Kaplan–Meier estimator under ranked set sampling with censoring and imperfect ranking

Abstract

Ranked set sampling (RSS) is a sampling design for studies in which it is inexpensive to rank experimental units using an auxiliary measurement, but costly or slow to obtain full measurements. This work extends RSS methodology to right-censored survival data, a setting common in medical and reliability studies. We develop Kaplan–Meier and Nelson–Aalen estimators for RSS under both perfect and imperfect ranking, establish their large-sample behavior, and show that they can provide more precise survival estimates than those based on simple random sampling. We also propose variance estimators, including a modification designed to improve stability when risk sets become sparse at long follow-up times. Simulation studies using Weibull and log-normal survival models examine the impact of set size, censoring, and ranking quality, and confirm substantial efficiency gains when the ranking variable is reasonably informative. The work also outlines future extensions, including RSS-based survival tests and finite-horizon functionals such as restricted or window mean lifetime. Overall, the results demonstrate that RSS can make survival studies more efficient and cost-effective without changing their scientific targets.

Keywords: ranked set sampling; Kaplan–Meier estimator; Nelson–Aalen estimator; right censoring; imperfect ranking; survival analysis.

1 Introduction

Ranked set sampling (RSS) is a cost-efficient design that uses inexpensive baseline ranking to select a more informative subset of individuals for full measurement (McIntyre, 1952; Dell and Clutter, 1972; Chen, Bai, and Sinha, 2004). In many modern studies it is easy to obtain a crude ordering of units at baseline (for example by biomarkers, imaging scores, clinical risk scores, device diagnostics, or engagement metrics), but costly or slow to observe the primary outcome. RSS exploits this imbalance: small sets of candidates are formed, ranked using the cheap information, and only one unit per set is fully measured. For uncensored outcomes, this is known to deliver estimators with substantially smaller variance than simple random sampling (SRS) of the same size, even when rankings are only moderately accurate.

Survival studies often have the same structure. Baseline variables such as disease stage, comorbidities, early biomarker panels, or engineering diagnostics can order subjects by expected time to event, whereas observing the actual event or censoring time requires long and expensive follow-up. In principle, an RSS design could let investigators follow fewer patients while retaining the precision of a larger SRS, or achieve much higher precision at fixed sample size. In practice, however, applied survival analysis remains almost entirely SRS based. Existing RSS work for censored data has focused mainly on point estimation of the Kaplan–Meier (KM) curve under random censoring (Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2014; Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2017), with relatively little attention to large-sample theory under imperfect ranking or to the variance estimators and design diagnostics that practitioners need in order to calibrate the gains from RSS.

The overarching goal of this work is to develop a rank-aware survival framework that shows when RSS can deliver real efficiency gains in time-to-event studies while retaining the same type of inferential output that practitioners are used to under SRS. We recast the KM and Nelson–Aalen estimators under independent random censoring in the empirical-process and martingale framework of Stute and Wang (1993), Stute (1995), and Andersen et al. (1993), and extend it to balanced RSS with both perfect and judgment ranking. The rank index is carried through the influence-function representation, yielding law of large numbers and process-level central limit theorems for KM and NA on $[0, \tau^* - \varepsilon]$ under SRS, perfectly ranked RSS, and judgment-ranked RSS in a common language. On top of this linearization, we derive rank-aware Greenwood-type plug-in variance estimators, show that they converge to the limiting covariance kernels in the RSS CLTs, and use them, together with Monte Carlo experiments, to quantify how relative efficiency depends on set size k, number of cycles m, censoring, and ranking quality (for example Dell–Clutter-style misranking). The resulting ordering has the variance under perfect RSS no larger than under judgment ranking, and both no larger than under SRS, with equality only at the extremes of perfect or noninformative ranking. Although this paper focuses on estimation and variance, the rank-indexed counting-process representation is directly reusable for two-sample log-rank and weighted log-rank tests and for time-integrated functionals such as restricted mean life and window mean life, so the results developed here provide the technical foundation for those extensions, which are outlined briefly in the discussion.

Collectively, these contributions support a framework in which survival can be estimated, compared, and interpreted in much the same way as under SRS, but with the potential for sizable efficiency gains whenever informative ranking is available at baseline.

2 Background and motivation

Ranked set sampling. RSS was introduced by McIntyre (1952) in agricultural yield studies, based on a simple observation: ranking small sets of units is often cheap, but fully measuring all of them is expensive. In a balanced RSS design, we repeatedly draw sets of size k, rank the k units using judgment or a cheap surrogate, and in the rth draw of the cycle we only measure the unit judged to have rank r, $r = 1, \dots, k$. Repeating this for m cycles yields

$n = mk$ fully observed units,

$$X_{[r]j}, \quad r = 1, \ldots, k, \quad j = 1, \ldots, m,$$

where $X_{[r]j}$ denotes the unit judged to be the rth order statistic in cycle j. Under perfect ranking, this order coincides with the true order in each set; under imperfect ranking (the practically relevant case), ranking is performed on a noisy concomitant $\tilde{X} = X + Z$ so the judged order only correlates with the truth.

The key fact, made precise in Dell and Clutter (1972), is that RSS forces the measured sample to be spread across the entire support of the underlying distribution. As a result, the RSS sample mean is unbiased and typically has substantially smaller variance than the SRS mean of the same size, sometimes by 30–50% or more depending on the distribution and k. Later work, summarized in Chen, Bai, and Sinha (2004), showed that this efficiency gain for the sample mean is robust: even when rankings are not perfect, RSS still outperforms SRS.

Why RSS is attractive for survival. In time-to-event studies, it is common to have a relatively cheap way to order subjects by their likely risk (baseline biomarkers, imaging, clinical scores, wearables) and a more expensive or longer process to obtain the actual event or censoring time. This creates exactly the same asymmetry that motivated RSS in agriculture: ranking is cheap, full follow-up is costly. An RSS design lets investigators (i) form small sets of eligible subjects, (ii) rank them by expected time-to-event using the auxiliary information, and (iii) fully follow only the selected unit from each set. This "rank-then-follow" workflow is realistic in oncology (imaging or marker-based ranking, then follow-up for recurrence), cardiovascular cohorts (lipids, blood pressure, ECG features to follow-up for myocardial infarction or stroke), infectious-disease surveillance (symptom profile to follow-up for time-to-confirmed diagnosis), business analytics (engagement to time-to-churn), and reliability studies (early sensor diagnostics or performance metrics to time-to-failure for components or devices), all domains in which events can be rare, delayed, or administratively censored.

Compared with classical stratified sampling, RSS has an important advantage in these settings. Stratification over a continuous auxiliary variable requires choosing cut-points; poor cut-points waste information. RSS instead uses local ranking within sets, preserving the full ordinal information of the auxiliary measure and avoiding a global partition. Under perfect ranking, this can deliver variance reductions approaching $1/k$. Under imperfect ranking, the gain decays smoothly with the surrogate's correlation, as documented in Mahdi Mahdizadeh and Strzałkowska-Kominiak (2014) and Mahdi Mahdizadeh and Strzałkowska-Kominiak (2017) for censored data. Thus, if a study can afford to form sets before committing to full follow-up, RSS is often more informative than SRS for the same number of subjects.

