HOKKAIDO UNIVERSITY

Title	Quantifying the Systematic Error of Single-Weibull Approximations for Competing Failure Modes
Author(s)	Herrmann, Tim; Dazer, Martin
Relation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan
Citation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)
Issue Date	2026-07-01
Doc URL	https://hdl.handle.net/2115/100449
Type	conference paper
File Information	C000088_1.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

10th International Symposium on Reliability Engineering and Risk Management (*ISRERM 2026*),
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

Quantifying the Systematic Error of Single-Weibull Approximations for Competing Failure Modes

Tim Herrmann[1,] * and Martin Dazer[2]

[1,2]*Institute of Machine Components, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany.*
E-mail: tim.herrmann@ima.uni-stuttgart.de
E-mail: martin.dazer@ima.uni-stuttgart.de
**Corresponding Author*

Fitting a single Weibull distribution to all failure data is a questionable simplification for systems with multiple competing failure modes. This paper quantifies the systematic error of this approach to determine the boundaries within which this simplification remains acceptable compared to exact competing risks models. To establish these boundaries, Monte-Carlo simulations are employed to evaluate the coverage probability of the lower 90% confidence bound for the B_{10}-lifetime. When all failure causes are known, the single-Weibull method only provides valid estimates if the competing modes share similar shape parameters. Otherwise, it systematically overestimates the lifetime. For masked failure data, a mathematically exact joint likelihood formulation exists. However, computational complexity and limited software support restrict its practical use. Therefore, engineers often use a standard simplification, treating masked failures as suspensions. This study compares the baseline single-Weibull approach against both the exact method and this suspension heuristic. Results show that at high masking, the heuristic introduces a severe positive bias. While single-Weibull provides more reliable coverage here, both simplifications collapse completely for fundamentally different failure modes. Finally, the paper provides a practical decision framework to establish strict validity limits for these heuristics. Based on expected failure physics, sample size, and masking, it defines where simplifications remain applicable versus where the exact approach is strictly required.

Keywords: Competing Failure Modes, Masked Failure Data, Weibull Analysis, System Reliability, Decision Framework.

1 Introduction

Weibull analysis is the standard tool for reliability assessment in industrial practice (Abernethy 2006). To account for estimation uncertainty, engineers typically rely on characteristic life metrics, particularly the lower confidence bound of the B_x-lifetimes. As this bound – particularly the B_{10}-lifetime – is a central decision criterion in industry, it serves as the primary comparative metric throughout this study. The fundamental assumption of this single-Weibull approach is that all failures originate from a single underlying cause. However, complex systems are actually subject to multiple competing failure modes (Nelson 1982). The statistically correct competing risks (CR) framework models each mode independently, but this requires assigning every observed failure to a specific cause. When this identification is entirely missing, engineers often fit a single Weibull distribution to all data, simplifying the analysis but ignoring the physical reality. A more complex challenge arises with partially identified causes, a scenario known as masked failure data. Here, the mathematically exact approach (Guess et al. 1991) is rarely supported by commercial software (ReliaSoft 2021) due to computational complexities and sensitivity to initial values. Consequently, engineers frequently rely on a simplified heuristic, treating masked failures as independent right-censored observations (suspensions). While the statistical theory for competing risks is well established (Meeker et al. 2022), the practical consequences of these common simplifications remain poorly quantified. This paper addresses this gap by systematically evaluating the coverage probability of the common confidence bounds. First, the estimation error of the single-Weibull approach is characterized across a broad parameter space without masking. Second, the analysis extends to masked data scenarios to directly compare the single-Weibull approach, the simplified suspension heuristic, and the mathematically exact competing risks benchmark. The study highlights the numerical and statistical pitfalls of standard practices and condenses the findings into a practical decision framework. This framework helps engineers identify the most appropriate statistical model for their specific scenario.

2 Methodology

This section describes the statistical models, the confidence interval estimation and the simulation framework.

