KCOR Supplementary Information (SI)

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

S1. Overview

This SI is organized as follows:

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

S2. Extended diagnostics and failure modes

This section describes the **observable diagnostics and failure modes** associated with the KCOR working assumptions and the corresponding diagnostics and identifiability criteria. No additional assumptions are introduced here. KCOR is designed to **fail transparently rather than silently**: when an assumption is violated, the resulting lack of identifiability or model stress manifests through explicit diagnostic signals rather than spurious estimates.

The KCOR framework separates **working assumptions**, **empirical diagnostics**, and **identifiability criteria**; these are summarized below in Tables @tbl:si_assumptions–@tbl:si_identifiability.

Table S1: KCOR working assumptions.

Assumption	Description	Role in KCOR
A1 Cohort stability	Cohorts are fixed at enrollment with no post-enrollment switching or informative censoring.	Ensures cumulative hazards are comparable over follow-up
A2 Shared external hazard environment	Cohorts experience the same background hazard over the comparison window.	Prevents confounding by cohort-specific shocks
A3 Time-invariant latent frailty	Selection operates through time-invariant unobserved heterogeneity inducing depletion.	Enables geometric normalization of curvature

Assumption	Description	Role in KCOR
A4 Adequacy of gamma frailty	Gamma frailty provides a reasonable approximation to observed depletion geometry.	Allows tractable inversion and normalization
A5 Quiet-window validity	A prespecified window exists in which depletion dominates other curvature sources.	Permits identification of frailty parameters

Table S2: Empirical diagnostics associated with KCOR assumptions.

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

Table S3: Identifiability criteria governing KCOR interpretation.

Criterion	Condition	Consequence if violated
I1 Diagnostic sufficiency	All required diagnostics pass.	KCOR interpretable
I2 Window alignment	Follow-up overlaps the hypothesized effect window.	Out-of-window effects not recoverable

Criterion	Condition	Consequence if violated
I3 Stability under perturbation	Estimates robust to tuning of windows and skips.	Interpretation limited
I4 Anchoring validity	Quiet window exhibits post-normalization linearity.	Anchoring invalid
I5 Conservative failure rule	Any failure → not identified.	Estimator remains valid, but results not reported

Failure of any interpretability or identifiability check limits the scope of inference but does not invalidate the KCOR estimator itself.

S3. Positive controls (injected harm/benefit)