RSS with censored survival. For right-censored data, the survival target is

$$S(t) = P(X > t),$$

but we only observe $Y = \min(X, C)$ and $\delta = I(X \leq C)$. Under SRS and independent censoring, the Kaplan–Meier estimator

$$\widehat{S}(t) = \prod_{u \leq t} \left(1 - \frac{dN(u)}{R(u)}\right)$$

is standard, and its properties are well understood: almost sure uniform convergence (Stute and Wang, 1993), central limit theorems for a wide class of functionals (Stute, 1995), and martingale representations (Fleming and Harrington, 1991; Andersen et al., 1993).

Mahdi Mahdizadeh and Strzałkowska-Kominiak (2014) were the first to carry this over to RSS with random censoring. In their setup, we still have a balanced RSS design with k ranks and m cycles, so the observed data are

$$(Y_{[r]j}, \delta_{[r]j}), \quad r = 1, \ldots, k, \quad j = 1, \ldots, m,$$

and within each rank r these pairs are i.i.d. but come from a rank-specific distribution (because higher-ranked units tend to have larger underlying X). Their RSS Kaplan–Meier estimator pools the rank-wise risk sets:

$$\widehat{S}_{\mathrm{RSS}}(t) = \prod_{u \le t} \left(1 - \frac{\sum_{r=1}^{k} dN_{[r]}(u)}{\sum_{r=1}^{k} R_{[r]}(u)} \right),$$

which is exactly the SRS product-limit form but with rank-indexed counting and at-risk processes. They showed that this estimator is unbiased at event times, uniformly consistent, and asymptotically normal, and that in simulations it can have noticeably smaller variance than the SRS KM for the same $n = mk$. Mahdi Mahdizadeh and Strzałkowska-Kominiak (2017) then relaxed the perfect-ranking assumption by letting judgment be based on a noisy concomitant; using bootstrap-type procedures they showed that efficiency degrades smoothly as ranking worsens, rather than collapsing immediately.

There are also side branches of this literature: parametric RSS survival with Type I censoring (Yu and Tam, 2002), PROS-based KM for incomplete rankings (Nematolahi et al., 2020), and several mean residual life (MRL) papers under generally fully observed RSS lifetimes (Elham Zamanzade, Parvardeh, and Asadi, 2019; Ehsan Zamanzade, M. Mahdizadeh, and Samawi, 2024; Elham Zamanzade, Ehsan Zamanzade, and Parvardeh, 2024). These all reinforce the same message: whenever it is possible to rank cheaply, RSS can buy efficiency.

What is still missing. Despite these contributions, RSS survival methods remain limited in several ways. Work with censoring has largely focused on point estimation of $S(t)$ under perfect ranking (Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2014) or on simulation and resampling under heuristic concomitant models (Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2017), with only partial large-sample theory. The KM and NA estimators have not been treated in a unified, process-level framework that accommodates both perfect and imperfect (judgment) ranking, discrete event times, and the martingale/empirical-process tools routinely used under SRS. In particular, there is little guidance on rank-aware variance estimation, on how efficiency depends on ranking quality, or on when RSS genuinely improves precision over SRS and when it effectively collapses back to SRS. These gaps make it difficult to recommend RSS as a routine design option for survival studies, even when informative baseline ranking is available.

3 Survival estimation under random censoring and RSS

This section develops the asymptotic framework for the rest of the paper. We start from the classical independent random-censorship model and the product-integral form of the Kaplan–Meier and Nelson–Aalen estimators, then extend it to balanced RSS with perfect and judgment ranking. Once each rank (true or judged) is treated as its own i.i.d. stratum with independent censoring, the usual SRS arguments—uniform laws, martingale representations, functional CLTs, and Greenwood plug-ins—apply rank-wise and can then be averaged (see Supplements A.1–A.2).

3.1 Baseline random-censorship model

Let X be the nonnegative lifetime and C the nonnegative censoring time. We observe i.i.d.

$$\{(Y_i, \delta_i) : i = 1, \dots, n\}, \qquad Y_i = \min(X_i, C_i), \quad \delta_i = I(X_i \le C_i),$$

under the assumptions (i) (X_i, C_i) i.i.d., (ii) $X \perp C$, and (iii) $F(t) = P(X \le t)$, $G(t) = P(C \le t)$ are defined on $[0, \tau)$ with strictly positive observed-time survival

$$S_Y(t) := P(Y \ge t) = S(t)K(t) > 0 \quad \text{on } [0, \tau^* - \varepsilon],$$

where $S = 1 - F$ and $K = 1 - G$, and τ^* is the right endpoint of Y (see Supplement A.1 for the distinction $\tau^* \le \tau$ and the interval of uniformity). The sub-distributions

$$H_1(t) = P(Y \le t, \delta = 1), \qquad H_0(t) = P(Y \le t, \delta = 0), \qquad H = H_0 + H_1$$

satisfy the standard identities $H_1(t) = \int_0^t K(u-)\,dF(u)$ and

$$d\Lambda_1(u) = \frac{dH_1(u)}{S_Y(u-)} = \frac{dF(u)}{S(u-)}.$$

These drive both the KM and NA estimators.

3.2 Counting-process setup

For each subject define the counting and at-risk processes

$$N_i(t) = I(Y_i \le t, \delta_i = 1), \qquad Y_i(t) = I(Y_i \ge t),$$

and their sums $N(t) = \sum_i N_i(t)$, $R(t) = \sum_i Y_i(t)$. With respect to the filtration generated by $\{Y_i(\cdot), N_i(\cdot)\}$, we have the Doob–Meyer decomposition

$$N_i(t) = \int_0^t Y_i(u)\,d\Lambda_1(u) + M_i(t),$$

with M_i a square-integrable martingale. Summing gives

$$N(t) = \int_0^t R(u)\,d\Lambda_1(u) + M(t),$$

so the Nelson–Aalen estimator

$$\widehat{\Lambda}_1(t) = \int_0^t \frac{dN(u)}{R(u)}$$

admits the martingale transform representation

$$\widehat{\Lambda}_1(t) - \Lambda_1(t) = \int_0^t \frac{1}{R(u)}\,dM(u),$$

and on $[0, \tau^* - \varepsilon]$ we have $R(u) \ge c_\varepsilon n$ almost surely for all large n (by a Glivenko–Cantelli argument; see Supplement A.1), so $\widehat{\Lambda}_1 - \Lambda_1 = O_p(n^{-1/2})$.

The Kaplan–Meier estimator,

$$\widehat{S}(t) = \prod_{u \le t}\Big(1 - \frac{dN(u)}{R(u)}\Big) = \prod_{u \le t}(1 - d\widehat{\Lambda}_1(u)),$$

is linked to NA through the product-integral map. A Taylor expansion of $\log(1-x)$ with $x = dN/R$ yields

$$\log \widehat{S}(t) = -\widehat{\Lambda}_1(t) + r_n(t), \qquad \sup_{t \le \tau^* - \varepsilon} |r_n(t)| = o_p(n^{-1/2}),$$

so, at the CLT scale, KM and NA carry the same randomness (see Supplement A.1 for the full product-integral linearization).

