KCOR: A Depletion-Neutralized Cohort Comparison Framework Using Gamma-Frailty Normalization Under Selection-Induced Cumulative Hazard Curvature

Manuscript metadata

Abstract

Retrospective cohort comparisons are widely used when randomized trials are infeasible, but selection-induced frailty heterogeneity can generate non-proportional hazards that invalidate standard survival estimands. We introduce the Kirsch Cumulative Outcomes Ratio (KCOR), a normalization-and-comparison framework designed to remove selection-induced depletion curvature from cumulative hazards prior to cohort comparison using minimal registry data (dates of birth, intervention, and event).

KCOR estimates cohort-specific depletion parameters during prespecified epidemiologically quiet periods, transforms observed cumulative hazards into a depletion-neutralized space via gamma-frailty inversion, and defines a cumulative hazard ratio as the target estimand. Using simulation studies that vary frailty heterogeneity and selection strength, we evaluate KCOR alongside commonly used hazard- and survival-based estimands under both null and non-null conditions. Under selection-only regimes, KCOR remains centered near the null with stable trajectories, whereas Cox regression and related estimands exhibit systematic non-null behavior driven by residual depletion effects.

KCOR is presented as a diagnostic and normalization framework rather than a causal identification strategy. Interpretation of KCOR trajectories is conditional on stated assumptions and validated through prespecified simulations, negative controls, and internal diagnostics.

Key messages

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

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

• Simulation studies show that KCOR remains stable under selection-only regimes, whereas proportional hazards–based Cox regression can produce non-null hazard ratios driven by selection-induced depletion rather than true signal.

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 rates and convert to discrete-time hazards using a midpoint approximation (Equation {#eq:hazard-from-mr-improved}). 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 {#eq:nls-objective}). The frailty variance parameter θ_d captures selection-induced depletion curvature.

- **Compute adjusted cumulative hazard**: Apply the gamma-frailty inversion (Equation {#eq:normalized-cumhazard}) to transform observed cumulative hazards into depletion-neutralized baseline cumulative hazards $\tilde{H}_{0,d}(t)$ for the full follow-up period.

- **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 (R² from linear fit), and parameter stability under boundary perturbations. If diagnostics fail, treat KCOR as not identified rather than reporting potentially misleading contrasts.

---

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

The examples in this section are intended to be illustrative rather than exhaustive, and are chosen to demonstrate common patterns that arise in retrospective cohort data. They should not be interpreted as defining the scope of applicability of the method, which is agnostic to the specific intervention or event under study.

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

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

To connect KCOR to common vaccine-effectiveness study designs, we additionally simulated rollout-style cohorts with time-localized epidemic hazard ('waves') and uptake correlated with latent frailty. In these settings, Cox regression can return extremely small p-values under a true null due to depletion and time-varying baseline hazard, whereas KCOR remains centered near the null cumulative contrast.

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

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

*The methods reviewed in this section primarily aim to improve estimation of the **observed hazard function** under non-proportionality; KCOR addresses a different problem—**identification and removal of selection-induced depletion geometry prior to cohort comparison**.* A large literature relaxes proportional hazards by allowing either (i) covariate effects to vary over time or (ii) the baseline hazard to be modeled flexibly. Representative examples include time-varying coefficient Cox models, flexible parametric survival models (e.g., spline-based log-cumulative hazard models), additive hazards models (Aalen-type formulations), landmarking and dynamic prediction approaches, and weighting-based approaches such as marginal structural models (MSMs) for time-varying confounding.[4–11] These methods are most effective when the scientific target is an instantaneous contrast (e.g., a time-specific hazard ratio) and when sufficiently rich covariate histories are available to support identification. In settings with rich, time-updated covariate information that fully captures frailty-relevant heterogeneity, weighting or structural modeling approaches may partially mitigate depletion effects. KCOR is designed for the complementary and common regime in which such covariates are unavailable, unreliable, or post-treatment. In contrast, KCOR targets cumulative contrasts under explicit depletion normalization in cumulative-hazard space, rather than instantaneous hazard ratios conditional on survival.

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

This manuscript does not attempt to review the extensive literature on non-proportional hazards exhaustively; instead, it positions KCOR as complementary to these methods by addressing a distinct identifiability problem not resolved by hazard-level flexibility alone.

Summary distinction. KCOR differs from flexible hazard and time-varying coefficient models not in descriptive flexibility, but in estimand and direction of inference. Existing approaches model the observed hazard trajectory and interpret fitted coefficients or functions, implicitly treating all systematic divergence as signal. KCOR instead first normalizes selection-induced depletion geometry in cumulative-hazard space and only then defines a cohort contrast. The target estimand is cumulative rather than instantaneous, and interpretability is enforced diagnostically rather than assumed through model flexibility.

1.3.1a Why flexible hazard models do not resolve depletion bias

Flexible hazard models—including spline-based parametric survival models, time-varying coefficient Cox models, and neural network survival estimators—can improve descriptive fit to observed hazard trajectories by allowing curvature in the hazard function over time. However, these methods do not distinguish whether observed curvature arises from genuine causal time-varying effects or from compositional changes due to selective depletion of susceptibles. Selection on frailty produces curvature in cumulative hazards even under a true null treatment effect, as heterogeneous cohorts experience differential depletion rates over time. *A model that perfectly fits the observed hazard trajectory can still yield a biased estimand if that trajectory is driven by selective depletion rather than a causal effect.* KCOR treats curvature as a **bias to remove** rather than a feature to model, explicitly estimating and neutralizing depletion geometry before defining cohort comparisons.

1.3.1b Time-varying confounding approaches under latent selection

A large literature addresses time-varying confounding using marginal structural models (MSMs), g-computation, and inverse-probability weighting (IPW). These approaches can be effective when the relevant time-varying confounders are **measured with sufficient fidelity** to support sequential exchangeability and correct model or weight specification. However, in retrospective administrative data, key determinants of both intervention uptake and near-term outcome risk are often **unobserved or incompletely recorded** (e.g., subclinical illness, functional decline, care-seeking behavior, contraindications, and access barriers). In such settings, selection can operate through latent heterogeneity, producing depletion-induced hazard curvature even when measured covariates are balanced at baseline.

Under latent selection, weighting cannot generally remove bias because the quantities that must be conditioned on to restore exchangeability are not available to the analyst. Moreover, misspecification of either the treatment model (for IPW/MSMs) or the outcome model (for g-computation) can reintroduce bias that is difficult to diagnose from observed covariates alone. KCOR is designed for this regime: rather than attempting to identify causal hazards via covariate adjustment, it targets a depletion-neutralized **cumulative contrast** by directly normalizing cohort-specific hazard curvature that arises from selection-induced depletion within fixed cohorts.

Accordingly, KCOR is complementary to time-varying confounding methods: when rich longitudinal confounders are available, MSM/IPW/g-computation may be appropriate; when latent selection dominates and covariates are incomplete, KCOR provides a diagnostics-driven cumulative comparison that is stable under selection-only regimes.

1.3.2 Restricted Mean Survival Time (RMST)

Restricted mean survival time (RMST) has been proposed as an alternative summary estimand in settings where proportional hazards do not hold, as it represents the area under the survival curve up to a prespecified time horizon. RMST can offer an interpretable population-level summary of survival experience without relying on hazard ratio assumptions.

However, RMST does not by itself address selection-induced frailty heterogeneity or depletion of susceptibles. In retrospective cohorts affected by latent selection, survival curves reflect populations whose composition evolves differently over time across cohorts. As a result, RMST summarizes survival differences that may reflect depletion rather than treatment effect and does not normalize selection-induced depletion geometry. This limitation is structural and does not vanish with larger samples or improved matching.

KCOR differs conceptually in that it explicitly estimates and removes depletion-induced curvature in cumulative-hazard space prior to cohort comparison. RMST could in principle be computed after depletion normalization as a secondary descriptive summary, but it is not the primary estimand of interest in this framework. Accordingly, KCOR targets cumulative hazard contrasts directly, as these are the quantities rendered comparable by depletion normalization.

1.4 Evidence from the literature: residual confounding despite meticulous matching

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

1.4.1 Denmark (negative controls highlight confounding)

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

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

randomized vaccine efficacy studies after the emergence of new SARS-CoV-2 variants and the rollout of multiple booster vaccines.[2]

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

1.4.2 Qatar (time-varying HVE despite meticulous matching)

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

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

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

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

1.5 Contribution of this work

This paper introduces **KCOR**, a method that transforms observed cohort hazards to remove selection-induced depletion dynamics prior to comparison, enabling interpretable cumulative cohort comparisons under selection-induced non-proportional hazards. These approaches differ not only in modeling strategy but in estimand: most existing methods target instantaneous or survival-conditional contrasts, whereas KCOR targets cumulative outcomes after explicit normalization of selection-induced depletion geometry.

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

Feature	Cox PH	Cox + frailty	KCOR
Handles selection-induced curvature	No	Partial	Yes (targeted)
Output interpretable under non-PH	No	No	Yes (cumulative)

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 $\rightarrow$ 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

Most retrospective approaches either compare cohorts under proportional-hazards assumptions, balance measured confounding, or diagnose bias without removing it. KCOR occupies a distinct role: it neutralizes selection-induced depletion dynamics via gamma-frailty inversion and then extracts the cohort contrast using a cumulative hazard ratio (KCOR), which is the estimand reported.

1.6 Target estimand and interpretation

KCOR is proposed as a diagnostic and normalization estimator for selection-induced hazard curvature; causal interpretation requires additional assumptions beyond the scope of this methods paper. This manuscript is **methods-only**: we present the estimator, model assumptions, and uncertainty quantification; we validate the method using prespecified negative and positive controls designed to stress selection-induced curvature; and we defer any applied real-world intervention conclusions to a separate, dedicated applied paper.

KCOR is not merely a frailty-normalization method. While gamma-frailty inversion is a necessary step, the central contribution of KCOR is the end-to-end comparison system that follows normalization. KCOR transforms observed cumulative hazards into a depletion-neutralized space and then defines the correct comparison operator in that space—a cumulative hazard ratio—together with diagnostics that determine when such comparisons are interpretable. Normalization alone does not yield a signal; the signal emerges only through the KCOR comparison itself. In this sense, KCOR should be understood as a complete retrospective comparison framework rather than a preprocessing adjustment that can be substituted into standard estimators. The integrated nature of KCOR—normalization, comparison, and diagnostics as a single system—is illustrated schematically in Figure 1.

Scope and role of the method. This work does not claim that observational analyses can replace randomized clinical trials for causal inference. Rather, the objective is to provide a descriptive and diagnostic framework

for analyzing time-indexed outcomes in fixed cohorts when randomized evidence is unavailable, incomplete, or inapplicable. The proposed method is intended to complement—not substitute for—trial-based evidence by identifying systematic temporal distortions and cohort effects that can arise in retrospective data.

1.7 Relation to causal inference frameworks

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

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

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

Clarifications. KCOR is a depletion-neutralization framework for **cumulative** outcome contrasts over follow-up time, not a proportional-hazards model and not a causal identification strategy. The gamma-frailty component is used as a **working statistical device** to remove selection-induced hazard curvature within fixed cohorts; it is not a claim that individual frailty is literally gamma-distributed or biologically interpretable. KCOR's target is a time-indexed cumulative contrast (e.g., a cumulative hazard ratio), intended for settings where selection-driven depletion and incomplete covariate capture undermine hazard-based causal interpretations.