S3.1 Positive controls: detect injected harm/benefit

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

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

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

{#eq:pos-control-injection}

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

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

S4. Control-test specifications and simulation parameters

S4.1 Reference implementation and default operational settings

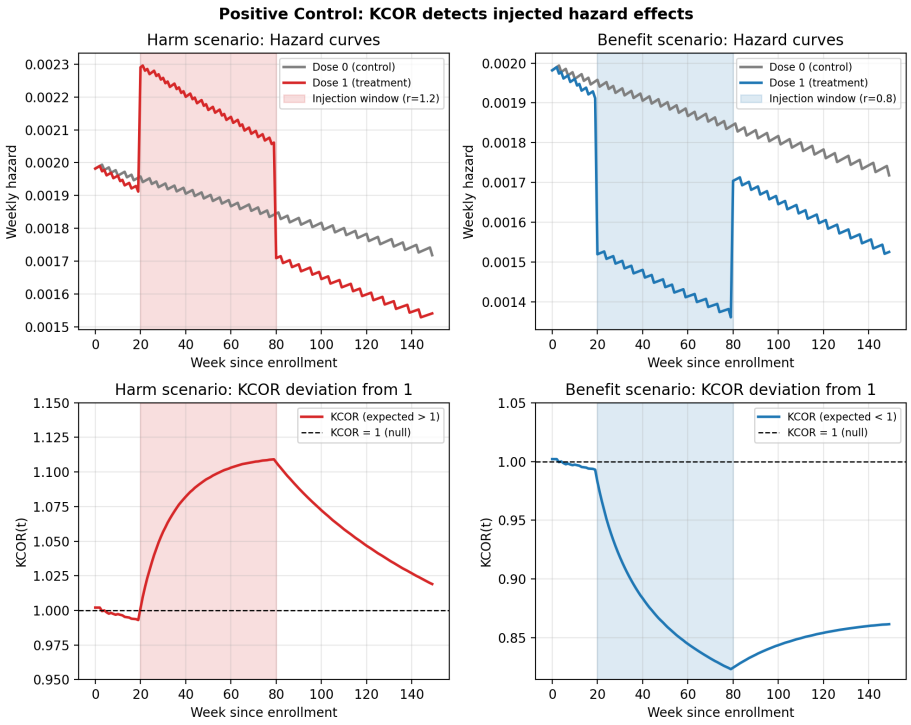

Figure S1: Positive control validation: KCOR correctly detects injected effects. Left panels show harm scenario (r=1.2), right panels show benefit scenario (r=0.8). Top row displays cohort hazard curves with effect window shaded. Bottom row shows KCOR(t) deviating from 1.0 in the expected direction during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

Table S4: Reference implementation and default operational settings.

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

S4.2 Negative control: synthetic gamma-frailty null

The synthetic negative control (Figure @fig:neg_control_synthetic) 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.

S4.3 Negative control: empirical age-shift construction

The empirical negative control (Figures @fig:neg_control_10yr and @fig:neg_control_20yr) 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.

S4.4 Positive control: injected effect

The positive control (Figure @fig:pos_control_injected and Table @tbl:pos_control_summary) 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.

S4.5 Sensitivity analysis parameters

The sensitivity analysis (Figure @fig:sensitivity_overview) varies:

- **Baseline weeks**: [2, 3, 4, 5, 6, 7, 8]
- **Quiet-start offsets**: [-12, -8, -4, 0, +4, +8, +12] weeks from 2023-01
- **Quiet-window end**: Fixed at 2023-52

- **Dose pairs**: 1 vs 0, 2 vs 0, 2 vs 1
- **Cohorts**: 2021_24

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

S4.6 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

S4.7 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

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

{#eq:si_individual_hazard_with_effect}

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

S5. Additional figures and diagnostics

S5.1 Fit diagnostics

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

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

Example diagnostic output from the reference implementation:

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

S5.2 Residual analysis

Fit residuals should be examined for. Define residuals:

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

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

S5.3 Parameter stability checks

Robustness of fitted parameters is assessed by:

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

S5.4 Quiet-window overlay plots

Quiet-window overlay plots: 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.

S5.5 Robustness to age stratification

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

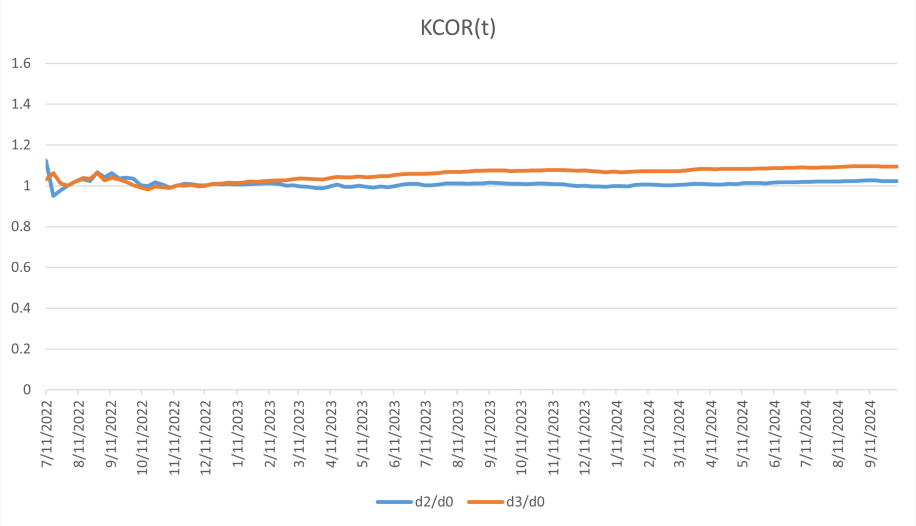

Figure S2: Birth-year cohort 1930: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR($t; t_0$)). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