3.3 Empirical sub-distributions and strong law (SRS)

Because the classes $\{I(Y \le t, \delta = d)\}$ are VC, the empirical sub-distributions

$$H_{1n}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le t, \delta_i = 1), \quad H_{0n}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le t, \delta_i = 0)$$

converge uniformly almost surely to (H_1, H_0) on $[0, \tau^* - \varepsilon]$. Thus $S_{Y,n} = 1 - H_n \to S_Y$ uniformly and $S_{Y,n}$ is bounded away from 0 eventually. Substituting into

$$\widehat{\Lambda}_1(t) = \int_0^t \frac{dH_{1n}(u)}{S_{Y,n}(u-)}, \quad \Lambda_1(t) = \int_0^t \frac{dH_1(u)}{S_Y(u-)},$$

and bounding the two add–subtract terms gives

$$\sup_{t \le \tau^* - \varepsilon} |\widehat{\Lambda}_1(t) - \Lambda_1(t)| \xrightarrow{\text{a.s.}} 0,$$

and, by continuity of the product-integral map,

$$\sup_{t \le \tau^* - \varepsilon} |\widehat{S}(t) - S(t)| \xrightarrow{\text{a.s.}} 0.$$

A detailed version of the add–subtract expansion is given in Supplement A.1.

3.4 Functional CLT and Greenwood variance (SRS)

Applying Rebolledo's martingale functional CLT to the stochastic integral $\int_0^t R(u)^{-1} dM(u)$ and using $R(u)/n \to S_Y(u-)$ yields, in $\ell^\infty([0, \tau^* - \varepsilon])$,

$$\sqrt{n}\,(\widehat{\Lambda}_1 - \Lambda_1) \Rightarrow \mathbb{G}_\Lambda, \qquad \mathrm{Cov}\big(\mathbb{G}_\Lambda(s), \mathbb{G}_\Lambda(t)\big) = \int_0^{s\wedge t} \frac{dH_1(u)}{S_Y(u-)^2}.$$

By the delta method for the product integral,

$$\sqrt{n}\,(\widehat{S} - S) \Rightarrow \mathbb{G}_S := -S\,\mathbb{G}_\Lambda, \quad \mathrm{Cov}\big(\mathbb{G}_S(s), \mathbb{G}_S(t)\big) = S(s)S(t) \int_0^{s\wedge t} \frac{dH_1(u)}{S_Y(u-)^2}.$$

This gives, at any fixed t,

$$\mathrm{AVAR}\big(\widehat{S}(t)\big) = \frac{S(t)^2}{n} \int_0^t \frac{dH_1(u)}{S_Y(u-)^2} + o(n^{-1}),$$

and the usual Greenwood plug-in

$$\widehat{\mathrm{AVAR}}\big(\widehat{S}(t)\big) = \widehat{S}(t)^2 \sum_{u \le t} \frac{dN(u)}{R(u)\big(R(u) - dN(u)\big)}$$

is consistent (the with-ties version); see Supplement A.1 for an explicit derivation.

3.5 Ranked set sampling with concomitants

In balanced RSS we observe $n = mk$ units arranged as k ranks over m cycles. Under perfect ranking, the lifetime among units measured at true rank r is the rth order-statistic law, written $F_{[r]}$, with survival $S_{[r]}$ and observed-time survival $S_{Y,[r]}(t) = S_{[r]}(t)K(t)$. We compute a KM/NA curve within each rank,

$$\widehat{\Lambda}_{1,[r]}(t) = \int_0^t \frac{dN_r(u)}{R_r(u)}, \qquad \widehat{S}_{[r]}(t) = \prod_{u \le t}\Big(1 - \frac{dN_r(u)}{R_r(u)}\Big),$$

and estimate the population survival by the equal-weight average

$$\widehat{S}_{\mathrm{RSS}}(t) := \frac{1}{k}\sum_{r=1}^{k} \widehat{S}_{[r]}(t).$$

McIntyre's RSS identity $\frac{1}{k}\sum_{r=1}^{k} S_{[r]}(t) = S(t)$ then gives $\widehat{S}_{\mathrm{RSS}} \to S$ uniformly on $[0, \tau^* - \varepsilon]$ as $m \to \infty$. The rank-wise i.i.d. structure and independence across ranks (because each cycle uses k independent candidate sets) are spelled out in Supplement A.2.

In practice, ranking is done on a proxy $\tilde{X} = X + Z$ (Dell–Clutter model), so we work with judged ranks $J = r$. As recalled in Supplement **??**, the judged-r lifetime cdf is a mixture

$$F_r^J(t) = \sum_{j=1}^{k} w_{rj} F_{[j]}(t), \qquad w_{rj} = P(T = j \mid J = r), \ \sum_j w_{rj} = 1,$$

and, under balanced selection $P(J = r) = 1/k$ and uniform true-rank frequencies $P(T = j) = 1/k$, the average over r recovers the population:

$$\frac{1}{k}\sum_{r=1}^{k} F_r^J(t) = F(t), \qquad \frac{1}{k}\sum_{r=1}^{k} S_r^J(t) = S(t).$$

Hence the judgment-rank KM,

$$\widehat{S}_{\mathrm{RSS},J}(t) := \frac{1}{k}\sum_{r=1}^{k}\widehat{S}_r^J(t),$$

is a consistent estimator of $S(t)$.

3.6 CLT and efficiency ordering under RSS

Because each rank contributes m i.i.d. observations, the SRS martingale functional CLT applies within rank:

$$\sqrt{m}\big(\widehat{S}_{[r]} - S_{[r]}\big) \Rightarrow \mathbb{G}_{S,[r]}, \qquad \mathrm{Cov}(\mathbb{G}_{S,[r]}(s), \mathbb{G}_{S,[r]}(t)) = S_{[r]}(s)S_{[r]}(t)\int_0^{s\wedge t}\frac{dH_{1,[r]}(u)}{S_{Y,[r]}(u-)^2}.$$

With $n = mk$ and independence across ranks,

$$\sqrt{n}\big(\widehat{S}_{\mathrm{RSS}} - S\big) = \frac{1}{\sqrt{k}}\sum_{r=1}^{k}\sqrt{m}\big(\widehat{S}_{[r]} - S_{[r]}\big) \Rightarrow \mathbb{G}_S^{(\mathrm{perf})}$$

with covariance kernel

$$\mathrm{Cov}\big(\mathbb{G}_S^{(\mathrm{perf})}(s), \mathbb{G}_S^{(\mathrm{perf})}(t)\big) = \frac{1}{k}\sum_{r=1}^{k} S_{[r]}(s)S_{[r]}(t)\int_0^{s\wedge t}\frac{dH_{1,[r]}(u)}{S_{Y,[r]}(u-)^2}.$$

The judgment-rank limit is identical except that $[r]$ is replaced by $J = r$ and the rank-specific laws are mixtures (see Supplement A.2 for the RSS add–subtract expansion mirroring the SRS one). Comparing the SRS kernel

$$S(s)S(t)\int_0^{s\wedge t}\frac{dH_1(u)}{S_Y(u-)^2}$$

with the two RSS kernels yields the pointwise ordering

$$V_{\mathrm{perf}}(t) \;\le\; V_{\mathrm{judg}}(t;\rho) \;\le\; V_{\mathrm{SRS}}(t),$$

with equalities at $\rho = 1$ (perfect ranking) and $\rho = 0$ (uninformative ranking). Thus, rank information never hurts first-order efficiency and helps whenever the proxy is even moderately informative.

