{A.8}$$

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

$$H_{\text{obs},d}(t) = \frac{1}{\theta_d} \log\bigl(1 + \theta_d\, H_{0,d}(t)\bigr)\,. \tag{A.9}$$

This is the gamma-frailty identity (see Equation (6) in the main text).

A.3 Inversion formula

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

$$H_{0,d}(t) = \frac{\exp\big(\theta_d H_{\text{obs},d}(t)\big) - 1}{\theta_d}. \tag{A.10}$$

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

A.3.1 Relationship to the Vaupel–Manton–Stallard gamma frailty framework

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

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

A.4 Variance propagation (sketch)

For uncertainty quantification, variance in $\mathrm{KCOR}(t)$ can be approximated via the delta method. Define:

$$KCOR(t) = \frac{\tilde{H}_{0,A}(t)}{\tilde{H}_{0,B}(t)}. \tag{A.11}$$

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

$$\mathrm{Var}(KCOR(t)) \approx KCOR(t)^2 \left[\frac{\mathrm{Var}(\tilde{H}_{0,A}(t))}{\tilde{H}_{0,A}(t)^2} + \frac{\mathrm{Var}(\tilde{H}_{0,B}(t))}{\tilde{H}_{0,B}(t)^2} - 2\frac{\mathrm{Cov}(\tilde{H}_{0,A}(t), \tilde{H}_{0,B}(t))}{\tilde{H}_{0,A}(t)\tilde{H}_{0,B}(t)} \right]. \tag{A.12}$$

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

Appendix B: Control-test specifications

Control-test specifications and simulation parameters are provided in the Supporting Information (SI), Section S4.

Frailty structure

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

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

Treatment effect

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

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

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

Quiet period and estimation

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

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

Evaluation criteria

The simulation is considered successful if:

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

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

Interpretation

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

Appendix C: Additional figures and diagnostics

Appendix C material (additional figures, diagnostics, and extended Czech illustrative application) is provided in the Supporting Information (SI), Sections S5–S6.

Appendix D: Diagnostics and Failure Modes for KCOR Assumptions

Appendix D material is provided in the Supporting Information (SI), Section S2.

Appendix E: Reference Implementation and Default Settings

Appendix E material (reference implementation and default operational settings) is provided in the Supporting Information (SI), Section S7.

:::