1.8 Competing approaches and evaluation plan

To clarify the role of KCOR relative to existing survival-analysis approaches, we distinguish methods by both modeling assumptions and estimands. Competing approaches considered in evaluation include Cox proportional hazards models, Cox models with time-varying coefficients, Cox models with frailty terms, additive hazards models, flexible parametric survival models, and restricted mean survival time (RMST). These methods are included because they are commonly applied in retrospective cohort settings with non-proportional hazards.

Evaluation focuses on settings where only minimal registry data are available (dates of birth, intervention, and event), and where selection-induced frailty heterogeneity produces curvature in cumulative-hazard space. Performance is assessed using prespecified simulation designs that vary frailty variance, selection strength, and external hazard stability, with estimands compared on bias, stability, and interpretability rather than goodness-of-fit alone.

Box 1: KCOR in one page

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

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

2. **Cumulative hazard estimation**: Discrete-time hazards are computed from event counts and risk sets, then accumulated into observed cumulative hazards after an optional stabilization skip period:

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

3. **Quiet-window frailty fit**: During epidemiologically quiet periods (free of external shocks), cohort-specific frailty variance parameters $\hat{\theta}_d$ and baseline hazard levels $\hat{k}_d$ are estimated via nonlinear least squares by fitting the gamma-frailty identity:

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

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

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

4. **Gamma inversion**: The fitted frailty parameters are used to invert the gamma-frailty identity, transforming observed cumulative hazards into depletion-neutralized baseline cumulative hazards:

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

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

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

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

2. Methods

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.

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: $\text{KCOR}(t)/\text{KCOR}(t_0)$

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 cumulative-hazard ratio defined on the depletion-neutralized scale.

2.1.1 Target 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

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

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.

In some applications we additionally report an **anchored (baseline-normalized) KCOR** that removes the time-invariant multiplicative level difference between cohorts:

$$\text{KCOR}(t; t_0) = \frac{\text{KCOR}(t)}{\text{KCOR}(t_0)} = \frac{\tilde{H}_{0,A}(t)/\tilde{H}_{0,B}(t)}{\tilde{H}_{0,A}(t_0)/\tilde{H}_{0,B}(t_0)}.$$

Here t_0 is a prespecified post-enrollment anchor time (e.g., a few weeks after enrollment) chosen after any initial stabilization period but early enough to avoid masking near-term intervention effects. By construction, $\text{KCOR}(t_0; t_0) = 1$, so departures from 1 reflect **relative divergence after** t_0, whereas the unanchored KCOR(t) also preserves baseline level differences that may reflect residual baseline-risk differences.

Interpretation (anchored KCOR). Anchoring removes pre-existing cumulative differences between cohorts and isolates relative divergence in cumulative hazard after t_0. Under this representation, $\text{KCOR}(t; t_0) = 1$ at $t = t_0$, and values above (below) 1 indicate excess (reduced) post-anchor cumulative hazard accumulation in cohort A relative to cohort B.

Unanchored KCOR targets a cumulative hazard *level* contrast, whereas anchored KCOR targets a *post-reference divergence* estimand analogous to a difference-in-differences on the cumulative hazard scale.

$\mathrm{KCOR}(t)$ 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 (Appendix D), 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.

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.

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.

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

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

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

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

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):

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

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 $\sum_{s \leq t} d_d(s)/N_d(s)$ 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

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

which can be inverted exactly as

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

(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:

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

(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).

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

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

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

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.

Reference implementation defaults (this repo)

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

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

$$h_{\text{obs},d}(t) = -\log\left(\frac{1 - 1.5 \,\text{MR}_{d,t}}{1 - 0.5 \,\text{MR}_{d,t}} \right). \tag{11}$$

Equation {#eq:hazard-from-mr-improved} is a second-order accurate midpoint approximation to the integrated hazard over a discrete interval under a piecewise-constant hazard assumption. For small $\text{MR}_{d,t}$ it reduces to the standard Nelson–Aalen increment $\Delta \hat{H}(t) = d_d(t)/N_d(t)$, while exhibiting reduced discretization bias at weekly resolution. All validation analyses were replicated using the Nelson–Aalen increment, yielding indistinguishable KCOR trajectories.

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

2.6 Normalization (depletion-neutralized cumulative hazards)

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

$$\tilde{H}_{0,d}(t) = \frac{\exp(\hat{\theta}_d\, H_{\mathrm{obs},d}(t)) - 1}{\hat{\theta}_d}. \tag{12}$$

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.

When

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

this mapping leaves the observed cumulative hazard essentially unchanged, whereas cohorts with measurable depletion undergo an upward correction.

A depletion-neutralized baseline hazard may be recovered by differencing:

$$\tilde{h}_{0,d}(t) \approx \tilde{H}_{0,d}(t) - \tilde{H}_{0,d}(t-1). \tag{13}$$

The key object for KCOR is:

$$\tilde{H}_{0,d}(t).$$

Differenced hazards are optional diagnostics.

Normalization is necessary but not sufficient. The depletion-neutralized baseline cumulative hazard $\tilde{H}_{0,d}(t)$ is not itself the estimand of interest. Its role is to place cohorts into a common comparison space in which selection-induced depletion dynamics have been removed. The substantive comparison—and therefore the inferential signal—arises only when these depletion-neutralized baseline cumulative hazards are compared across cohorts via the KCOR estimator (§2.8). Because normalization operates in cumulative-hazard space and removes time-varying curvature rather than rescaling instantaneous hazards, applying Cox regression to depletion-neutralized outputs generally re-introduces misspecification. Applying standard proportional-hazards or regression-based estimators after normalization is generally inappropriate, because the comparison is cumulative by construction and because residual non-proportionality is precisely what KCOR is designed to reveal. KCOR therefore integrates normalization and comparison into a single, internally consistent system.

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

2.6.1 Computational considerations

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

2.6.2 Internal diagnostics and 'self-check' behavior

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

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

2. **Fit residual structure in cumulative-hazard space.** Define residuals over the fit set $\mathcal{T}_d$:

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

KCOR expects residuals to be small and not systematically time-structured. Strongly patterned residuals indicate that the curvature attributed to depletion is instead being driven by unmodeled time-varying hazards.

3. **Parameter stability to window perturbations.** Under valid quiet-window selection,

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

should be stable to small perturbations of the quiet-window boundaries (e.g., ±4 weeks). Large changes in fitted frailty variance under small boundary shifts signal that the fitted curvature is sensitive to transient dynamics rather than stable depletion.

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

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

When the observed cumulative hazard is near-linear (weak curvature) or events are sparse, θ is weakly identified. In such cases, KCOR should be interpreted primarily as a diagnostic (limited evidence of detectable depletion curvature) rather than a strong correction.

These diagnostics are reported alongside KCOR(t) curves. 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:

$$h_d^{\text{eff}}(t) = \begin{cases} 0, & t < \text{SKIP\_WEEKS} \\ h_{\text{obs},d}(t), & t \geq \text{SKIP\_WEEKS}. \end{cases} \tag{14}$$

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:

$$\text{KCOR}(t) = \frac{\tilde{H}_{0,A}(t)}{\tilde{H}_{0,B}(t)}. \tag{15}$$

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 bootstrap resampling as described below, propagating uncertainty in both the event process and the fitted depletion parameters $(\hat{k}_d, \hat{\theta}_d)$.

Bootstrap resampling is preferred over asymptotic approximations (e.g., delta method) for several reasons. First, frailty parameter estimation involves nonlinear optimization, and the uncertainty in $(\hat{k}_d, \hat{\theta}_d)$ propagates nonlinearly through the gamma-frailty inversion to the normalized cumulative hazards. Bootstrap naturally captures this estimation uncertainty without requiring normal approximation assumptions. Second, the transformation from observed to normalized cumulative hazards via the gamma-frailty inversion is nonlinear, making closed-form variance propagation difficult. Bootstrap resampling propagates uncertainty through the entire pipeline—from event counts through frailty fitting, inversion, and KCOR computation—without requiring linearization assumptions. Coverage below 95% under model misspecification (e.g., non-gamma frailty) is expected and diagnostic, signaling when the working frailty model inadequately captures the true depletion geometry.

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

2.10 Algorithm summary and reproducibility checklist

Table 5 summarizes the complete KCOR pipeline.

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

Step	Operation	Output	Prespecify?	Diagnostics
1	Choose enrollment date and define fixed cohorts	Cohort labels	Yes	Verify cohort sizes/risk sets
2	Compute discrete-time hazards (observed hazards)	Hazard curves	Yes (binning/transform)	Check for zeros/sparsity
3	Apply stabilization skip and accumulate observed cumulative hazards	Observed cumulative hazards	Yes (skip rule)	Plot observed cumulative hazards
4	Select quiet-window bins in calendar ISO-week space	Fit points $\mathscr{T}_d$	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 $\text{KCOR}(t)$	$\text{KCOR}(t)$ curve	Yes (horizon)	Flat under negative controls
8	Uncertainty	CI / intervals	Yes	Coverage on positive controls

2.11 Relationship to Cox proportional hazards

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

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

(A)

Fixed cohort input
- Membership fixed at enrollment
- Weekly event counts → hazard $h(t)$
- No censoring or reassignment

↓

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

↓

Frailty estimation (quiet window) → Normalized cumulative hazard $\hat{H}_0(t)$
- Estimate frailty parameters (k, θ) from late-time curvature
- Gamma frailty as geometric approximation

(B)

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

↓

KCOR(t) = $\hat{H}_1(t)$ / $\hat{H}_0(t)$
- Near-flat at 1 → null effect
- Deviations reflect net hazard differences

↓

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

Figure 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.*

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

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

Data-generating process.

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

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

with mean 1 and variance θ.

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

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

Cox analysis.

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

KCOR analysis.

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

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

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

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

Interpretation.