2.1 *Competing risks model*

In a competing risks setting, each failure mode is described by an independent two-parameter Weibull distribution with shape parameter β_k and scale parameter η_k:

$$F_k(t) = 1 - \exp\left(-\left(\frac{t}{\eta_k}\right)^{\beta_k}\right) \tag{1}$$

Under a series system assumption, the first occurring failure mode determines the system lifetime (Nelson 1982). For two modes A and B, the system reliability is simply the product of the reliabilities, $R_{sys}(t) = R_A(t) \cdot R_B(t)$, the system failure probability follows as $F_{sys}(t) = 1 - R_{sys}(t)$ and the system probability density function (PDF) is:

$$f_{sys}(t) = f_A(t) \cdot R_B(t) + f_B(t) \cdot R_A(t) \tag{2}$$

where $f_A \cdot R_B$ describes failures caused by mode A given that mode B has not yet occurred, and $f_B \cdot R_A$ correspondingly for mode B. The system characteristic lifetime, such as the B_{10}-quantile, is obtained numerically by solving $F_{sys}(t) = 0.10$.

When the exact cause of failure is unknown (masked data), the statistically exact approach formulates a joint system likelihood (Usher and Hodgson 1988). As established by Guess et al. (1991), this assumes the masking probability is independent of the true failure cause (i.e., masking occurs equally likely for mode A or B). By partitioning the dataset into index sets for known causes (A and B) and masked observations (M), the likelihood contribution for ambiguous failures ($i \in M$) is formulated as the sum of the probabilities of failing from either mode A or mode B:

$$L = \prod_{i \epsilon A} f_A(t_i) R_B(t_i) \cdot \prod_{i \epsilon B} f_B(t_i) R_A(t_i) \cdot \prod_{i \epsilon M} \left(f_A(t_i) R_B(t_i) + f_B(t_i) \cdot R_A(t_i) \right) \tag{3}$$

Maximizing this exact function, referred to as the Joint System CR, requires simultaneous optimization of all four parameters ($\beta_A, \eta_A, \beta_B, \eta_B$). To ensure numerical stability and convergence within this highly coupled optimization space, the parameter estimates obtained from a single-Weibull fit are utilized as the starting values. This exact method is evaluated against the two common industry simplifications: the Separate CR approach, which models the modes independently by treating masked failures as right-censored data, and the single-Weibull baseline, which fits one distribution to all data by effectively ignoring the underlying competing risks structure.

2.2 *Confidence interval estimation*

Confidence intervals for the B_x-quantile are obtained on the logarithmic scale using the delta method (Meeker et al. 2022). For the single-Weibull and Separate CR approaches, the covariance matrix is derived by inverting the 2×2 Hessian matrix H of the negative log-likelihood function ($Cov(\theta) = H^{-1} \approx F^{-1}$). For the Joint System CR, a full 4×4 Hessian matrix must be inverted to capture the cross-correlations of the coupled parameters.

The variance of the logarithmic B_x estimate is obtained by propagating the uncertainty of the parameter vector $\hat{\theta} = \left[\hat{\beta}, \hat{\eta}\right]^T$ ($\hat{\theta} = \left[\hat{\beta}_A, \hat{\eta}_A, \hat{\beta}_B, \hat{\eta}_B\right]^T$ for Joint System CR) through the quantile function:

$$Var\left(ln\left(B_x(\hat{\theta})\right)\right) = \left(\frac{\partial\, ln\left(B_x(\hat{\theta})\right)}{\partial \theta}\right)^T F^{-1} \left(\frac{\partial\, ln\left(B_x(\hat{\theta})\right)}{\partial \theta}\right) \tag{4}$$

The one-sided lower confidence bound is then calculated as

$$CI_{lower} = exp\left(ln\left(B_x(\hat{\theta})\right) - Z_\alpha \sqrt{Var\left(ln\left(B_x(\hat{\theta})\right)\right)}\right) \tag{5}$$

where Z_α is the standard normal quantile for the desired confidence level.

