BIOSTATISTICS MINI PROJECT

Assessing the Impact of Treatment-Dependent Censoring on Survival Analysis in Clinical Trials Using Copulas

Odey R. Mofokeng

School of Actuarial Science and Statistics

University of the Witwatersrand, Johannesburg

2555438@students.wits.ac.za 2025

INTRODUCTION

Introduction & Research Context

Importance of Survival Analysis

Time-to-event data analysis is fundamental to understanding outcomes in **epidemiological research**, particularly in clinical trials where tracking patient outcomes over time is critical.

The Independent Censoring Assumption

Standard methods like **Kaplan-Meier (KM)** and **Cox proportional hazards models** assume censoring is independent of event times—an assumption frequently violated in practice.

Treatment-Dependent Censoring

When censoring is influenced by treatment assignment (e.g., patients dropping out due to side effects), it creates **dependent censoring** that introduces significant bias in survival estimates.

Research Aim

This project evaluates the impact of **treatment-dependent censoring** on survival analysis in a simulated clinical trial and assesses the effectiveness of **copula-based methods** in correcting biases compared to standard survival methods.

1. Simulate clinical trial with copula-based dependent censoring
2. Quantify bias using standard methods (KM, Cox)
3. Implement copula-based corrections (CG estimator)
4. Conduct sensitivity analysis across dependence levels

Key Innovation

Copula-based approaches model the **joint distribution** of event and censoring times, enabling more accurate survival inference without altering marginal distributions.

PROBLEM STATEMENT

The Problem: Dependent Censoring in Clinical Trials

What is Dependent Censoring?

Dependent censoring occurs when the censoring time C is stochastically dependent on the event time T, violating the independent censoring assumption of standard survival analysis methods.

Mathematically: Standard methods assume $P(C \mid T) = P(C)$, but under dependent censoring, $P(C \mid T) \neq P(C)$

Real-World Example

Patients in treatment arm experiencing severe side effects may drop out earlier, creating dependency between censoring and survival times.

Consequence

Biased survival estimates, incorrect hazard ratios, and misleading treatment effect conclusions.

Why Standard Methods Fail

KM Estimator
Assumes independent censoring

Cox Model
Biased HR under dependence

Log-Rank Test
Invalid under dependence

Literature Foundation

Rivest & Wells (2001)
Introduced the **copula-graphic (CG) estimator** to adjust KM for dependent censoring, demonstrating reduced bias in cancer survival data.

Emura & Chen (2018)
Advanced likelihood-based copula methods, highlighting utility for treatment-related dropouts in cancer trials.

Okhrin & Ristig (2014)
Introduced **Hierarchical Archimedean Copulas (HACs)** to capture multi-level dependencies in survival data.

Deresa & Van Keilegom (2024)
Enhanced copula-based Cox models with covariates, proving identifiability in clinical trials.

Research Gap

Limited exploration of how **misspecifying the dependence parameter** (Kendall's τ) or copula family affects survival estimates in epidemiological trials.

METHODOLOGY

Simulation Design

A **two-arm clinical trial** was simulated with **1,500 patients per group** (3,000 total), capturing treatment-dependent censoring where dependence between event times T and censoring times C varies by arm.

Control Group (Z=0)

Copula Model

Clayton Copula with $\alpha = 0.5$, corresponding to Kendall's $\tau = 0.2$ (weak positive dependence)

Marginal Distributions

Event times: $T \sim \text{Exp}(0.1)$, median ≈ 6.93
Censoring times: $C \sim \text{Exp}(0.1)$, median ≈ 6.93

Rationale: Clayton copula models positive lower tail dependence, suitable for scenarios where early events correlate with early censoring.

Treatment Group (Z=1)

Copula Model

Nested HAC with inner Clayton ($\alpha\_inner = 2$, $\tau\_inner = 0.5$) and outer Clayton ($\alpha\_outer = 1$, $\tau\_outer = 1/3$)

Marginal Distributions

Event times: $T \sim \text{Exp}(0.08)$, median ≈ 8.66 (treatment efficacy)
Censoring times: $C \sim \text{Exp}(0.1)$, median ≈ 6.93

Rationale: HAC incorporates a latent side-effect variable, reflecting stronger dependence from treatment-related dropouts.