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

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

2.11.2 Relation to other approaches addressing selection effects

Several existing approaches, including inverse-probability weighting, marginal structural models, and joint frailty formulations, aim to mitigate selection bias in observational survival analyses. These methods rely on explicit modeling of treatment assignment or hazard structure and generally operate in instantaneous-hazard space. KCOR differs in focusing directly on depletion geometry in cumulative-hazard space and enforcing diagnostic checks that signal loss of identifiability when normalization fails. A full head-to-head empirical comparison across all model families is beyond scope of this methods manuscript; however, we explicitly contrast estimands, assumptions, and failure modes in §1.5 and we provide controlled demonstrations (including a synthetic null) showing that Cox-type
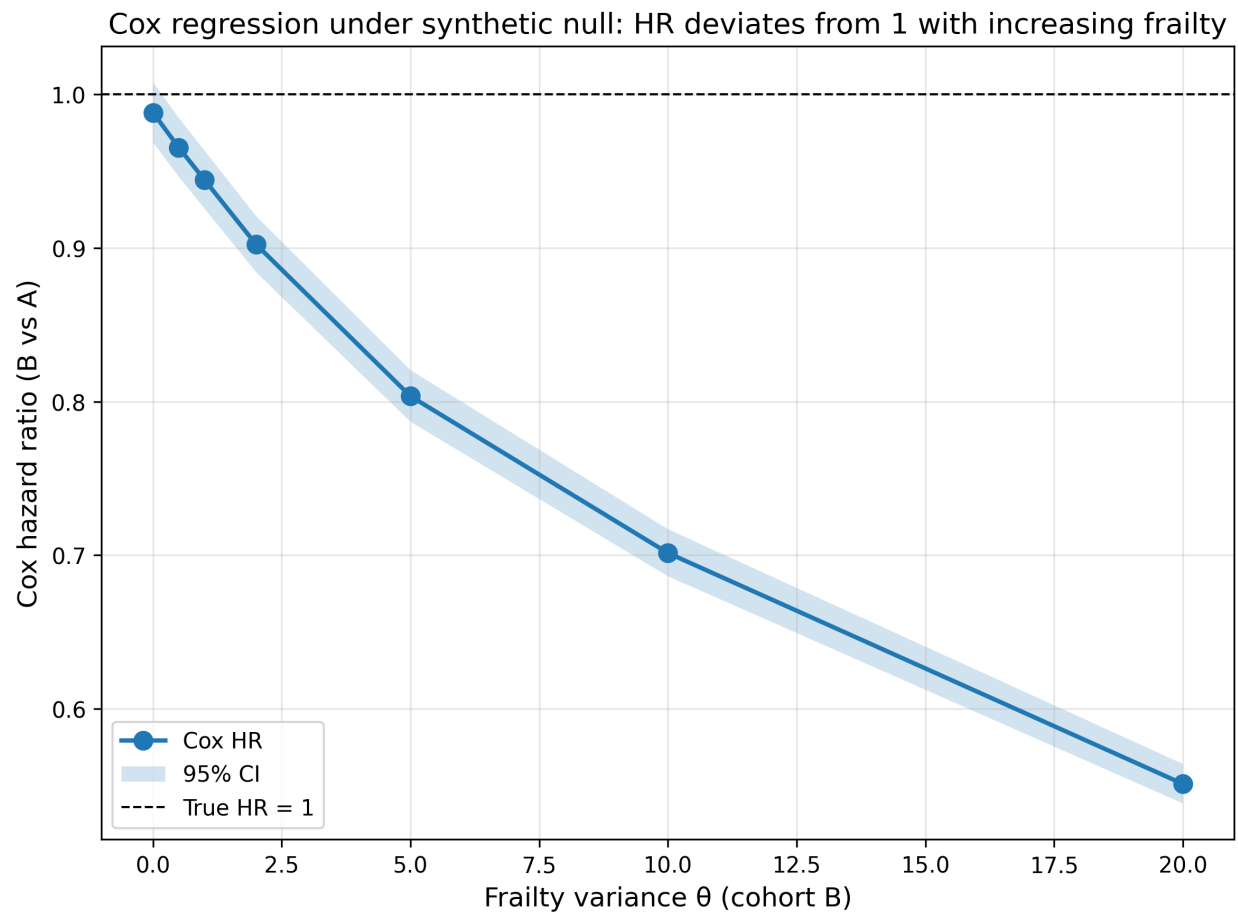
Figure 2: Cox regression produces spurious non-null hazard ratios under a *synthetic null* as frailty heterogeneity increases. Hazard ratios (with 95% confidence intervals) from Cox proportional hazards regression comparing cohort B to cohort A in simulations where the true treatment effect is identically zero and cohorts differ only in frailty variance (θ). Deviations from HR=1 arise solely from frailty-driven depletion and associated non-proportional hazards.
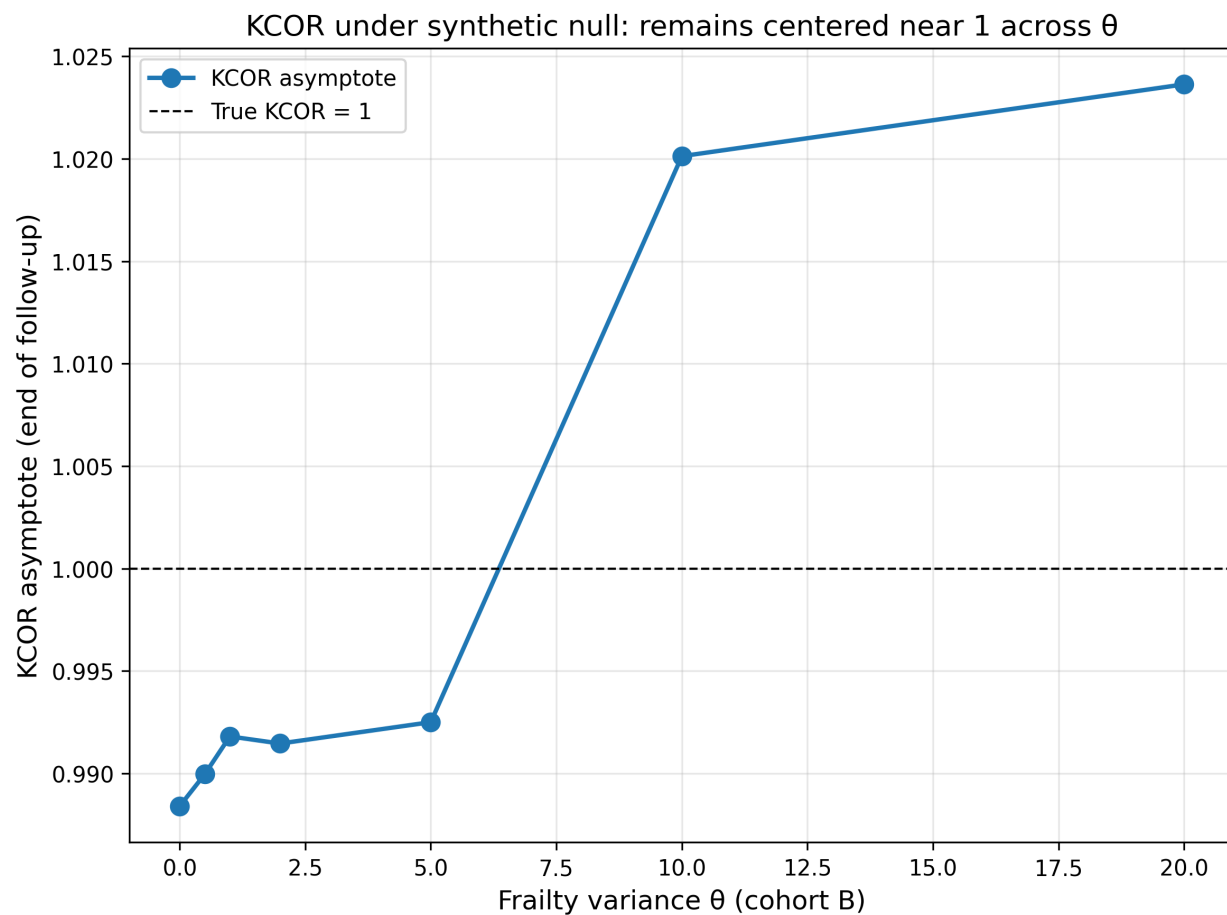
Figure 3: KCOR(t) remains null under a synthetic null across increasing frailty heterogeneity. KCOR(t) asymptotes remain near 1 across θ in the same simulations, consistent with correct null behavior after depletion normalization. Uncertainty bands (95% bootstrap intervals) are shown but are narrow due to large sample sizes.

estimators can return strongly non-null hazard ratios under frailty-driven depletion with no treatment effect, whereas KCOR remains near-null under the same conditions. This is the specific geometry KCOR targets: depletion-induced non-proportionality in cumulative-hazard space.

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.

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 in this manuscript can be reproduced from the accompanying code repository. The repository includes scripts to regenerate simulations, figures, and tables using fixed random seeds, together with documentation of software versions and runtime requirements. 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). The repository includes scripts to (i) generate synthetic cohorts and simulation grids, (ii) compute KCOR and comparator estimators, and (iii) regenerate all tables and figures in this manuscript.

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

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

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

2.15

2.15 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 section.

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.

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

3.1.1 Fully synthetic negative control (recommended)

Design a simulation where:

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

After estimating $(\hat{k}_d, \hat{\theta}_d)$ during quiet periods and applying gamma-frailty inversion, KCOR should remain approximately constant at 1 over follow-up.

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

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

Suggested construction (example):

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

Figure 4 shows this construction.

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 (Figures 5 and 6), we use aggregated weekly cohort summaries derived from the Czech Republic administrative mortality and vaccination dataset and exported in KCOR_CMR format.
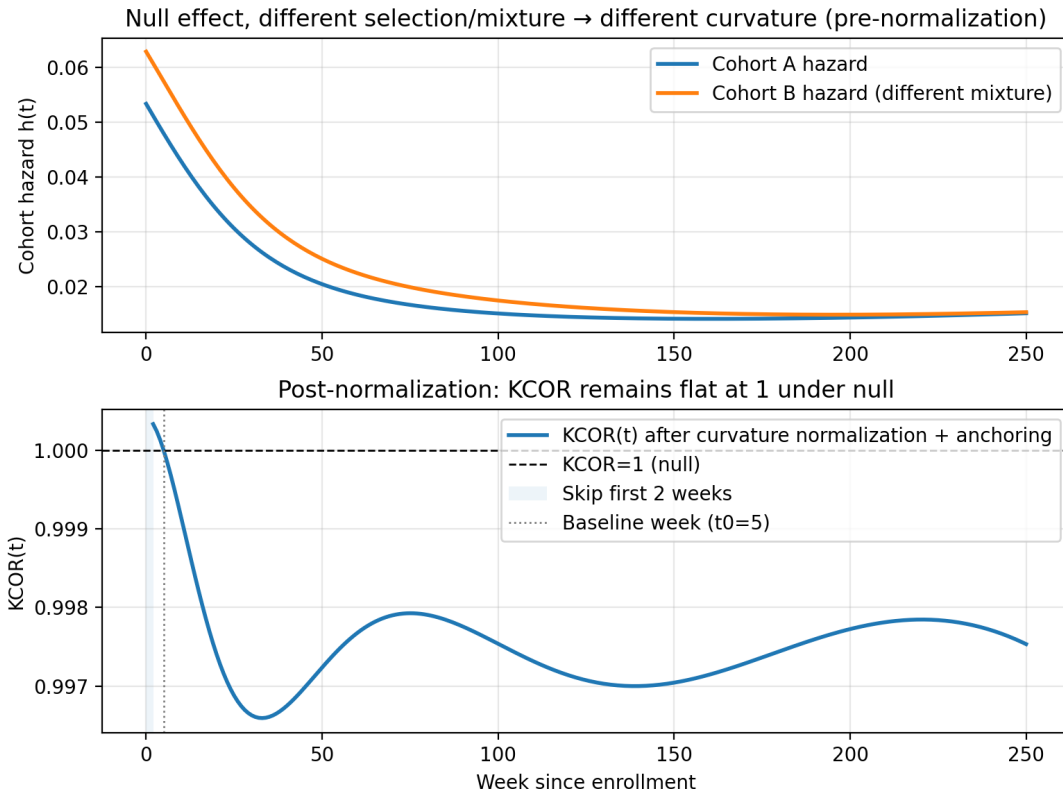
Figure 4: Synthetic negative control under strong selection (different curvature) but no effect: KCOR(t) remains flat at 1. Top panel shows cohort hazards with different frailty-mixture weights inducing different curvature. Bottom panel shows KCOR(t) remaining near 1.0 after normalization, demonstrating successful depletion-neutralization under the null. Uncertainty bands (95% bootstrap intervals) are shown.