3.7 Variance estimation for rank-aware KM

Because the RSS KM is an equal-weight average of k independent KM curves, the natural plug-ins are simple averages of within-rank Greenwoods:

$$\widehat{\mathrm{Var}}\{\widehat{\Lambda}_{\mathrm{RSS}}(t)\} = \frac{1}{k^2}\sum_{r=1}^{k}\sum_{u\le t}\frac{dN_r(u)}{R_r(u)^2}, \quad \widehat{\mathrm{Var}}\{\widehat{S}_{\mathrm{RSS}}(t)\} = \frac{1}{k^2}\sum_{r=1}^{k}\widehat{S}_{[r]}(t)^2\sum_{u\le t}\frac{dN_r(u)}{R_r(u)\{R_r(u)-dN_r(u)\}},$$

with the same formulas for judgment ranks using $\widehat{S}_r^J$ and $H_{1,r}^J$. These converge to the kernels in the RSS CLTs above (see Supplement A.2).

3.8 Simulation study

We assess the finite-sample behavior of these estimators under two superpopulations, chosen to match the theory as closely as possible while still producing appreciable rankwise heterogeneity.

(i) Log-AFT superpopulation. We generate lifetimes from

$$X = \exp(\mu - \beta Z + \varepsilon), \qquad Z \sim N(0,1),\ \varepsilon \sim N(0,\sigma_\varepsilon^2),$$

with (Z,ε) independent and baseline values $\mu = 0$, $\beta = 1.5$, $\sigma_\varepsilon = 0.4$. A large i.i.d. sample from this model is first generated to (a) fix four evaluation times at survival levels

$$S(t_{0.25}) \approx 0.75, \quad S(t_{0.50}) \approx 0.50, \quad S(t_{0.75}) \approx 0.25, \quad S(t_{0.90}) \approx 0.10,$$

and (b) calibrate ranking noise. To mimic practical RSS, we rank on a noisy concomitant

$$\tilde{X} = Z + U, \qquad U \sim N(0, \sigma^2_{\tilde{X}}),$$

and choose $\sigma_{\tilde{X}}$ so that $|\operatorname{Corr}(\tilde{X}, X)|$ is close to a prescribed $\rho_{\text{target}} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. This is the analogue of the Dell–Clutter $\rho = \operatorname{Corr}(\tilde{X}, X)$ used in the theory.

Balanced RSS samples are then generated exactly as in the model: for each cycle we create k independent candidate sets of size k, rank within the rth set by W and measure only the unit judged to be rank r, $r = 1, \ldots, k$. Repeating over m cycles yields $n = mk$ observations and m observations per judged rank, so the rank-aware KM is $\widehat{S}_{\text{RSS}}(t) = \frac{1}{k}\sum_{r=1}^{k} \widehat{S}^J_r(t)$.

Censoring is imposed independently by $C \sim \operatorname{Exp}(-\log(1 - p_{\text{cens}})/\mathbb{E}X)$ with $p_{\text{cens}} \in \{0, 0.1, 0.3, 0.5\}$, and the same censoring law is used for the SRS benchmark of size $n = mk$. For each replicate we compute: (i) the SRS KM, (ii) the rank-aware RSS KM, and (iii) the corresponding Greenwood plug-ins from the previous subsection.

The grid is

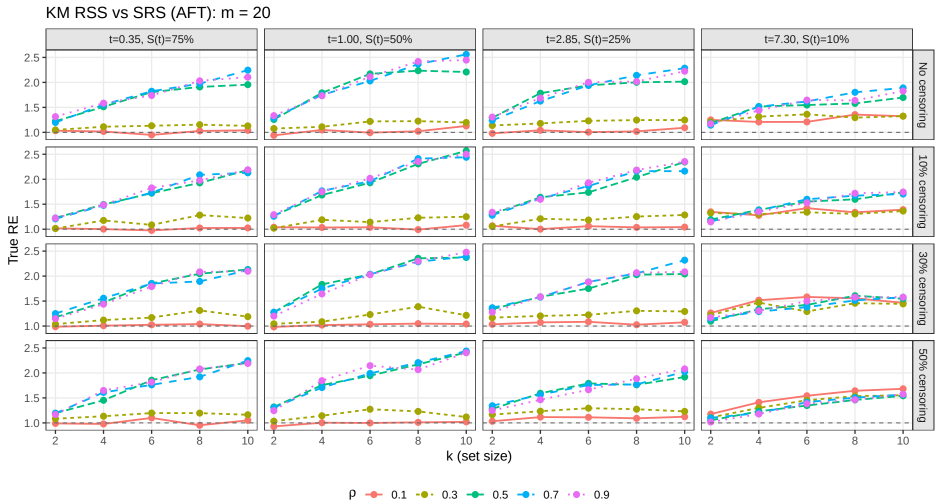
$$k \in \{2, 4, 6, 8, 10\}, \quad m \in \{20, 50\}, \quad \rho_{\text{target}} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}, \quad p_{\text{cens}} \in \{0, 0.1, 0.3, 0.5\}.$$

For every cell we run $B_{\text{MC}} = 10{,}000$ Monte Carlo replicates, which keeps the relative Monte Carlo error of variance estimates around 1.4% and makes variance ratios

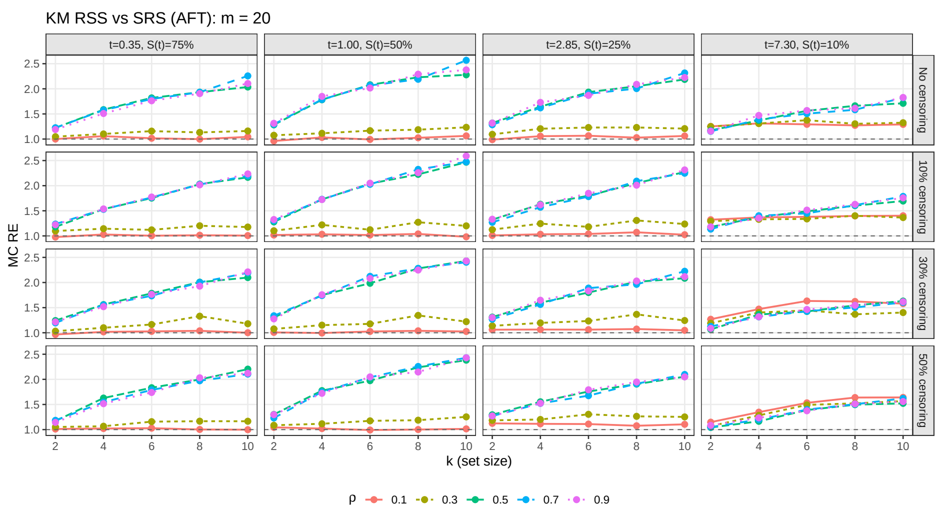
$$\text{RE}_{\text{MC}}(t) = \frac{\widehat{V}_{\text{SRS,MC}}(t)}{\widehat{V}_{\text{RSS,MC}}(t)}, \qquad \text{RE}_{\text{GW}}(t) = \frac{\overline{\widehat{\operatorname{Var}}(\widehat{S}_{\text{SRS}}(t))}}{\overline{\widehat{\operatorname{Var}}(\widehat{S}_{\text{RSS}}(t))}},$$

stable. A secondary run with $B_{\text{true}} = 4{,}000$ replicates is used only to smooth the theoretical ratio $\text{RE}_{\text{true}}(t) = V_{\text{SRS,true}}(t)/V_{\text{RSS,true}}(t)$ over the grid. Because the SRS and RSS samples in a replicate always share (n, p_{cens}), the efficiency comparison is fair.