2.3 *Simulation framework*

The performance of the evaluated methods is assessed through a Monte-Carlo simulation study. For each parameter configuration, failure times for both failure modes are drawn independently from their respective Weibull distributions. The observed system lifetime is determined as the minimum of the two competing failure times, and the corresponding failure cause is recorded. This process is repeated to generate individual samples of size N, reflecting the natural competition between failure modes under realistic stochastic sampling variability. Aggregating a large number of these samples provides stable estimates of the coverage probability, which are then evaluated against the analytically known true system quantile.

Three simulation studies are conducted. The parameter ranges, target quantiles, and confidence levels were specifically chosen to reflect the typical spectrum of failure behaviors and standard benchmarks in industrial reliability engineering. Furthermore, sample sizes and masking ratios were varied extensively to ensure a comprehensive overview of realistic data conditions. First, to isolate the fundamental effect of varying shape parameters, a discrete full factorial design evaluates 28 pairwise combinations of $\beta \in \{0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0\}$ (with $\beta_A \leq \beta_B$) across nine sample sizes $N \in [15, 300]$, using fixed scale parameters, no masking, and 2000 replications. Second, to generalize these findings across the entire parameter space and account for scale parameter interactions, a sensitivity analysis utilizes Latin Hypercube Sampling (LHS) to explore 5000 configurations in the continuous parameter space

$(\beta_A, \beta_B, \eta_A, \eta_B \in [0.5, 5.0], N \in [10, 300])$, evaluated across combinations of target quantiles $B_x \in \{0.01, 0.05, 0.1\}$ and nominal confidence levels $CI_{lb} \in \{0.90, 0.95, 0.99\}$. Finally, to isolate the masking effect, a second LHS generates 750 configurations ($\beta_A, \beta_B, \eta_A, \eta_B \in [0.5, 5.0]$), each evaluated at discrete sizes $N \in \{20, 50, 100, 300\}$ across four masking ratios $m \in \{0\%, 10\%, 25\%, 50\%\}$ with 2000 replications. Masked data are generated by independently withholding the cause-of-failure label with a fixed probability m, ensuring independence from actual failure times and modes, to directly compare the single-Weibull, Separate CR, and Joint System CR methods.

3 Characterization of the Single-Weibull Approximation Error

To establish a baseline, this section characterizes the systematic error introduced by the single-Weibull approach when masking is absent. The primary metric is the coverage probability of the lower one-sided 90% confidence bound for the B_{10}-lifetime.

Figure 2 illustrates the coverage probability of the single-Weibull approach across the shape parameter space for different sample sizes. As expected, the approach performs well when both competing failure modes share similar or identical shape parameters. In these regions, the combined failure behavior closely resembles a single Weibull distribution. Even for perfectly identical shape parameters, the observed coverage is slightly below the nominal target of 90% (e.g., 82% at $N = 20$). This is a known consequence of the finite-sample properties of Maximum Likelihood Estimation and the normal approximation used within the delta method, which inherently compute slightly narrow confidence intervals for small datasets (Meeker et al. 2022). As the difference between the shape parameters increases, the performance of the single-Weibull approach drops significantly. The method systematically overestimates the true reliability, causing the lower confidence bound to frequently fall above the true B_{10}-lifetime. Consequently, coverage drops to unacceptable levels, approaching 0% in extreme cases. For reference, the dedicated competing risks model (note that the Separate CR and Joint System CR approaches are mathematically identical in the absence of masking) was also evaluated across the same grid. The results confirm that the competing risks model maintains coverage probabilities close to the nominal 90% across the entire parameter space.

To explain this coverage drop, Figure 3 shows the relative bias of the B_{10} point estimates for both methods. The single-Weibull bias is almost exclusively a function of $\Delta\beta$ and is virtually independent of the sample size N. This explains the surprising coverage behavior observed in Figure 2, where larger sample sizes sometimes produce worse coverage. Because the systematic bias remains constant, the sample size dictates the width of the confidence interval. At small sample sizes, the wide confidence intervals hide the underlying bias, resulting in an artificially high coverage. As N increases, the variance shrinks, and the narrow confidence intervals fully expose the systematic overestimation, causing the coverage probability to drop significantly. The competing risks model, in contrast, shows negligible bias regardless of the parameter combination or sample size.