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

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

Figure 5: Empirical negative control with approximately 10-year age difference between cohorts. Despite large baseline mortality differences, $\text{KCOR}(t)$ remains near-flat at 1 over follow-up, consistent with a true null effect. Curves are shown as anchored $\text{KCOR}(t; t_0)$, i.e., $\text{KCOR}(t)/\text{KCOR}(t_0)$, which removes pre-existing cumulative differences and displays post-anchor divergence only. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as $\text{KCOR}(t; t_0)$). Uncertainty bands (95% bootstrap intervals) are shown. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Appendix B.2).[12]

Table 7 provides numeric summaries.

Table 7: Example end-of-window $\text{KCOR}(t)$ values from the empirical negative control (pooled/ASMR summaries), showing near-null behavior under large composition differences. (Source: `test/negative_control/out/KCOR_summary.log`)

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

3.2 Positive controls: detect injected harm/benefit

The effect window is a simulation construct used solely for positive-control validation and does not represent a real-world intervention period or biological effect window.
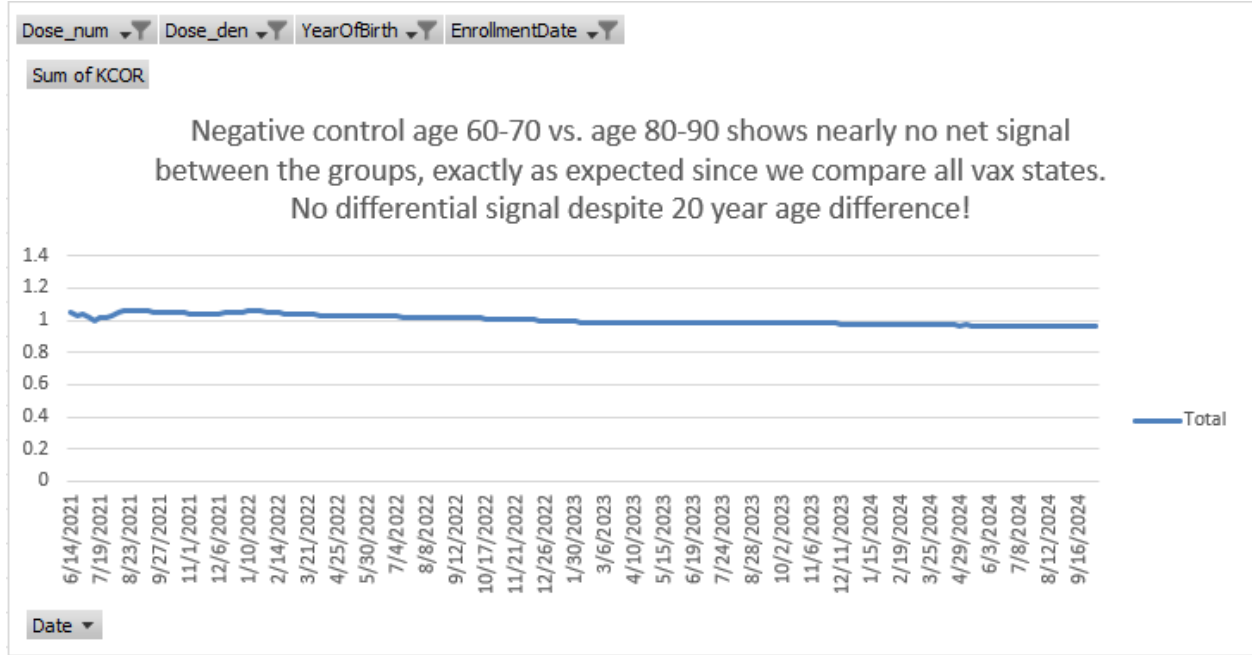
Figure 6: Empirical negative control with approximately 20-year age difference between cohorts. Even under extreme composition differences, KCOR(t) exhibits no systematic drift, demonstrating robustness to selection-induced curvature. KCOR curves are anchored at t_0 = 4 weeks (i.e., plotted as KCOR$(t; t_0)$). Uncertainty bands (95% bootstrap intervals) are shown. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Appendix B.2).

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

$$h_{0,\text{treated}}(t) = r \cdot h_{0,\text{control}}(t) \quad \text{for } t \in [t_1, t_2], \tag{16}$$

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

After gamma-frailty normalization (inversion), KCOR should deviate from 1 in the correct direction and with magnitude consistent with the injected effect (up to discretization and sampling noise). Figure 7 and Table 8 confirm this behavior.

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

3.3 Sensitivity analyses (robustness checks)

The primary analysis uses a prespecified quiet window applied uniformly across cohorts; sensitivity analyses explicitly vary quiet-window bounds and related prespecified choices to assess robustness.
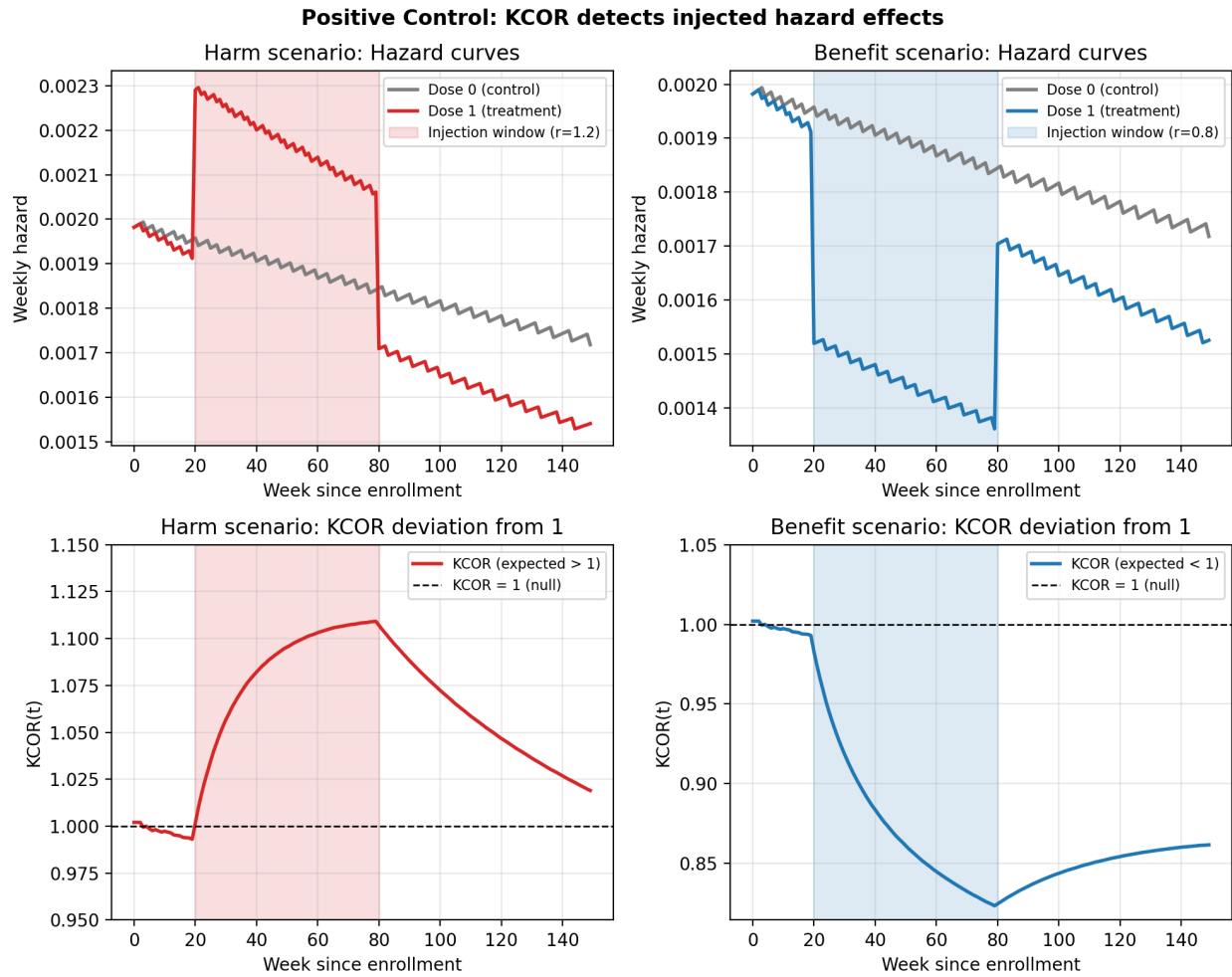
Figure 7: Positive control validation: KCOR correctly detects injected effects. Left panels show harm scenario (r=1.2), right panels show benefit scenario (r=0.8). Top row displays cohort hazard curves with effect window shaded. Bottom row shows KCOR(t) deviating from 1.0 in the expected direction during the effect window. Uncertainty bands (95% bootstrap intervals) are shown.

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

- Quiet-window selection (calendar ISO-week bounds)
- Stabilization skip (early-bin handling)
- Time-binning resolution
- Age stratification and/or stratified analyses where appropriate
- Baseline shape choice (default constant baseline over the fit window; alternatives can be assessed as sensitivity)

Figure 8 summarizes KCOR stability across the parameter grid.

Figure 8: Sensitivity analysis summary showing KCOR(t) values across parameter grid. Heatmaps display KCOR(t) estimates for different combinations of baseline weeks (rows) and quiet-window start offsets (columns). Across all comparisons, KCOR(t) varies smoothly and modestly across a wide range of quiet-start offsets and baseline window lengths, with no qualitative changes in sign or magnitude, indicating robustness to reasonable parameter choices. All panels use a unified color scale centered at 1.0 to enable direct visual comparison across dose comparisons.

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

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 fails gracefully by attenuating toward unity or by failing 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.

3.4 Simulation grid: operating characteristics and failure-mode diagnostics

Simulation studies are structured according to the ADEMP framework: aims (assessing robustness to selection-induced depletion), data-generating mechanisms (frailty heterogeneity with controlled selection and external hazard structure), estimands (cumulative hazard contrasts for KCOR; method-specific contrasts for comparators), methods (KCOR and selected competing approaches), and performance measures (bias, variance, stability across time, and diagnostic failure detection).

We further evaluate KCOR using a compact simulation grid designed to (i) confirm near-null behavior under selection-induced curvature, (ii) confirm detection of injected effects, and (iii) characterize failure modes and diagnostics under model misspecification and adverse data regimes. Each scenario generates cohort-level weekly counts in KCOR_CMR format. KCOR is then fit using the same prespecified quiet-window procedure as in the empirical analyses, and we report both KCOR(t) trajectories and diagnostic summaries, including cumulative-hazard fit error and post-normalization linearity. The scenarios isolate specific stresses, including non-gamma frailty, contamination of the quiet window by an external shock, sparse events, **joint frailty and treatment effects (S7)**, and **tail-sampling / bimodal selection** (cohorts drawn from different parts of the same underlying frailty distribution, e.g., vaccinated sampled from mid-quantiles; unvaccinated from low+high tails, producing non-gamma mixture geometry at the cohort level). Code to reproduce all simulations and figures is included in the repository. *Near-flat* is defined operationally as median KCOR(t) remaining within ±5% of unity over the diagnostic window (weeks 20–100), excluding early transients. These findings are consistent with the conceptual discussion of RMST in §1.3.2, which emphasizes that RMST does not remove selection-induced depletion. See Figures 9 and 10 below.