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

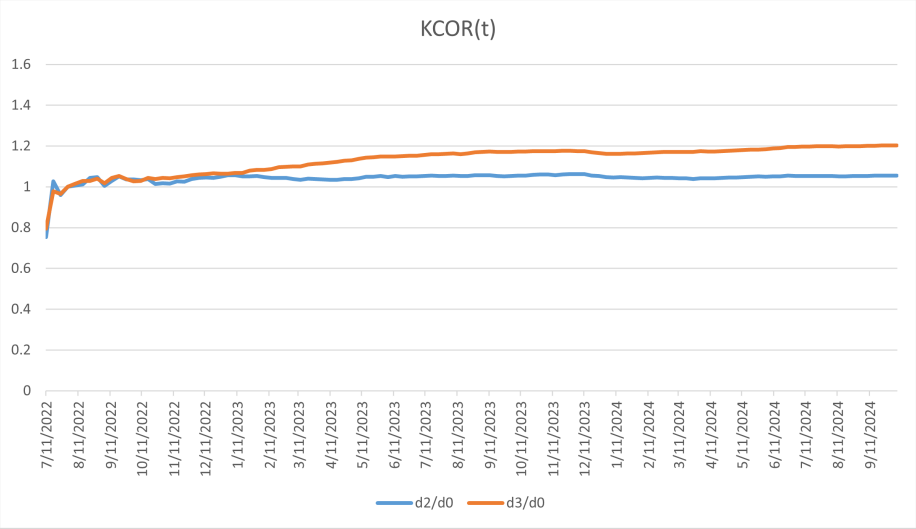

Figure S3: Birth-year cohort 1940: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR$(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

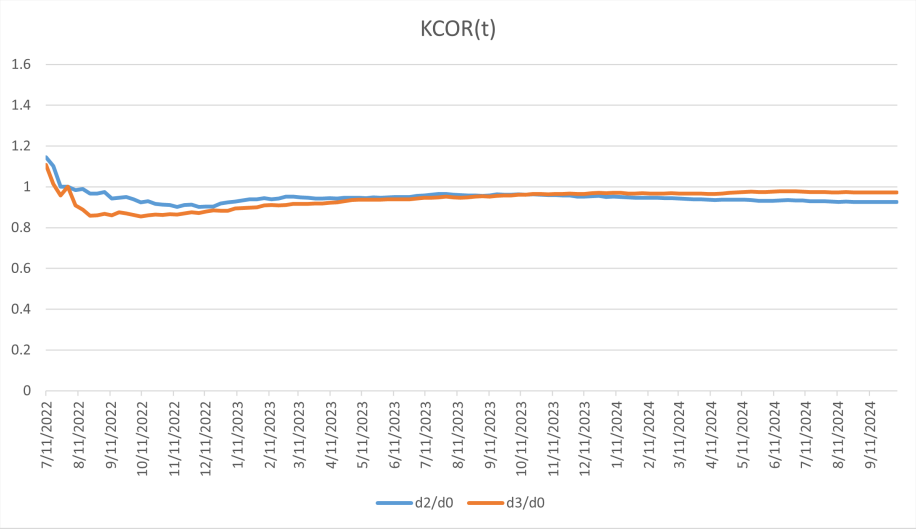

Figure S4: Birth-year cohort 1950: KCOR(t) trajectories comparing dose 2 and dose 3 to dose 0 for cohorts enrolled in ISO week 2022-26 and evaluated over calendar year 2023. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR$(t; t_0)$). This figure is presented as an illustrative application demonstrating estimator behavior on registry data and does not support causal inference. X-axis units are weeks since enrollment.