Finally, the sensitivity of these findings to different target quantiles (B_1, B_5, B_{10}) and confidence levels (90%, 95%, 99%) is evaluated using a continuous LHS study. Figure 4 visualizes the coverage probability of the single-Weibull approach (bold) alongside the exact competing risks model (italic). To isolate these effects, configurations are grouped by $\Delta\beta$, following the primary trend observed in Figure 2. The sample sizes (N) below each column indicate how many configurations fell into the respective category. The results show a clear trend. First, lower quantiles (such as B_1) show a significantly larger coverage deficit than B_{10}. This occurs because lower quantiles require extrapolation further into the lower tail of the distribution, where a single Weibull fit, which averages out the distinct early and late failure dynamics, deviates the most from the true system behavior. Second, higher nominal confidence levels compute wider intervals. This increased width acts as an artificial buffer that absorbs a portion of the systematic bias, making 95% or 99% confidence bounds appear slightly more robust against model misspecification. Overall, the pattern confirms that the error is mostly driven by $\Delta\beta$, and the choice of a 90% confidence interval for the B_{10}-lifetime provides a solid baseline for the subsequent masking analysis.

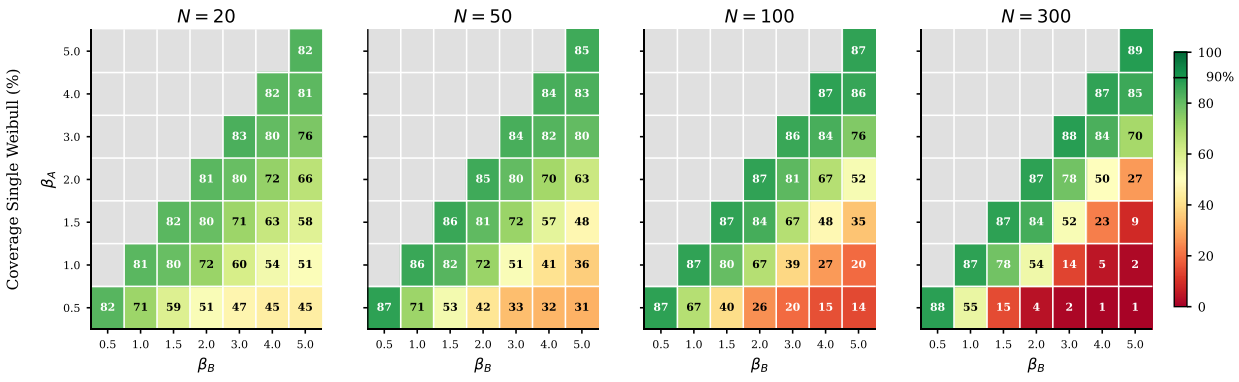

Figure 2. Coverage probability of the single-Weibull 90% confidence bound for the B_{10}-lifetime across varying shape parameters and sample sizes *N*.