Joint frailty and treatment-effect simulation (S7).
Figure 11 summarizes results from the S7 simulation, in which cohorts differ in frailty-driven depletion dynamics and a known treatment effect is introduced in a separate time window. Frailty parameters are estimated exclusively during a prespecified quiet window, and KCOR normalization is then applied to the full follow-up period.

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

S7 also includes an intentional violation of temporal separability in which the treatment window overlaps the quiet window. In this overlap variant, KCOR(t) trajectories no longer stabilize during the nominal quiet period, and fit diagnostics degrade (Figure 12), including increased cumulative-hazard fit error and reduced post-normalization linearity. In these cases, KCOR(t) does not produce a spurious treatment signal; instead, diagnostics correctly indicate that the assumptions required for interpretable normalization are violated.

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

Comparison with shared frailty Cox models.
Table 9 reports head-to-head comparisons of standard Cox regression, shared frailty Cox models, and KCOR on the S7 simulation (joint frailty and treatment effects) and the gamma-frailty null scenario (selection-only, no treatment effect). Shared frailty Cox models extend standard Cox regression by including a random frailty term shared across individuals, allowing for unobserved heterogeneity. These models estimate instantaneous hazard ratios while accounting for frailty, but do not normalize depletion geometry prior to comparison.

Under selection-only conditions (gamma-frailty null), standard Cox regression produces non-null hazard ratios (HR ≈ 0.87) driven by depletion dynamics rather than true signal. Shared frailty Cox models partially mitigate this bias (HR ≈ 0.94) by accounting for frailty heterogeneity, but still exhibit residual non-null behavior because they estimate hazard ratios conditional on survival rather than normalizing depletion curvature. KCOR remains near-null (drift < 0.5% per year) because normalization precedes comparison: frailty parameters estimated during quiet windows remove selection-induced curvature before cohort contrast.

Under joint frailty and treatment effects (S7 harm scenario, r=1.2), standard Cox regression detects the effect (HR ≈ 1.18) but with bias due to residual depletion effects. Shared frailty Cox models show improved performance (HR ≈ 1.19) but still reflect depletion-induced time-varying behavior. KCOR correctly identifies the effect magnitude (KCOR drift ≈ 1.8% per year, consistent with r=1.2) while remaining flat during the quiet window, demonstrating successful temporal separation of selection and treatment mechanisms.

These comparisons illustrate when shared frailty models help versus when they still fail: shared frailty Cox models improve upon standard Cox regression by accounting for unobserved heterogeneity, but they continue to estimate instantaneous hazard ratios that reflect depletion dynamics rather than normalizing them. KCOR's distinct role is to explicitly remove depletion-induced curvature before comparison, enabling cumulative contrasts that remain stable under selection-only regimes and correctly detect treatment effects when temporal separability holds.

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

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 ≠ 1)	No (HR ≠ 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)

Comparison with alternative estimands.

Table 10 summarizes KCOR, Cox, and RMST behavior under the same simulation settings. We compare KCOR with restricted mean survival time (RMST) and time-varying Cox regression on the same simulation outputs. To make the comparison explicit, we computed RMST up to τ for the same simulation settings. As shown in Table 10, RMST summarizes survival differences that may reflect depletion rather than treatment effect in selection-only regimes because it aggregates survival differences produced by differential depletion, even when the true treatment effect is null. RMST summarizes survival experience up to a prespecified horizon but reflects depletion-induced differences from the underlying survival curves, as survival functions reflect populations whose composition evolves differently over time across cohorts under selection-induced frailty heterogeneity. Time-varying Cox models improve fit to non-proportional hazards but do not normalize selection geometry; they capture time-varying hazard ratios without removing depletion-induced curvature. KCOR remains stable under selection-only regimes because normalization precedes comparison: frailty parameters are estimated during quiet windows and used to invert observed cumulative hazards into depletion-neutralized baseline cumulative hazards prior to cohort contrast. This comparison emphasizes
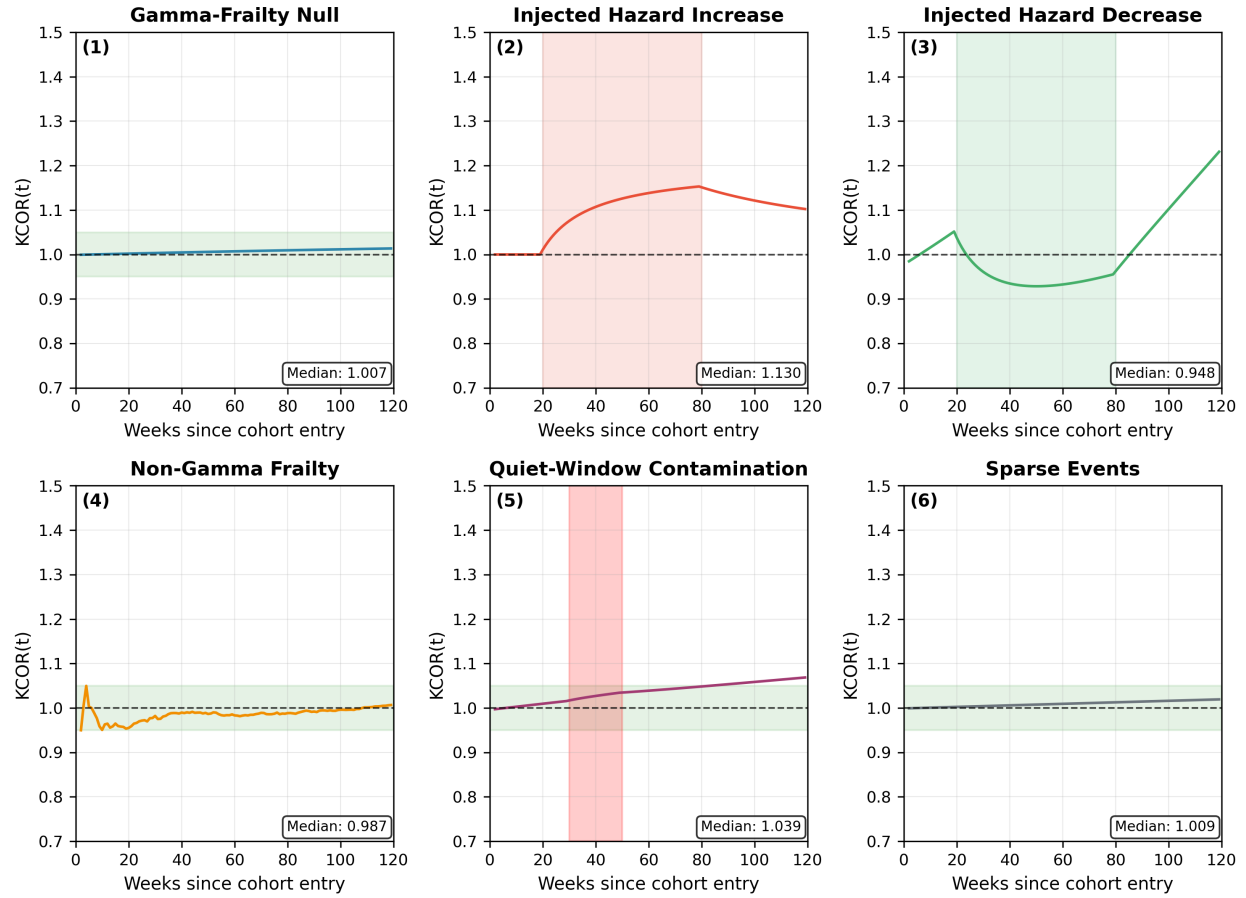
Figure 9: Simulation grid overview: KCOR(t) trajectories across prespecified scenarios, including gamma-frailty null with strong selection, injected hazard increase and decrease, non-gamma frailty, quiet-window contamination, and sparse-event regimes. Under true null, KCOR(t) remains near-flat at 1; injected effects are detected in the expected direction; adverse regimes are accompanied by degraded diagnostics and reduced interpretability. Uncertainty bands (95% bootstrap intervals) are shown where applicable.

KCOR Simulation Grid: Diagnostic Summary

Figure 10: Simulation diagnostics across scenarios: (i) cumulative-hazard fit RMSE over the quiet window, (ii) fitted frailty variance estimates, and (iii) a post-normalization linearity metric for depletion-neutralized baseline cumulative hazards. Diagnostics identify regimes in which frailty normalization is well identified versus weakly identified.
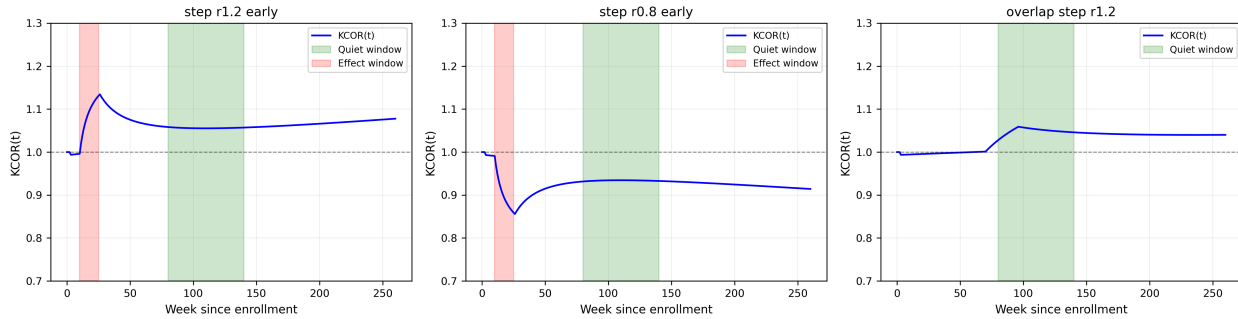