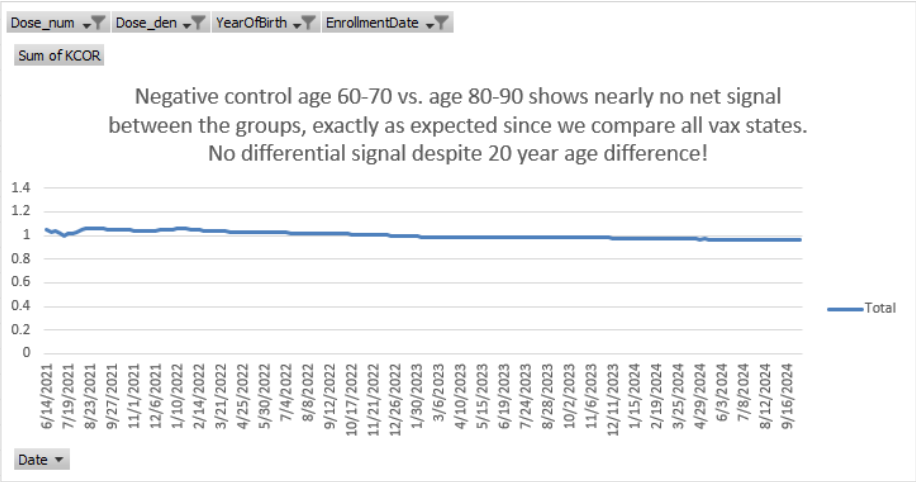

Figure S5: Empirical negative control with approximately 20-year age difference between cohorts. Even under extreme composition differences, KCOR(t) exhibits no systematic drift, consistent with robustness to selection-induced curvature. KCOR curves are anchored at $t_0 = 4$ weeks (i.e., plotted as KCOR($t; t_0$)). Uncertainty bands (95% bootstrap intervals) are shown. Data source: Czech Republic mortality and vaccination dataset processed into KCOR_CMR aggregated format (negative-control construction; see Supplementary Information, SI). X-axis units are weeks since enrollment.

S6. Extended Czech empirical application

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

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

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

S6.1.1 Illustrative empirical context: COVID-19 mortality data The COVID-19 vaccination period provides a natural empirical regime characterized by strong selection heterogeneity and non-proportional hazards, making it a useful illustration for the KCOR framework. During this period, vaccine uptake was voluntary, rapidly time-varying, and correlated with baseline health status, creating clear examples of selection-induced non-proportional hazards. The Czech Republic national mortality registry data exemplify this regime: voluntary uptake led to asymmetric selection at enrollment, with vaccinated cohorts exhibiting minimal frailty heterogeneity while unvaccinated cohorts retained substantial heterogeneity. This asymmetric pattern reflects the healthy vaccinee effect operating through selective uptake rather than treatment. KCOR normalization removes this selection-induced curvature, enabling interpretable cumulative comparisons. While these examples illustrate KCOR's application, the method is general and applies to any retrospective cohort comparison where selection induces differential depletion dynamics.

S6.1.2 Frailty normalization behavior under empirical validation Across examined age strata in the Czech Republic mortality dataset, fitted frailty parameters exhibit a pronounced asymmetry across cohorts. Some cohorts show negligible estimated frailty variance:

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

{#eq:si_theta_near_zero}

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 @tbl:si_diagnostic_gate summarizes these diagnostic checks across age strata.

Table S5: 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: In this application, unvaccinated cohorts exhibit frailty heterogeneity, while Dose 2 cohorts show near-zero estimated frailty across all age bands, consistent with selective uptake prior to follow-up:

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

{#eq:si_theta_positive}

for Dose 0 cohorts and

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

{#eq:si_theta_near_zero_dose2}

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 @tbl:si_frailty_variance demonstrates selection-induced heterogeneity, unadjusted cumulative outcome contrasts are expected to conflate depletion effects with any true treatment differences; see Table @tbl:si_raw_hazards for raw cumulative hazards reported as a pre-normalization diagnostic. KCOR normalization removes the depletion component, enabling interpretable comparison of the remaining differences.

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

Table S6: 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 @tbl:si_raw_hazards reports raw outcome contrasts for ages 40+ (YOB $\leq$ 1980) where event counts are stable.

Diagnostic checks: - **Dose ordering:** the fitted frailty variance is positive for Dose 0 and collapses toward zero for Dose 2 across all age strata, consistent with selective uptake. - **Magnitude separation:** Dose 2 estimates are effectively zero relative to Dose 0, indicating near-linear cumulative hazards rather than

forced curvature. - **Age coherence:** the fitted frailty variance decreases at older ages as baseline mortality rises and survivor populations become more homogeneous; monotonicity is not imposed. - **Stability:** No sign reversals, boundary pathologies, or numerical instabilities are observed. - **Falsifiability:** Failure of any one of these checks would constitute evidence against model adequacy.

Table S7: Ratio of observed cumulative mortality hazards for unvaccinated (Dose 0) versus fully vaccinated (Dose 2) Czech cohorts enrolled in 2021_24. (Note: the all-ages row reflects aggregation effects and is not directly comparable to age-stratified rows.)

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

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

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

S6.1.3 Illustrative application to national registry mortality data
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 @fig:si_allages 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.

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

S7. Computational environment and runtime notes

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

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

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