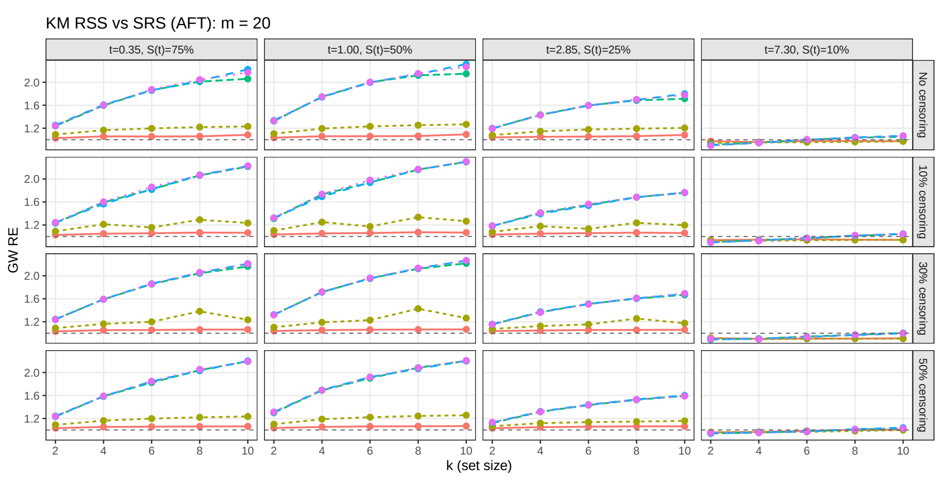
Here we present the simulation results graphically to show how the relative efficiency changes with evaluation time, k, m, ρ and censoring level.



(a) True RE, $m = 20$

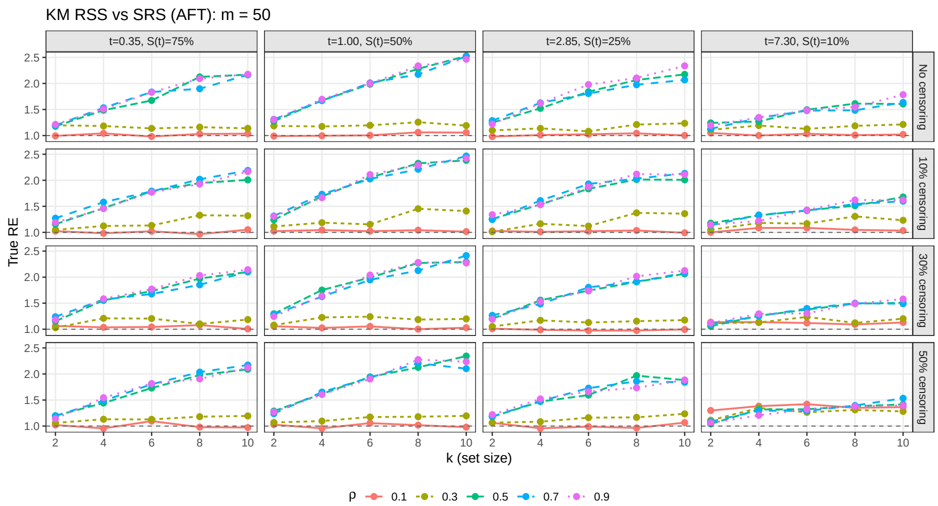


(b) MC RE, $m = 20$

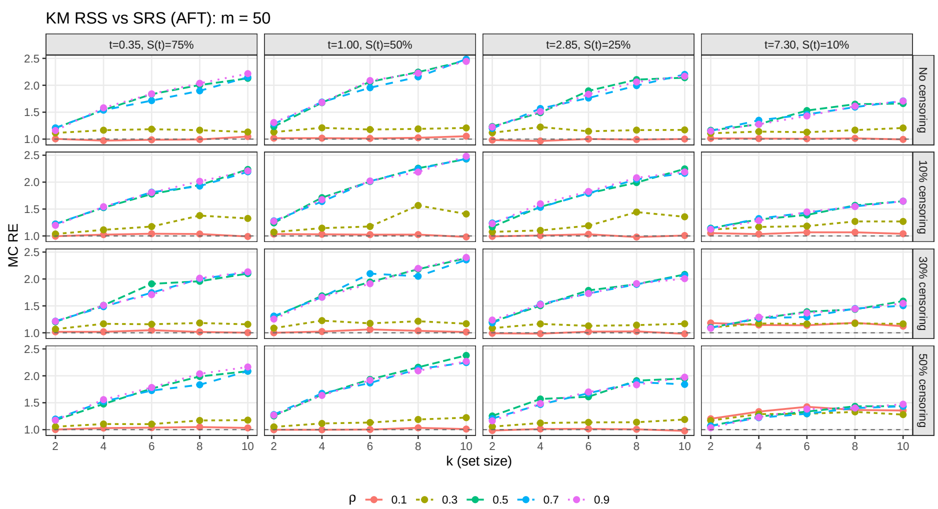


(c) GW RE, $m = 20$

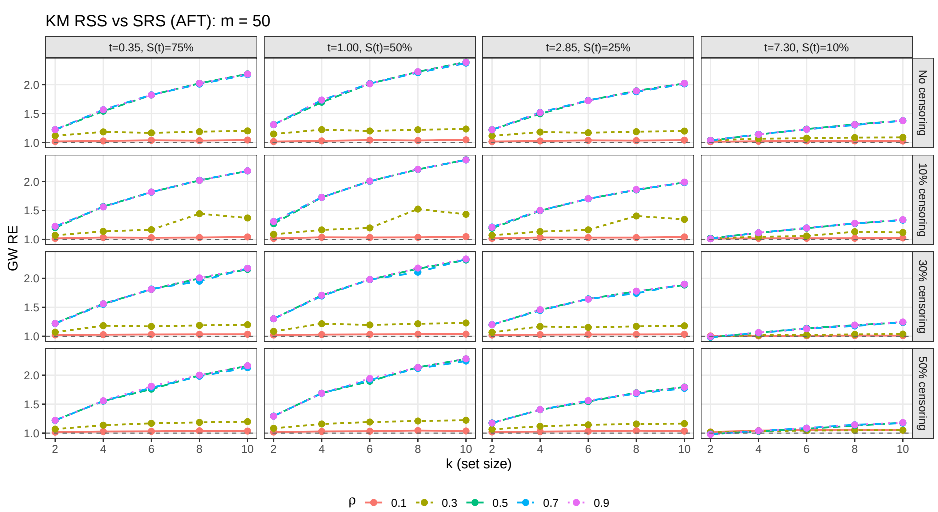
Figure 1: AFT superpopulation: relative efficiency (RSS vs. SRS) across set size k, ranking quality ρ (legend), censoring fractions (rows), and evaluation times (columns). The dashed line marks RE = 1.



(a) True RE, $m = 50$



(b) MC RE, $m = 50$



(c) GW RE, $m = 50$

Figure 2: AFT superpopulation: relative efficiency (RSS vs. SRS) with $m = 50$ cycles. Layout as in Fig. 1.