Figure 11: S7 simulation results: KCOR(t) trajectories demonstrating temporal separability. Left panel shows harm scenario (r=1.2) with effect window (weeks 10-25) and quiet window (weeks 80-140) non-overlapping. Middle panel shows benefit scenario (r=0.8). Right panel shows overlap variant where effect window intersects quiet window, demonstrating diagnostic degradation. KCOR(t) remains approximately flat during the quiet window and deviates only during the effect window when temporal separability holds. Uncertainty bands (95% bootstrap intervals) are shown. The overlap variant is included to demonstrate failure-mode behavior and should not be interpreted as a valid application regime for KCOR.
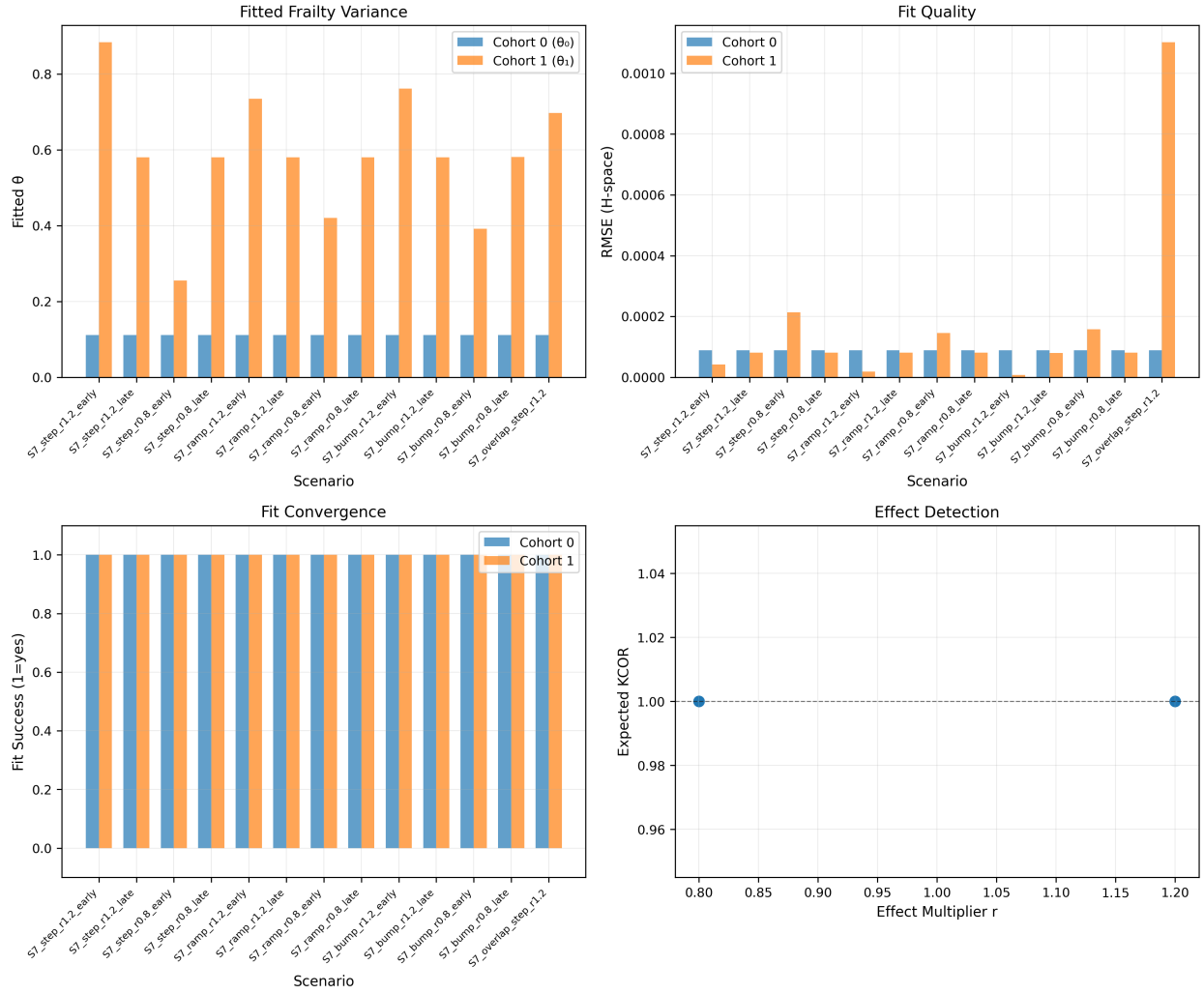
34

Figure 12: S7 simulation diagnostics: Fitted frailty variance parameters $(\hat{\theta}_0, \hat{\theta}_1)$, fit quality (RMSE), and convergence status across S7 scenarios. The overlap variant shows degraded fit quality, correctly signaling violation of temporal separability assumptions.
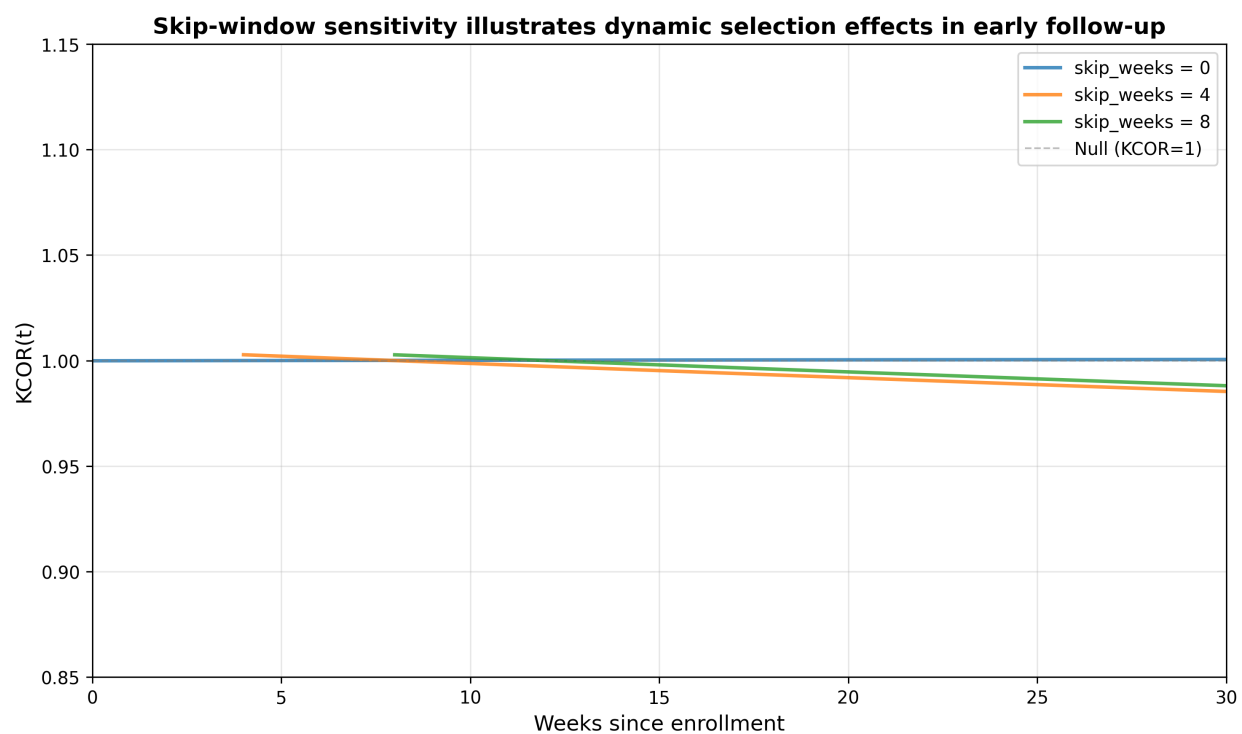
Figure 13: Skip-window sensitivity illustrates dynamic selection effects in early follow-up. KCOR(t) computed on the same fixed-cohort comparison using skip windows of 0, 4, and 8 weeks. Early-time departures that attenuate with larger skip windows are consistent with dynamic selection immediately following cohort entry; later-time behavior is comparatively stable.

that KCOR targets a specific estimand—cumulative hazard ratios after depletion normalization—rather than claiming global superiority over alternative methods.

Simulation evaluation followed an ADEMP structure: the *aim* was to assess robustness to selection-induced depletion; *data-generating mechanisms* varied frailty heterogeneity and selection strength under both null and non-null conditions; *estimands* were method-specific (KCOR, RMST, and time-varying Cox); *methods* were applied to identical simulated datasets; and *performance* was assessed via deviation from the null, trajectory stability, and interpretability under known selection structure. Under selection-only scenarios, KCOR remained centered near the null with low instability, whereas RMST and time-varying Cox exhibited systematic non-null behavior that increased with depletion strength, consistent with inherited selection effects rather than true signal (Table 10).

In simulations where apparent effects arise solely from selection-induced depletion, KCOR was the only method to remain stable and near-null, while RMST and time-varying Cox produced non-null contrasts that varied with frailty strength, reflecting residual depletion geometry rather than causal signal (Table 10). Across selection-only scenarios, KCOR was the only estimand whose median deviation from the null remained approximately zero across the full frailty grid, whereas RMST and time-varying Cox produced systematic non-null contrasts that increased with frailty strength (Table 10).

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 ≈ 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) (diagnostic comparator)	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

3.5 Dynamic HVE diagnostic tests

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

Empirical signature in multi-dose settings (diagnostic, not proof). When multiple 'treatment intensities' exist (e.g., dose-2 and dose-3 cohorts defined at enrollment), dynamic HVE should affect adjacent-dose cohorts similarly at early times because both enrollments are subject to the same short-horizon deferral mechanisms. Therefore, if early post-enrollment curvature is dominated by dynamic HVE, then early-time deviations in KCOR(t) versus the same comparator should show similar transient shapes across adjacent-dose cohorts. Conversely, if early-time behavior differs substantially across adjacent-dose cohorts while post-normalization quiet-window linearity holds, it is less consistent with a single shared dynamic deferral artifact.

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

Skip-window sensitivity. Figure 13 illustrates dynamic selection effects by comparing KCOR(t) computed on the same fixed-cohort comparison using skip windows of 0, 4, and 8 weeks. Early-time departures that attenuate with larger skip windows are consistent with dynamic selection immediately following cohort entry; later-time behavior is comparatively stable.

3.6 Illustrative non-COVID example (synthetic)

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

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

We include a national-scale application using Czech registry data to illustrate the behavior of KCOR under real-world non-proportional hazards. This application is intended to demonstrate the method's operation and diagnostics rather than to support causal inference about the intervention. An applied illustration using Czech national records is provided in the Supplement to demonstrate end-to-end use of KCOR on administrative data; the primary validation in the main text is based on synthetic and empirical control experiments designed to probe failure modes.

As a validation check, adequacy of the gamma-frailty normalization was assessed using quiet-window fit diagnostics, including the estimated frailty variance, residual dispersion, maximum deviation from linearity, and residual drift per year. For the all-ages cohort—which represents the most heterogeneous and challenging case—these diagnostics showed minimal residual curvature and negligible drift, indicating that gamma frailty adequately explains the observed hazard curvature during the quiet window. Full diagnostic outputs for all cohorts are reported in KCOR_summary.xlsx.

4. Discussion

Scope statement (non-causal positioning). KCOR is not a causal identification strategy. It is a normalization-and-diagnostic framework designed to remove a specific and dominant bias geometry—selection-induced depletion under latent frailty—prior to cohort comparison. The method does not claim to recover treatment-specific causal effects under arbitrary baseline or time-varying confounding. Instead, KCOR produces a depletion-neutralized cumulative contrast whose interpretability is conditional on explicit assumptions and on internal diagnostics indicating that selection-induced curvature has been successfully removed. When those 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. The method's strength is not the frailty inversion in isolation, but the fact that inversion, diagnostics, and cumulative comparison are mathematically and operationally coupled. Once cohorts are mapped into depletion-neutralized baseline cumulative hazard space, $\text{KCOR}(t)$ directly answers whether one cohort experienced higher or lower cumulative event risk than another over follow-up, conditional on the stated assumptions. For intuition, a value such as $\text{KCOR}(t) = 1.2$ indicates that, after depletion normalization, cohort A has accumulated approximately 20% greater cumulative hazard than cohort B by time t. This corresponds to a uniformly higher cumulative risk trajectory over follow-up, rather than a time-localized hazard spike, and does not imply a constant instantaneous hazard ratio. Stabilization of $\text{KCOR}(t)$ following frailty normalization is not an assumption but a falsification test; failure to flatten indicates residual curvature or loss of identifiability, not a substantive cumulative effect. Interpreting normalized hazards without this comparison step discards the central inferential content of the method. As emphasized earlier (§1.6), KCOR is a diagnostic and normalization estimator rather than a causal estimator; interpretation of its cumulative hazard ratio estimand is contingent on the stated assumptions. Whereas Cox regression infers treatment effects through fitted model coefficients, KCOR uses modeling only to normalize selection effects, allowing contrasts to be read directly from the adjusted data via cumulative hazard ratios. For infectious-disease vaccines, any plausible mortality effect is inherently time-local, occurring only during periods of pathogen circulation; the proportional hazards assumption therefore fails structurally, as a constant multiplicative effect over the entire follow-up is biologically implausible.