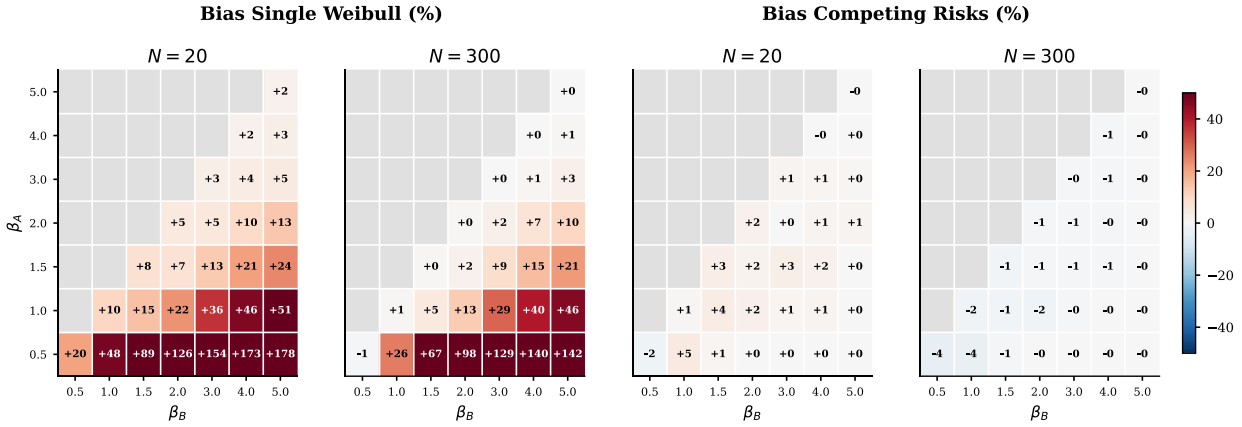

Figure 3. Relative bias of the B_{10} point estimates for the single-Weibull and competing risks models at $N \in \{20, 300\}$.

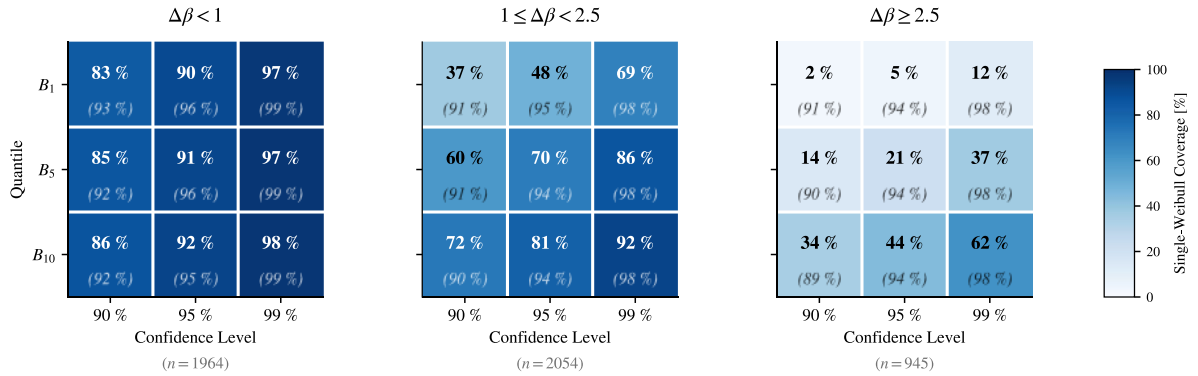

Figure 4. Coverage probability of the single-Weibull approach for different target quantiles and confidence levels, grouped by shape parameter difference $\Delta\beta$ (*Single-Weibull: bold, Competing Risks: italic*).

4 Impact of Masked Failure Data on Estimation Accuracy

In industrial practice, the exact cause of a system failure is not always identified. This lack of information, known as masked failure data, poses a significant challenge for reliability analysis. As introduced in Section 2, the Joint System CR approach models masked observations through a likelihood that sums the contributions of all known failure modes. However, this exact formulation explicitly assumes that all relevant failure modes are known and accounted for. Because it distributes the full probability of each masked failure exclusively among the modeled causes ($f_A R_B + f_B R_A$), any unmodeled, unknown failure mode (e.g., a new manufacturing defect) would be misattributed, which distorts the parameters of the known modes. Furthermore, this joint estimation requires simultaneous 4-dimensional optimization, which is not supported by standard commercial reliability software. Consequently, engineers commonly resort to the simpler Separate CR approach. While this suspension heuristic avoids explicit misattribution, it inevitably leads to information loss. This section investigates how both competing risks models compare against the masking-invariant single-Weibull baseline.

The practical impact of masked data on the statistical estimation is visualized in Figure 5, which shows the distribution of the lower confidence bounds for the B_{10}-lifetime for dissimilar β. Because the single-Weibull approach (red dashed line) ignores failure causes, its estimation remains completely unaffected by masking. By treating masked failures as suspensions, the Separate CR simplification (blue line) mathematically assumes the component survived both modes, completely ignoring the actual system failure. Consequently, as the masking ratio increases, its distribution flattens (indicating higher variance) and shifts significantly to the right (large positive bias). In contrast, the exact Joint System CR method (green line) correctly processes the probabilistic failure contributions. Its distribution stays centered near the true B_{10}-lifetime, eliminating the bias with only a minor variance increase due to the missing labels.