Main findings under the AFT superpopulation. Across all panels, relative efficiency (RSS vs. SRS) increases with both ranking quality ρ and set size k. Even with moderately informative ranks ($\rho \approx 0.5$), $k \in \{6, 8, 10\}$ delivers clear gains, while strong ranking ($\rho \approx 0.9$) can approach a twofold variance reduction as k grows. Improvements are largest when the curve is evaluated early ($S(t) \in \{75\%, 50\%\}$), taper at $S(t) = 25\%$, and are smallest at $S(t) = 10\%$, where small risk sets and late-time variability dominate. Increasing censoring shifts all curves downward but does not eliminate the advantage of RSS; for $\rho \geq 0.5$ and $k \geq 6$ the gains remain visible at the first three time points even with 30–50% censoring. Comparing $m = 20$ with $m = 50$ shows little qualitative change, consistent with the theory that, at fixed $n = mk$, first-order efficiency is driven mainly by how n is allocated across ranks (through k) and by ranking quality (through ρ). The three variance ratios agree well: the Monte Carlo ratio $\mathrm{RE}_{\mathrm{MC}}$ closely tracks the larger-run benchmark $\mathrm{RE}_{\mathrm{true}}$, and the Greenwood plug-in ratio $\mathrm{RE}_{\mathrm{GW}}$ aligns with both except at $S(t) = 10\%$, where its instability under small risk sets and many ties suggests the need for a more tail-robust plug-in.

(ii) Weibull superpopulation. To check that the same efficiency pattern holds in a fully parametric family, we run a smaller simulation with exponential lifetimes

$$X \sim \mathrm{Weibull}(\nu = 1, \theta_1 = 1)$$

and independent Weibull censoring with the same shape but scales chosen so that $p_{\mathrm{cens}} \in \{0, 0.1, 0.3, 0.5\}$. Evaluation times are set at survival levels $S(t) \in \{0.75, 0.50, 0.25, 0.10\}$. Imperfect ranking again follows the Dell–Clutter rule

$$\widetilde{X} = X + Z, \qquad Z \sim N(0, \sigma_Z^2),$$

with σ_Z^2 chosen to achieve $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. We use the same balanced RSS design over $k \in \{4, 6, 8, 10\}$ and $m \in \{20, 50\}$ with $B_{\mathrm{MC}} = 1{,}000$ replicates. In this Weibull case we also compute asymptotic KM variances from the analytic kernels; the resulting relative efficiencies closely match the Monte Carlo ratios, confirming that the RSS variance reduction is not specific to the AFT generator.

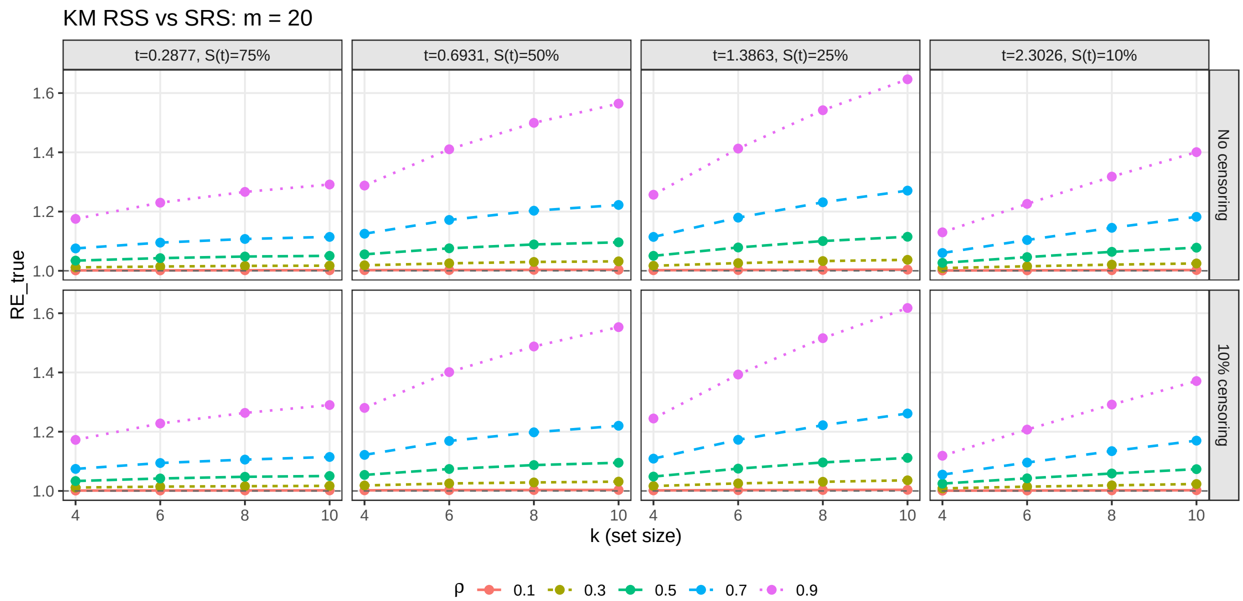


Figure 3: Weibull superpopulation (exponential baseline). True RE (RSS vs. SRS) across set size k, ranking quality ρ (legend), censoring fractions (rows), and evaluation times (columns); dashed line marks RE $= 1$. $m = 20$ cycles.

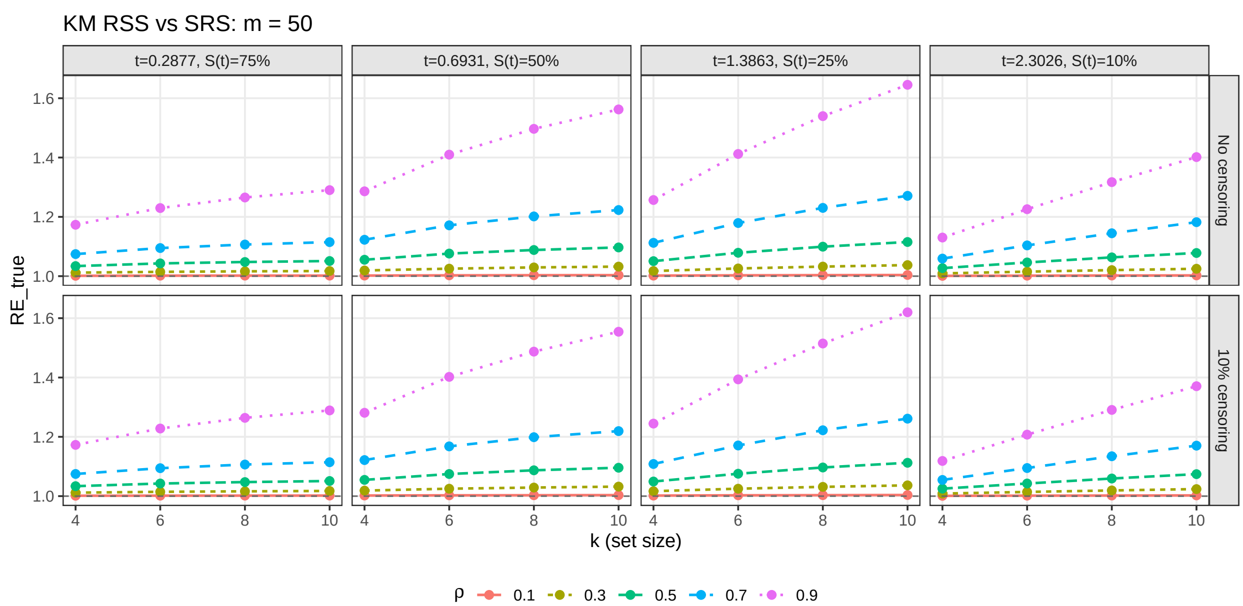


Figure 4: Weibull superpopulation (exponential baseline). True RE (RSS vs. SRS) across set size k, ranking quality ρ (legend), censoring fractions (rows), and evaluation times (columns); dashed line marks RE $= 1$. $m = 50$ cycles.