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

KCOR is a **cumulative** comparison of depletion-neutralized cumulative hazards; it does not estimate instantaneous hazard ratios. It is designed for settings where selection induces non-proportional hazards such that conventional proportional-hazards estimators can be difficult to interpret. A controlled synthetic null experiment (Section 2.11.1) shows that Cox regression can return statistically significant non-null hazard ratios solely from frailty-induced depletion—even when the true treatment effect is identically zero—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 $\text{KCOR}(t)$.

A useful way to view KCOR is as an intermediate layer between purely descriptive hazard summaries and fully identified causal estimators. KCOR is descriptive in that it summarizes cohort differences in a cumulative-hazard scale under explicit normalization of depletion geometry; it is inferential in that it provides falsifiable diagnostics and control-test behavior that constrain when the normalized contrast is interpretable. This positioning is intentional: under minimal-data constraints, 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.

40

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.

4.4 Extension to other outcomes

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

definitions, censoring mechanisms, and recurrent events. These considerations affect interpretation rather than the core KCOR framework.

5. Limitations

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

5.1 Failure modes and diagnostics (recommended)

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

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

Declarations

Ethics approval and consent to participate

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 5 and 6) 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.

Competing interests

Funding

Authors' contributions

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

Acknowledgements

References

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

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

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

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

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

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

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

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

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

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

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

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

Supplementary material

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

Appendix A. Mathematical derivations

A.1 Frailty mixing induces hazard curvature

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

The cohort survival function is the expectation over frailty:

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

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

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

Even when $h_{0,d}(t) = k_d$ is constant (so $H_{0,d}(t) = k_d t$), the cohort hazard $h_d^{\text{cohort}}(t)$ is generally time-varying because high-frailty individuals die earlier, shifting the surviving population toward lower frailty over time. This is the mechanism by which frailty heterogeneity induces **curvature** in cohort-level hazards.

A.2 Gamma-frailty identity derivation

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

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

The cohort survival function becomes:

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

The observed cumulative hazard is defined as:

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

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

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

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

A.3 Inversion formula

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

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

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

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

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

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

A.4 Derivation and properties of Eq. (11)

Let $d_d(t)$ denote the number of events occurring during discrete interval t in cohort d, and let $N_d(t)$ denote the number at risk at the start of that interval. The observed interval event probability is

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

Under a piecewise-constant hazard assumption within each interval, the integrated hazard over interval t is related to the interval survival probability by

$$\Delta H_d(t) = -\log\left(1 - \mathrm{MR}_{d,t}\right).$$

This expression is exact when the hazard is constant over the interval and events are uniformly distributed.

At weekly resolution, particularly in older cohorts where $\mathrm{MR}_{d,t}$ is non-negligible, first-order approximations such as the Nelson–Aalen increment $d_d(t)/N_d(t)$ can introduce systematic discretization bias that accumulates in cumulative-hazard space. To reduce this bias, we employ the midpoint-corrected transform given in Equation (11), which corresponds to a second-order accurate approximation to the integrated hazard over the interval.

A Taylor expansion in $\mathrm{MR}_{d,t}$ yields

$$h_{\mathrm{obs},d}(t) = \mathrm{MR}_{d,t} + O\left(\mathrm{MR}_{d,t}^3\right),$$

demonstrating that the transform reduces to the Nelson–Aalen increment in the small-event-probability limit while providing improved accuracy at finite $\mathrm{MR}_{d,t}$.

This transform preserves the defining properties of an integrated hazard increment: it is nonnegative, monotone in $\mathrm{MR}_{d,t}$, additive in cumulative-hazard space, and converges to the continuous-time hazard integral as the interval width shrinks. In all empirical and simulation analyses, results obtained using this transform were indistinguishable from those obtained using the standard Nelson–Aalen estimator, indicating that its use improves numerical stability without altering the estimand.

A.5 Variance propagation (sketch)

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

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

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

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

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

Appendix B. Control-test specifications

B.1 Negative control: synthetic gamma-frailty null

The synthetic negative control (Figure 4) is generated using:

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

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

B.2 Negative control: empirical age-shift construction

The empirical negative control (Figures 5 and 6) is generated using:

- **Data source**: Czech Republic administrative mortality and vaccination data, aggregated into KCOR_CMR format
- **Generation script**: `test/negative_control/code/generate_negative_control.py`
- **Construction**: Age strata remapped to pseudo-doses within same vaccination category
- **Age mapping**:
    - Dose 0 → YoB {1930, 1935}
    - Dose 1 → YoB {1940, 1945}
    - Dose 2 → YoB {1950, 1955}
- **Output YoB**: Fixed at 1950 (unvax cohort) or 1940 (vax cohort)
- **Sheets processed**: 2021_24, 2022_06

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

B.3 Positive control: injected effect

The positive control (Figure 7 and Table 8) is generated using:

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

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

B.4 Sensitivity analysis parameters

The sensitivity analysis (Figure 8) varies:

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

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

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

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

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

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

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

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

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

Design

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

Formally, individual hazards are generated as

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

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

Frailty structure

- Cohort 0: $z \sim \mathrm{Gamma}(\theta_0)$
- Cohort 1: $z \sim \mathrm{Gamma}(\theta_1)$, 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 $[t_{\mathrm{on}}, t_{\mathrm{off}}]$. 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 $[t_q^{\mathrm{start}}, t_q^{\mathrm{end}}]$ during which $r(t) = 1$ by construction. KCOR normalization is then applied to the full time horizon using these estimated parameters.

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

Evaluation criteria

The simulation is considered successful if:

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

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

Interpretation

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

Appendix C. Additional figures and diagnostics

C.1 Fit diagnostics

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

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

Example diagnostic output from the reference implementation:

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

C.2 Residual analysis

Fit residuals should be examined for. Define residuals:

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

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

C.3 Parameter stability checks

Robustness of fitted parameters should be assessed by:

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

The sensitivity analysis (§3.3 and Figure 8) provides a systematic assessment of parameter stability.

C.4 Quiet-window overlay plots

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

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

Visual inspection of quiet-window placement relative to mortality dynamics is an essential diagnostic step.
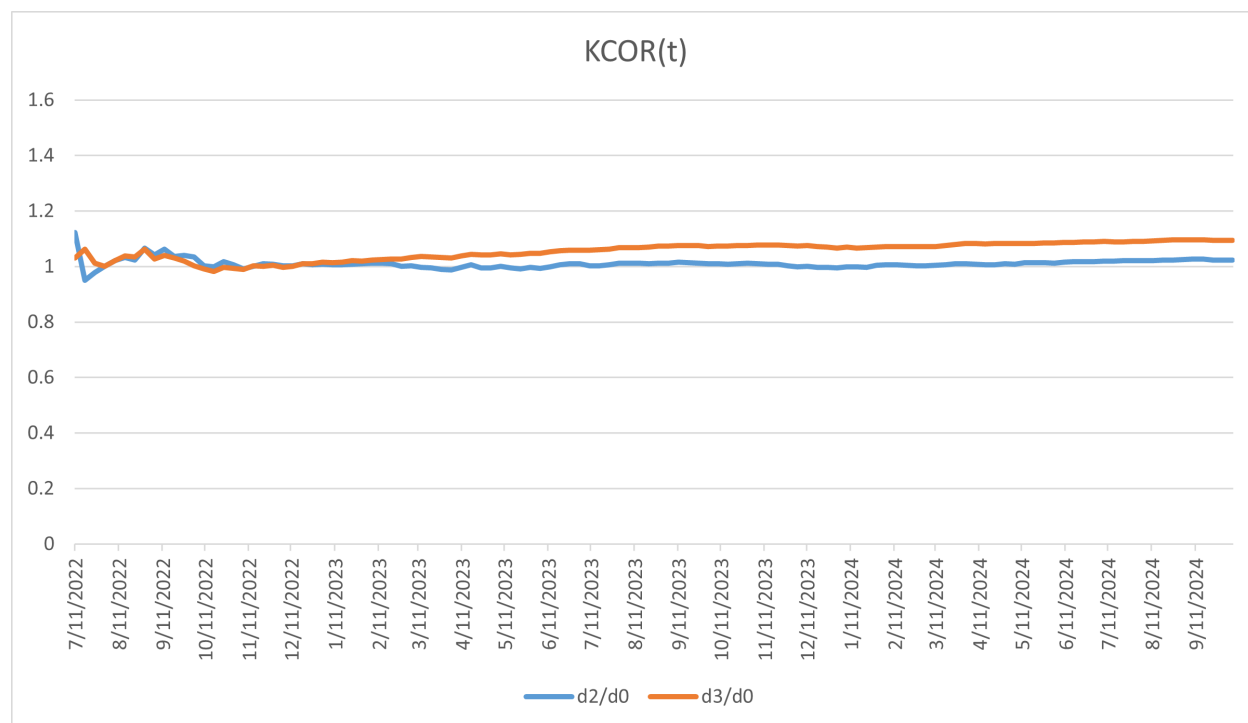
C.5 Robustness to age stratification

Figure 14: Birth-year cohort 1930: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as $\text{KCOR}(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference.