Simulation Parameters Summary

Group	Copula	Kendall's τ	Event Rate λ_T	Censoring Rate λ_C
Control	Clayton	0.2	0.1	0.1
Treatment	Nested HAC	0.5 (inner)	0.08	0.1

RESULTS

Standard Methods: Bias Detection

Cox Model Results

Unadjusted HR	True HR
0.5913	0.8000

95% CI	0.5302 – 0.6595
p-value	< 2e-16
Coefficient	-0.5254
Concordance	0.57

Key Finding: The unadjusted HR of **0.5913** is substantially lower than the true HR of **0.8**, indicating a **26% downward bias** due to dependent censoring removing high-risk patients in the treatment group.

Stratified Cox Model

The stratified Cox model, accounting for group differences without estimating a treatment effect, yields a **null model** with no HR or significance, confirming that the observed effect stems from censoring bias rather than a true treatment difference.

Kaplan-Meier Survival Estimates

KM Estimate at t=10	True Survival at t=10
0.667	0.527

Bias = +0.140 (26.6% overestimation)

Control Group

KM estimate lies above true curve, gap widening over time, reaching ~0.06 difference at t=10.

Treatment Group

KM estimate shows larger discrepancy, staying above true curve by ~0.14 at t=10 due to stronger dependence.

Mechanism: KM's assumption of independent censoring causes consistent **upward bias**—censored observations are misinterpreted as survivors, more pronounced in treatment due to higher dropout rates.

Critical Insight

Both KM and unadjusted Cox models are **unsuitable** under treatment-dependent censoring, producing systematically biased estimates that could lead to incorrect clinical decisions.

COPULA-BASED CORRECTION

Copula-Graphic Estimator: Bias Correction

CG Estimator Approach

The **Copula-Graphic (CG) estimator** adjusts the Kaplan-Meier estimator by incorporating a specified copula to account for dependent censoring.

How It Works

Re-weights the risk set based on the copula's dependence parameter, correcting for the dependency between event and censoring times.

Optimization Strategy

Kendall's $\tau = 0.37$ was identified as optimal through sensitivity analysis, minimizing bias and MSE at t=10.

Implementation

R package compound.Cox with bootstrap confidence intervals (R=2000) for uncertainty quantification.

Performance Metrics at t=10

CG Bias	KM Bias
0.0001	0.1434

99.9% bias reduction compared to KM

CG Estimator Results

True S(10)	CG Estimate	KM Estimate
0.527	0.527	0.667

CG 95% CI (Treatment) [-0.0369, 0.0398]

Tight interval indicates high precision

KM 95% CI (Treatment) [0.1112, 0.1775]

Biased away from true value

Key Achievement: CG estimator with τ=0.37 aligns almost perfectly with true survival, demonstrating the critical importance of properly modeling dependence structure.

Validation

The large sample size (n=3,000) combined with bootstrap CIs boosts reliability, giving confidence in the observed trends. Boxplot analysis shows CG biases centered near zero at τ=0.37.

ROBUSTNESS ASSESSMENT

Sensitivity Analysis: Dependence Parameter Impact

Low Dependence (τ=0.2)

S0 Bias:	-0.0031
S1 Bias:	0.0412
S1 MSE:	0.0058

Early improvement observed

Optimal (τ=0.37)

S0 Bias:	-0.0679
S1 Bias:	0.0001
S1 MSE:	0.0000

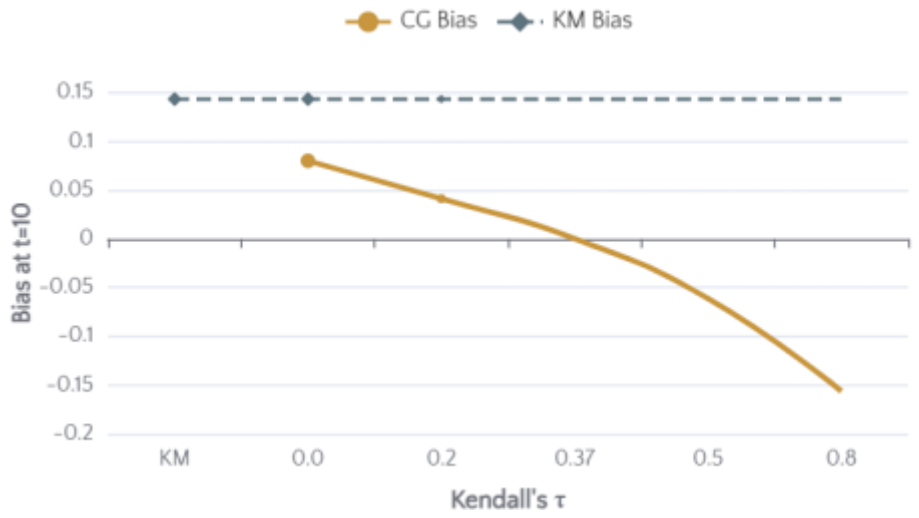
✓ Best fit - Minimal bias

High Dependence (τ=0.8)