3.9 Contributions and novelty

Relative to existing work on KM under censoring (Stute and Wang, 1993; Stute, 1995) and on RSS with survival outcomes (Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2014; Mahdi Mahdizadeh and Strzałkowska-Kominiak, 2017), this paper takes as its starting point process-level CLTs for the KM and NA estimators on $[0, \tau^* - \varepsilon]$ under independent random censoring, expressed directly in terms of H_1 and S_Y and derived via the counting-process/product-integral framework of Andersen et al. (1993) and the Hadamard differentiability of the product integral (Gill and Johansen, 1990). The resulting linearizations are automatically robust to ties and yield Greenwood-type plug-ins that handle discrete event times.

We then extend these SRS results to balanced RSS with perfect ranking, obtaining laws of large numbers and functional CLTs both within rank and for the equal-weight rank average, with the limiting covariance written as the simple average of the k within-rank kernels. This provides the process-level theory needed for simultaneous bands and time-indexed tests for RSS KMs, filling a gap in earlier work that was mainly simulation-based.

Imperfect (judgment) ranking is handled via the Dell–Clutter concomitant model. Pushing the induced true/judged mixing through the KM/NA machinery gives LLNs, FCLTs, and variance kernels that depend monotonically on a ranking-quality parameter $\rho \in [0, 1]$. Within this framework we establish the efficiency ordering

$$V_{\text{perf}}(t) \leq V_{\text{judg}}(t; \rho) \leq V_{\text{SRS}}(t),$$

with equality only at $\rho = 1$ and $\rho = 0$, using conditional-variance arguments for the KM influence function.

Finally, we clarify the target when the true-rank distribution of measured units is not uniform: the rank average estimates $S^\star(t) = \sum_j P(T = j)\, S_{[j]}(t)$, which matches $S(t)$ only when $P(T = j) = 1/k$. We also indicate precisely where rank-invariant censoring and independence across ranks/cycles are used. A practical consequence is that, at fixed $n = mk$, first-order relative efficiency depends on the set size k and on ranking quality ρ, but not on the number of cycles m.

4 Discussion and future directions

The results above develop a rank-aware KM/NA theory for balanced RSS with both perfect and judgment ranking, provide Greenwood-type plug-in variance estimators, and show through simulations that the predicted variance ordering between SRS, judgment-ranked RSS, and perfectly ranked RSS is visible in finite samples. The process-level CLTs and their rank-wise decompositions give a template for extending standard survival tools to RSS designs.

A first direction is to stabilize variance estimation at late follow-up while retaining the simple rank-averaged estimator $\widehat{S}_{\mathrm{RSS}}(t) = k^{-1} \sum_r \widehat{S}_r(t)$. The simulations suggest that rank-average Greenwood inflates when per-rank risk sets are small. Tail-robust alternatives include variance-stabilizing transforms (for example the complementary log-log of $\widehat{S}_{\mathrm{RSS}}(t)$ with delta-method standard errors), mild shrinkage of the rank-average Greenwood toward a pooled-risk-set Greenwood that ignores ranks and acts as a conservative lower bound, and rank-wise multiplier bootstraps that reweight the counting and at-risk processes within each rank by positive, mean-one weights. These approaches preserve the RSS structure while improving tail calibration.

The same rank-indexed counting-process machinery should also support rank-aware versions of common two-sample survival tools and integrated functionals. Replacing the SRS counting and at-risk processes in log-rank and weighted log-rank statistics by their rank-wise RSS analogues, and averaging the corresponding variance contributions across ranks, should yield tests whose null limits follow from the martingale arguments already used here for KM, with imperfect ranking inflating the variance smoothly toward the SRS limit under the null. Likewise, integrating the rank-averaged KM process over time and applying a functional delta method ought to provide asymptotic normality and plug-in variances for restricted mean life and window mean life under RSS, enabling Wald-type inference for integrated survival when curves differ mainly early or mainly late.

Finally, the simulation patterns suggest several design implications. At fixed $n = mk$, increasing the set size k typically yields clearer efficiency gains than increasing the number of cycles m, and improving the quality of the concomitant (increasing the ranking correlation) is often more beneficial than collecting additional cycles with weak ranking. Under heavy censoring it is natural to restrict inference to $[0, \tau^* - \varepsilon]$ for a reported choice of ε, reflecting the time range where at-risk processes remain well behaved. Translating these qualitative patterns into simple design guidelines, and illustrating them on large public survival datasets from reliability or clinical research, is a natural next step for applied work building on this theory.

References

Andersen, Per K. et al. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. New York: Springer. ISBN: 978-0-387-97873-3.

Chen, Zhidong, Zhaojun Bai, and Bikas K. Sinha (2004). *Ranked Set Sampling: Theory and Applications*. New York: Springer.

Dell, Thomas R. and J. L. Clutter (1972). “Ranked set sampling theory with order statistics background”. In: *Biometrics* 28.3, pp. 545–553.

Fleming, Thomas R. and David P. Harrington (1991). *Counting Processes and Survival Analysis*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons. ISBN: 978-0-471-53385-8.

Gill, Richard D. and Søren Johansen (1990). “A survey of product-integration with a view toward multivariate product-limit estimation”. In: *Annals of Statistics* 18.4, pp. 1501–1555. DOI: 10.1214/aos/1176347879.

Mahdizadeh, Mahdi and Ewa Strzałkowska-Kominiak (2014). “On the Kaplan–Meier estimator based on ranked set samples”. In: *Journal of Statistical Computation and Simulation* 84.12, pp. 2577–2591. DOI: 10.1080/00949655.2013.794348.

Mahdizadeh, Mahdi and Ewa Strzałkowska-Kominiak (2017). “Resampling based inference for a distribution function using censored ranked set samples”. In: *Computational Statistics* 32.4, pp. 1285–1308.

McIntyre, G. A. (1952). “A method of unbiased selective sampling using ranked sets”. In: *Australian Journal of Agricultural Research* 3, pp. 385–390.

Nematolahi, Samane et al. (2020). “Improved Kaplan–Meier estimator in survival analysis based on partially rank-ordered set samples”. In: *Computational and Mathematical Methods in Medicine* 2020, Article ID 7827434, 11 pages. DOI: 10.1155/2020/7827434.

Stute, Winfried (1995). “The central limit theorem under random censorship”. In: *Annals of Statistics* 23.2, pp. 422–439.

Stute, Winfried and Jane-Ling Wang (1993). “The strong law under random censorship”. In: *Annals of Statistics* 21.3, pp. 1591–1607.

Yu, Philip L. H. and Cindy Y. Tam (2002). “Ranked set sampling in the presence of censored data”. In: *Environmetrics* 13.4, pp. 379–396. DOI: 10.1002/env.552.