Notably, the Joint System CR optimization proved to be highly sensitive to the choice of initial parameters. When using the biased Separate CR estimates as starting values, the numerical optimization frequently failed to produce a valid, invertible Hessian matrix. In contrast, initializing the algorithm with the Single Weibull parameters provided significantly more stable starting points. This extreme sensitivity highlights the numerical instability of the joint likelihood surface under masked data, explaining why statistical research often pivots to alternative, more robust approaches like the Expectation-Maximization (EM) algorithm (Craiu and Duchesne 2004).

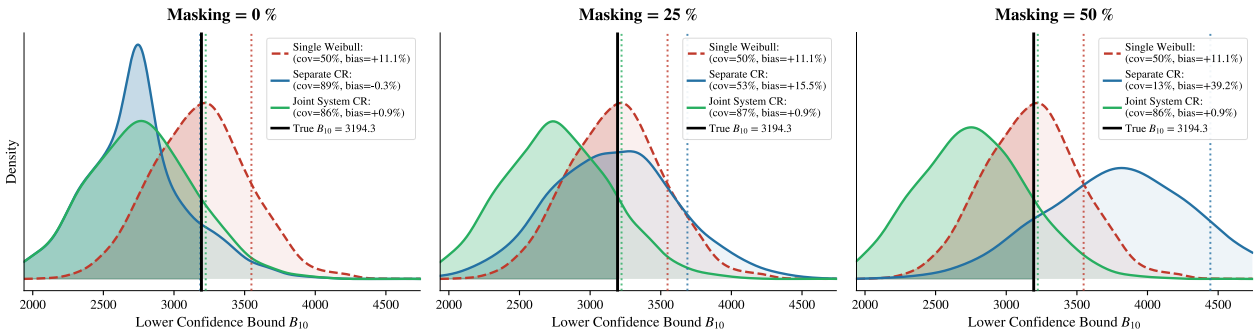

Figure 5. Distribution of the lower 90% confidence bounds for the B_{10}-lifetime under varying masking levels.

Figure 6 expands this observation across the continuous parameter space to show the average coverage probability as a function of the shape parameter difference $\Delta\beta$ and sample size N. The red line represents the single-Weibull approach, which drops sharply as $\Delta\beta$ increases. The shaded band around it indicates the standard deviation of the coverage across the evaluated parameter space, showing large variation depending on the specific Weibull parameters. The Separate CR model (dashed lines) shows almost no variance across the parameter space, indicating that its error is highly systematic and driven by the masking ratio rather than $\Delta\beta$. A surprising effect can be observed regarding the sample size: as N increases, the coverage probability of the heavily masked Separate CR models (e.g., 25% and 50%) actually decreases. However, because the underlying point estimate is heavily biased by the censoring simplification, this narrow interval completely misses the true value. The Joint System CR (dotted lines) perfectly maintains the nominal 90% coverage across all masking levels and parameter combinations, confirming it as the mathematical gold standard. However, if this exact method is unavailable, engineers must choose between the two simplified heuristics. As the data shows, the masking-invariant single-Weibull approach eventually overtakes the Separate CR simplification at a complex crossing point to provide a safer coverage probability. The following section introduces a comprehensive decision framework that precisely locates this crossing point and defines the exact application boundaries for both heuristics.

5 Decision Framework

To bridge statistical theory and industrial application, Figure 7 provides a practical decision framework for selecting the most robust estimation method. Since exact shape parameters are rarely known a priori, the continuous $\Delta\beta$ parameter space from Section 4 is translated into three practical engineering categories based on expected failure physics: two competing wear-out modes (e.g., fatigue and abrasion, $2 < \beta_A, \beta_B < 5$), a mix of wear-out and early-life failures (e.g., fatigue vs. manufacturing defects, $\beta > 2$ vs. $\beta < 1$), and remaining mixed configurations.