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

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

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

C.1.1 Illustrative empirical context: COVID-19 mortality data

The COVID-19 vaccination period provides a natural empirical regime characterized by strong selection heterogeneity and non-proportional hazards, making it a useful illustration for the KCOR framework. During this period, vaccine uptake was voluntary, rapidly time-varying, and correlated with baseline health status, creating clear examples of selection-induced non-proportional hazards. The Czech Republic national mortality registry data exemplify this regime: voluntary uptake led to asymmetric selection at enrollment, with vaccinated cohorts exhibiting minimal frailty heterogeneity while unvaccinated cohorts retained substantial heterogeneity. This asymmetric pattern reflects the healthy vaccinee effect operating through selective uptake rather than treatment. KCOR normalization removes this selection-induced curvature, enabling interpretable cumulative comparisons. While these examples illustrate KCOR's application, the method is general and applies to any retrospective cohort comparison where selection induces differential depletion dynamics.
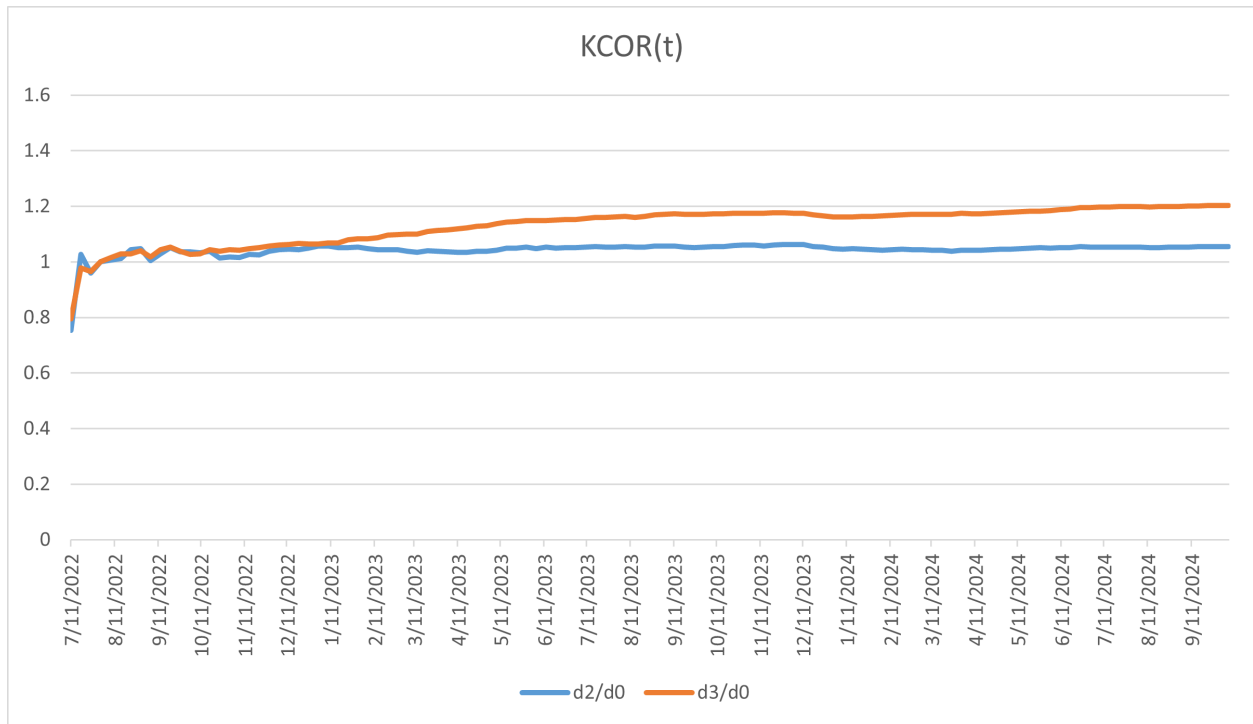
Figure 15: Birth-year cohort 1940: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as $\text{KCOR}(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference.
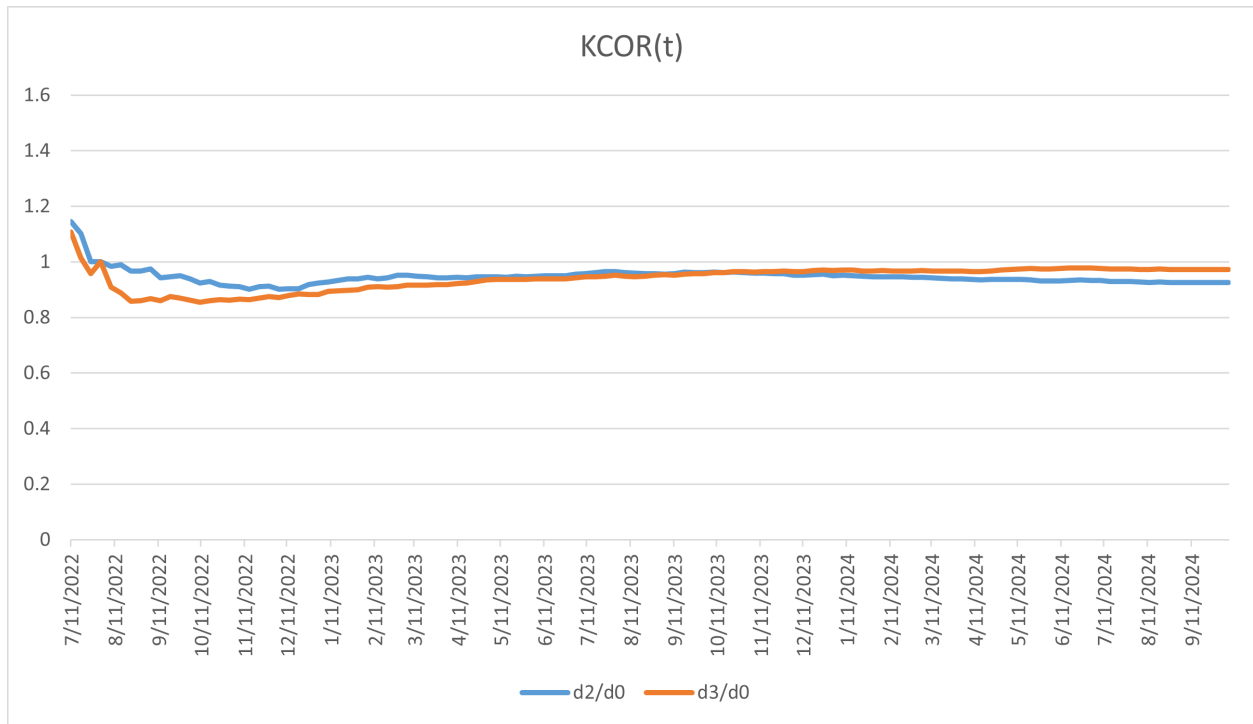
Figure 16: Birth-year cohort 1950: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as $\text{KCOR}(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference.

C.1.2 Frailty normalization behavior under empirical validation

Across examined age strata in the Czech Republic mortality dataset, fitted frailty parameters exhibit a pronounced asymmetry across cohorts. Some cohorts show negligible estimated frailty variance:

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

while others exhibit substantial frailty-driven depletion. This pattern reflects differences in selection-induced hazard curvature at cohort entry rather than any prespecified cohort identity.

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

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

Table 12: 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 14 reports raw outcome contrasts for ages 40+ (YOB ≤ 1980) where event counts are stable.

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

Table 13: Diagnostic gate for Czech application: KCOR results reported
only where diagnostics pass.

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

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

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

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

for Dose 0 cohorts and

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

for Dose 2 cohorts. Estimated frailty heterogeneity can appear larger at younger ages because baseline hazards are low, so proportional differences across latent risk strata translate into visibly different short-term hazards before depletion compresses the risk distribution. At older ages, higher baseline hazard and stronger ongoing depletion can reduce the apparent dispersion of remaining risk, yielding smaller fitted θ even if latent heterogeneity is not literally smaller. Frailty variance is largest at younger ages, where low baseline mortality amplifies the impact of heterogeneity on cumulative hazard curvature, and declines at older ages where mortality is compressed and survivors are more homogeneous. Because Table 12 demonstrates selection-induced heterogeneity, unadjusted cumulative outcome contrasts are expected to conflate depletion effects with any true treatment differences; Table 14 therefore reports raw cumulative hazards solely as a pre-normalization diagnostic. KCOR normalization removes the depletion component, enabling interpretable comparison of the remaining differences.

Table 14: Ratio of observed cumulative mortality hazards for
unvaccinated (Dose 0) versus fully vaccinated (Dose 2) Czech cohorts
enrolled in 2021_24.

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

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

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

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

C.1.3 Illustrative application to national registry mortality data

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

Figure 17 shows KCOR(t) trajectories for dose 2 and dose 3 relative to dose 0 for an all-ages analysis. We deliberately present an all-ages analysis as a high-heterogeneity stress test, since aggregation across age induces substantial baseline hazard and frailty variation. To assess whether apparent stability could arise from cancellation across strata, we also present narrow birth-year cohorts spanning advanced ages (1930, 1940, 1950) in Figures 14–16. Across aggregation levels, KCOR(t) remains stable over the evaluation window after depletion normalization, consistent with effective removal of selection-induced curvature in a real-data setting. These figures are presented as illustrative applications demonstrating estimator behavior on registry data and do not support causal inference; no hypothesis testing is performed.
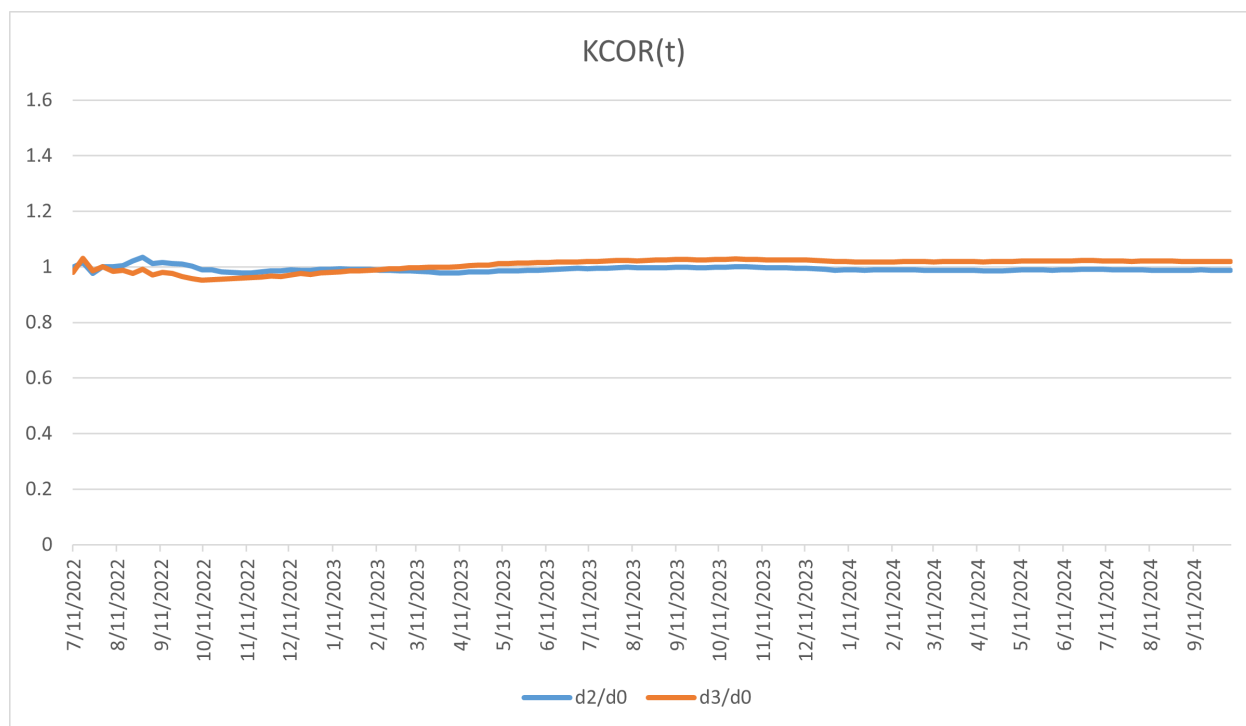

Figure 17: All-ages stress test: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR($t; t_0$)). This figure is presented as an illustrative application demonstrating estimator behavior under extreme heterogeneity and does not support causal inference.

Appendix D — Diagnostics and Failure Modes for KCOR Assumptions

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

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

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

Diagnostic signals of violation.

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

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

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

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

Diagnostic signals of violation.

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

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

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

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

Diagnostic signals of violation.

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

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

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

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

Diagnostic signals of violation.

- Poor fit of the gamma-frailty cumulative-hazard model during the quiet window.
- Systematic residual curvature after frailty normalization.

- Strong sensitivity of results to minor model or window perturbations.

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

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

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

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

Diagnostic signals of violation.

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

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

Under a valid quiet window, depletion-neutralized baseline cumulative hazards are expected to be approximately linear and $KCOR(t)$ trajectories to stabilize rather than drift. Persistent $KCOR(t)$ slope or $\hat{\theta}_d$ instability indicates contamination of the quiet window by external shocks or time-varying effects and signals loss of identifiability rather than evidence of cohort differences.

D.6 Diagnostic coherence across assumptions

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

D.7 Identifiability under sparse data

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

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

D.8 Summary: Diagnostic enforcement rather than assumption inflation

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