Zamanzade, Ehsan, M. Mahdizadeh, and Hani M. Samawi (2024). “Nonparametric estimation of mean residual lifetime in ranked set sampling with a concomitant variable”. In: *Journal of Applied Statistics* 51.13, pp. 2512–2528. DOI: 10.1080/02664763.2023.2301334.

Zamanzade, Elham, Afshin Parvardeh, and Majid Asadi (2019). “Estimation of mean residual life based on ranked set sampling”. In: *Computational Statistics & Data Analysis* 135. Asymptotic Gaussian process; efficiency over SRS, pp. 35–55.

Zamanzade, Elham, Ehsan Zamanzade, and Afshin Parvardeh (2024). “Estimation of a decreasing mean residual life based on ranked set sampling with an application to survival analysis”. In: *International Journal of Biostatistics* 20.2, pp. 571–583. DOI: 10.1515/ijb-2023-0051.

A Technical supplements

A.1 Asymptotic toolkit for SRS

Let

$$\tau := \sup\{t \geq 0 : S(t) > 0\}, \qquad \tau^* := \sup\{t \geq 0 : S_Y(t) > 0\},$$

where $S(t) = P(X > t)$ and $S_Y(t) = P(Y \geq t) = S(t)K(t)$. Since $Y = \min(X, C)$, we have $\tau^* \leq \tau$. All uniform laws and functional CLTs in the paper are stated on $[0, \tau^* - \varepsilon]$ for some fixed $\varepsilon > 0$ so that S_Y is bounded away from 0 and the at–risk process is uniformly positive.

Write $H(t) = P(Y \leq t)$ and $H_n(t) = n^{-1} \sum_{i=1}^n I(Y_i \leq t)$, so $S_Y = 1 - H$ and $S_{Y,n} = 1 - H_n$. Because $\{I(Y \leq t) : t \in [0, \tau^* - \varepsilon]\}$ is a VC class,

$$\sup_{t \leq \tau^* - \varepsilon} |H_n(t) - H(t)| \xrightarrow{\text{a.s.}} 0, \qquad \sup_{t \leq \tau^* - \varepsilon} |S_{Y,n}(t) - S_Y(t)| \xrightarrow{\text{a.s.}} 0.$$

Choosing $c_\varepsilon > 0$ with $\inf_{t \leq \tau^* - \varepsilon} S_Y(t) \geq 2c_\varepsilon$ gives, for all large n,

$$\inf_{t \leq \tau^* - \varepsilon} S_{Y,n}(t) \geq c_\varepsilon, \qquad R(t) = n S_{Y,n}(t) \geq c_\varepsilon n \quad (t \leq \tau^* - \varepsilon),$$

which is the positivity bound used in the counting–process arguments.

Let $\Lambda_n \to \Lambda$ uniformly on $[0, \tau^* - \varepsilon]$, with Λ of bounded variation, and define product integrals

$$S_n(t) = \prod_{u \leq t} (1 - d\Lambda_n(u)), \qquad S(t) = \prod_{u \leq t} (1 - d\Lambda(u)).$$

By Hadamard differentiability of the product integral (Gill and Johansen, 1990; Fleming and Harrington, 1991),

$$\log S_n(t) - \log S(t) = -\{\Lambda_n(t) - \Lambda(t)\} + o\big(\|\Lambda_n - \Lambda\|_\infty\big)$$

uniformly on $[0, \tau^* - \varepsilon]$. Applied with $\Lambda_n = \widehat{\Lambda}_1$ and $\Lambda = \Lambda_1$ and the usual $\|\widehat{\Lambda}_1 - \Lambda_1\|_\infty =$

$O_p(n^{-1/2})$ (from the martingale/empirical–process arguments of Stute and Wang, 1993; Stute, 1995 and standard counting–process theory) this gives

$$\sup_{t \le \tau^* - \varepsilon} \left| \log \widehat{S}(t) + \widehat{\Lambda}_1(t) \right| = o_p(n^{-1/2}),$$

so the KM and NA estimators have the same $\sqrt{n}$–scale limit.

For the empirical sub-distributions

$$H_{1n}(t) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \le t, \delta_i = 1), \qquad H_1(t) = P(Y \le t, \delta = 1),$$

the VC property implies a joint functional CLT for (H_{1n}, H_n) in $\ell^\infty([0, \tau^* - \varepsilon])$. A standard add–subtract argument and linearization of $x \mapsto 1/x$ (see Stute and Wang, 1993; Stute, 1995) yield a tight Gaussian limit for $\sqrt{n}\{\widehat{\Lambda} - \Lambda\}$ and, via the product–integral map above, for $\sqrt{n}\{\widehat{S} - S\}$; this is the process-level CLT quoted in the main text.

On $[0, \tau^* - \varepsilon]$ we have $R(u)/n \to S_Y(u-)$ and $n^{-1} dN(u) \Rightarrow dH_1(u)$. A consistent variance estimator for $\widehat{\Lambda}(t)$ is

$$\widehat{\mathrm{Var}}\{\widehat{\Lambda}(t)\} = \sum_{u \le t} \frac{dN(u)}{R(u)^2},$$

see Fleming and Harrington (1991, Sec. IV.1). Combining this with the product–integral delta method gives the with–ties Greenwood estimator

$$\widehat{\mathrm{Var}}\{\widehat{S}(t)\} = \widehat{S}(t)^2 \sum_{u \le t} \frac{dN(u)}{R(u)\{R(u) - dN(u)\}},$$

used throughout the paper.

A.2 RSS layout, mixing, and Greenwood plug-ins

In balanced RSS with set size k, each cycle draws k independent candidate sets of size k, ranks within each set, and measures only the unit judged rank r in the rth set. Over m

cycles, the r–labelled observations are i.i.d. within rank and independent across $r = 1, \ldots, k$, which is the layout assumed in the RSS LLNs and CLTs.

Let $T \in \{1, \ldots, k\}$ be the true rank and $J \in \{1, \ldots, k\}$ the judged rank under the Dell–Clutter concomitant model. For any r,

$$F_r^J(t) := P(X \le t \mid J = r) = \sum_{j=1}^{k} P(T = j \mid J = r)\, F_{[j]}(t),$$

so F_r^J is a mixture of the true–rank laws $F_{[j]}$. If $P(J = r) = 1/k$ and $P(T = j) = 1/k$ for all r, j, then

$$\frac{1}{k} \sum_{r=1}^{k} F_r^J(t) = \sum_{j=1}^{k} P(T = j)\, F_{[j]}(t) = F(t),$$

whereas in general the rank average converges to $F^\star(t) = \sum_{j=1}^{k} P(T = j) F_{[j]}(t)$, as described in the main text.

For rank r, let $\widehat{S}_{[r]}(t)$ be the KM curve and let $\widehat{\mathrm{Var}}_{[r]}(t)$ denote its with–ties Greenwood variance. The RSS estimator is the equal–weight average

$$\widehat{S}_{\mathrm{RSS}}(t) = \frac{1}{k} \sum_{r=1}^{k} \widehat{S}_{[r]}(t),$$

so its variance is the average of the k rank-specific variances. The plug-in

$$\widehat{\mathrm{Var}}\{\widehat{S}_{\mathrm{RSS}}(t)\} = \frac{1}{k^2} \sum_{r=1}^{k} \widehat{\mathrm{Var}}_{[r]}(t)$$

therefore converges to the limit variance in the RSS CLT. The same construction applies with judgment ranks ($[r]$ replaced by $J = r$).