Figure 7 evaluates the coverage achieved in 75% of the parameter scenarios, indicating the superior method in blue (Separate CR) and red (single-Weibull). Cells are hatched if this baseline falls below 80%. This 80% limit is a critical industrial boundary: dropping from the nominal 90% target to an 80% coverage effectively doubles the risk (from 10% to 20%) that the estimated lower bound will overestimate the true component lifetime. Because extreme variations in failure characteristics within these broad categories significantly degrade estimation accuracy, enforcing a strict 90% worst-case assurance would cause all simplified methods to be universally inadequate across all masking levels. Therefore, this 75% scenario threshold offers a realistic engineering compromise, defining where heuristics remain practically viable while clearly marking where they break down.

This framework establishes two primary guidelines. First, for similar failure mechanisms (Both Wear-Out), the exact Joint System CR approach is not strictly necessary. Barring extreme cases with highly divergent wear-out characteristics, engineers can apply the single-Weibull approach for higher masking levels. Second, for fundamentally

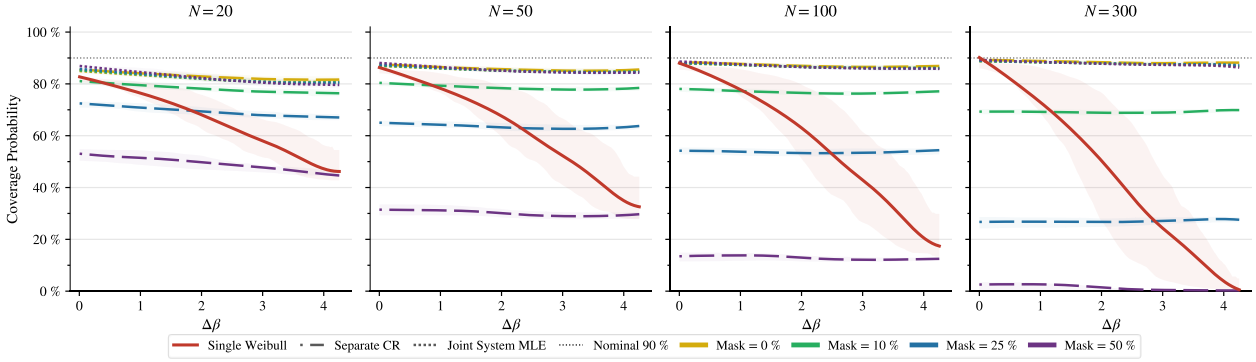

Figure 6. Average coverage probability across varying masking levels, shape parameter differences ($\Delta\beta$) and sample sizes N.