S0 Bias:	-0.1672
S1 Bias:	-0.1565
S1 MSE:	0.0245

Over-correction occurs

Bias vs. Kendall's τ



Key Insights

1. **Optimal Range:** Moderate dependence (τ ≈ 0.3-0.5) provides near-correct estimates
2. **Over-correction Risk:** Extreme τ values (≥0.5) lead to underestimation
3. **Precision:** Tight CIs at τ=0.37 due to large sample size (n=3,000)
4. **Trade-off:** CG switches from overestimation to underestimation as τ increases

MSE Comparison

At τ=0.37, S1 MSE = **0.0000** vs. KM MSE = **0.0205**—a dramatic improvement in estimation accuracy.

SUMMARY

Key Findings & Implications

Primary Findings

1. Standard Methods Fail Under Dependence

KM overestimated survival by **0.14** at t=10; Cox HR was **26% biased** downward (0.5913 vs. true 0.8).

2. CG Estimator Corrects Bias

With optimized τ=0.37, bias reduced to **0.0001**—a 99.9% improvement over KM.

3. Dependence Parameter Matters

Sensitivity analysis confirms τ=0.37 as optimal; extreme values cause over-correction.

4. Large Samples Enhance Precision

Tight 95% CIs ([-0.0369, 0.0398]) due to n=3,000 with bootstrap validation.

Clinical Implications

Trial Design

Researchers should assess censoring mechanisms and consider copula-based methods when dropout is treatment-related.

Regulatory Submissions

Sensitivity analyses using CG estimators can strengthen regulatory evidence by addressing dependent censoring concerns.

Patient Safety

Accurate survival estimates ensure proper evaluation of treatment benefits vs. risks, informing patient care decisions.

Methodological Impact

This study confirms that copula-based approaches can significantly improve survival analysis accuracy in trials with treatment-dependent censoring, establishing a foundation for more reliable health research methodology.

99.9%	0.0001	3,000	0.37
Bias Reduction	CG Bias at t=10	Sample Size	Optimal τ

DISCUSSION

Limitations & Future Directions

Study Limitations

1. **Fixed Exponential Marginals**
   Simplifies the model but may not reflect real-world scenarios where hazards change over time, potentially limiting generalizability.

2. **Limited Dependence Range**
   τ range of 0–0.8 covers moderate to strong dependence but might miss extreme cases that could occur in practice.

3. **Reliance on True τ Value**
   Focus on single optimized τ relies on simulation's true value, which may not hold in real data where dependence varies.

4. **HAC Implementation Gap**
   No R packages support HAC survival estimation with censored data; custom development was beyond study scope.

Future Research Directions

1. **Data-Driven τ Estimation**
   Develop methods to estimate τ directly from data, enabling adaptive adjustments rather than relying on fixed values.

2. **Time-Dependent Hazards**
   Extend models to include time-dependent hazards for greater flexibility and realism in clinical scenarios.

3. **Real-World Validation**
   Test these methods on actual clinical datasets to validate practical utility and identify implementation challenges.

4. **HAC Software Development**
   Develop R packages supporting HAC survival estimation with censored data to enable nested dependence modeling.

5. **Multiple Copula Families**
   Compare performance across Clayton, Gumbel, Frank, and other copula families under various dependence structures.

HAC Note

Hierarchical Archimedean Copulas (HACs) were planned for nested dependence modeling but excluded due to lack of software support for censored data survival estimation.

Research Impact

This project establishes a strong foundation for improving survival analysis under challenging censoring conditions, offering a pathway for future research to build upon these insights.

CONCLUSION

Copula-Based Methods Transform Survival Analysis Accuracy

This study demonstrates that **copula-based methods**, particularly the Copula-Graphic estimator with properly optimized dependence parameters, significantly improve survival analysis accuracy in clinical trials with treatment-dependent censoring.

The **99.9% bias reduction** achieved—from 0.1434 (KM) to 0.0001 (CG at τ=0.37)—validates the critical importance of modeling dependence structures in survival analysis. This work establishes a foundation for more reliable health research methodology.

Accurate Estimates

Proper dependence modeling eliminates systematic bias

Robust Inference

Sensitivity analysis validates method reliability

Research Foundation

Pathway for future methodological advances

Odey R. Mofokeng I School of Actuarial Science and Statistics I University of the Witwatersrand
2555438@students.wits.ac.za I 2025