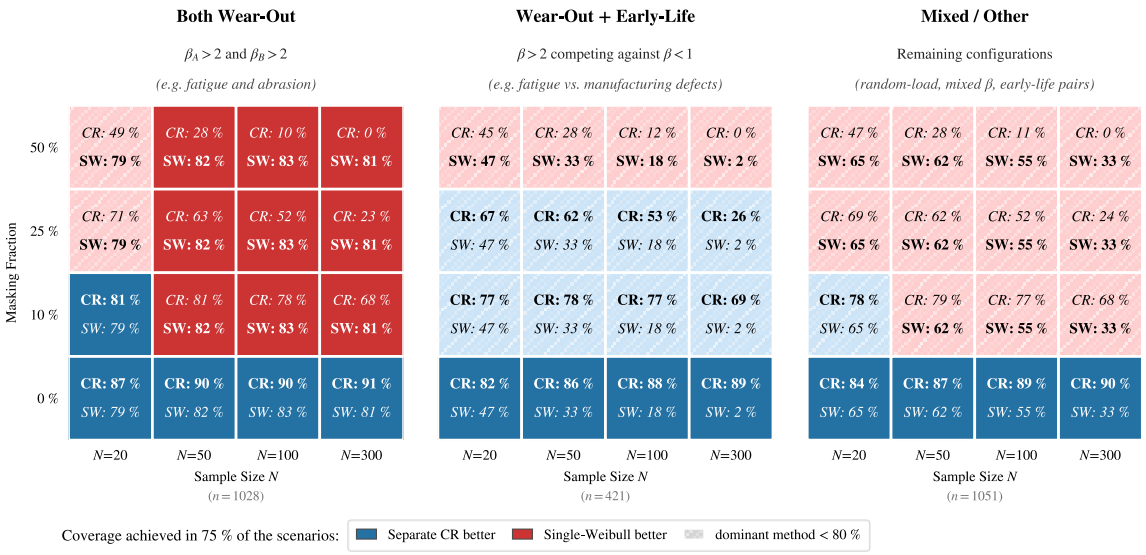

Figure 7. Decision framework defining the application boundaries of the Separate CR and single-Weibull simplifications, displaying the coverage achieved in 75% of the parameter scenarios.

different failure modes (Wear-Out + Early-Life), the Separate CR model outperforms the heavily biased single-Weibull approach. However, except at minimal masking (< 10%), both heuristics degrade so severely (e.g., achieving only 26% coverage at 25% masking) that they are practically unusable. Consequently, when significant masking coincides with dissimilar failure modes, the simplified methods must be abandoned in favor of the exact Joint System CR approach.

6 Conclusion

This paper quantified the systematic error of applying the single-Weibull approach and the Separate CR simplification to systems with multiple competing failure modes. By evaluating the coverage probability of the lower 90% confidence bound for the B_{10}-lifetime, the study revealed critical limitations and unexpected results in standard reliability practices. Under ideal conditions with known failure causes, the single-Weibull approach provides valid estimates only if competing modes share similar shape parameters. If the failure modes differ significantly, it systematically overestimates the reliable lifetime. This severe positive bias is independent of the sample size. Consequently, larger sample sizes actually worsen the coverage probability, because the narrower confidence intervals fully expose the underlying error. Under masked data scenarios, the mathematically exact Joint System CR approach perfectly maintains the nominal coverage. However, its strict closed-world assumption and the limitations of commercial software frequently force engineers to use the Separate CR simplification. This suspension heuristic discards critical failure information, artificially stretching the estimated lifetime. As a result, at high masking ratios, the Separate CR model performs significantly worse than the single-Weibull approach, a systematic error that is surprisingly amplified by larger sample sizes. The derived decision framework establishes strict boundaries for these simplifications: even for systems with similar failure mechanisms (e.g., concurrent wear-out), the simplified methods only provide acceptable estimates provided the underlying wear-out characteristics do not differ drastically. If extreme parameter variations exist, or if fundamentally different failure modes coincide with significant masking, both standard heuristics collapse completely, resulting in severe coverage deficits. In these inadequate scenarios, the exact Joint System CR approach becomes strictly necessary. Ultimately, engineers must carefully match their statistical models to the expected physical failure characteristics, sample size, and data quality. However, whenever analytical expertise and software capabilities allow, the exact Joint System CR approach is strongly recommended as the only guaranteed optimal solution for any masked scenario, ideally initialized with single-Weibull parameters to ensure robust numerical convergence.

References

Abernethy, R. B., *The New Weibull Handbook*, 5th ed., Robert B. Abernethy, 2006.

Craiu, R. V. and Duchesne, T., Inference Based on the EM Algorithm for the Competing Risks Model with Masked Causes of Failure, *Biometrika*, 91(3), 543-558, 2004.

Guess, F. M., Usher, J. S., and Hodgson, T. J., Estimating system and component reliabilities under partial information on cause of failure, *J. Stat. Plann. Inference*, 29(1-2), 75-85, 1991.

Meeker, W. Q., Escobar, L. A., and Pascual, F. G., *Statistical Methods for Reliability Data*, 2nd ed., John Wiley & Sons, 2022.

Nelson, W., *Applied Life Data Analysis*, John Wiley & Sons, 1982.

ReliaSoft, *Weibull++ User's Guide*, Version 2021, Hottinger Bruel & Kjaer UK Ltd, 2021.

Usher, J. S. and Hodgson, T. J., Maximum likelihood analysis of component reliability using masked system life-test data, *IEEE Transactions on Reliability*, 37(5), 550-555, 1988.