Ordinal Outcome Regression with Censored Covariates in the Presence of a Cured Fraction

אוניברסיטת חיפה
UNIVERSITY OF HAIFA
جامعة حيفا

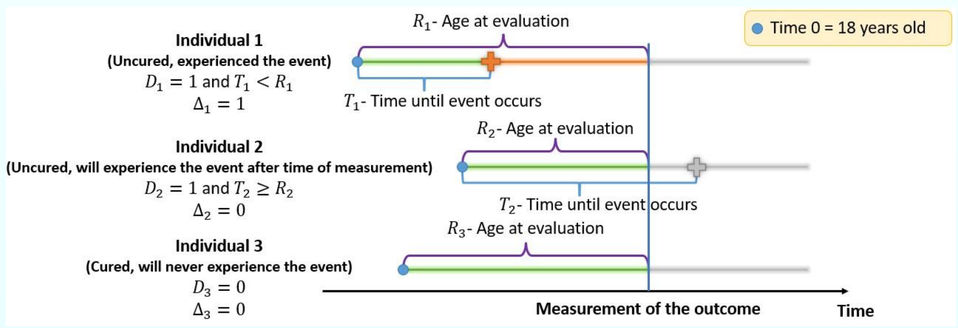
Sahar Ziv ; **Bella Vakulenko-Lagun**

Department of Statistics, University of Haifa, Israel

Motivation

In survival analysis standard assumption is that every individual will eventually experience an event of interest. This assumption fails when a **cured fraction** exists in a population, that is some individuals who will never experience the event regardless of follow-up length and therefore are considered "cured".

When data from such a population are collected cross-sectionally, cured individuals are **indistinguishable** from uncured individuals who have simply not yet experienced the event by the time of sampling, making cured status latent and event times censored.



Naïve analysis methods of such data, that ignore the presence of a cured fraction and censored covariates, can lead to **biased** results and **misleading** conclusions and cannot estimate effects for uncured individuals whose time-to-event is censored.

Our work proposes a new statistical approach to address these problems by developing a **two-stage estimation method**, which first, models the cure probability and time-to-event distribution for uncured individuals. Second, we estimate covariate effects on an ordinal outcome probabilities thorough a mixture likelihood, separately for all groups.

Our Contribution

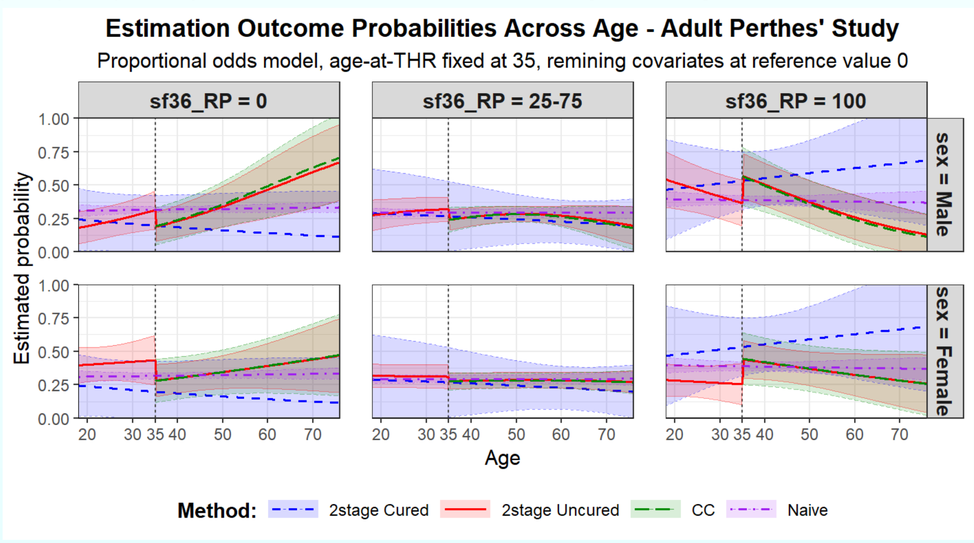
Our two-stage approach provides **consistent** estimation of covariate effects on an ordinal outcome across all subpopulations – cured, uncured with an observed event and uncured with an unobserved time-to-event. Naïve analysis fails to separate cured from uncured individuals and produces biased misleading results.

Moreover, our method provides a solution to the **Positive and Unlabeled (PU) data** problem, with applications in disease diagnosis from Electronic Health Records, presence-only data in ecology, epidemiology and econometrics, and recommendation systems and computer vision tasks in machine learning.

Results of Data Analysis

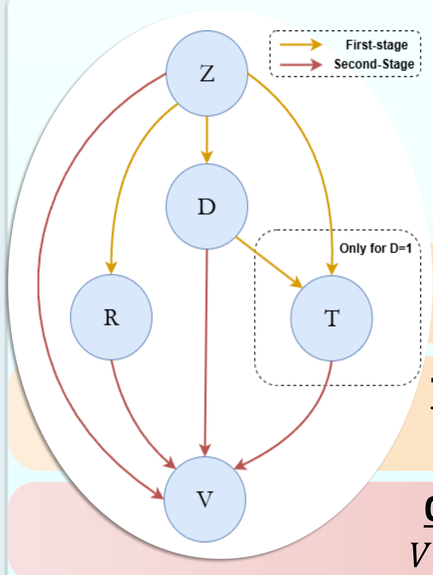
We applied our approach to the adult Perthes' disease study – a web-based survey of $n = 1{,}181$ adults diagnosed with Perthes' disease in childhood.

Our measured ordinal outcome is **SF-36 Role Physical (RP) score**, measuring role limitations due to physical health, collapsed from five possible values (0, 25, 50, 75, 100) to three ordered levels: **0, 25-75, 100**.



- At **THR (age 35)**, probability of **RP=100 jumps** then declines with age.
- **Cured subgroup** follows a distinct trajectory with **wide uncertainty**
- Naïve produces narrow but **incorrect** estimates.
- CC results are **similar** to our approach post-THR results.

Our Proposed Approach



To estimate covariate effects on the time-varying ordinal outcome while distinguishing between cured and uncured subpopulations, we propose a two-stage estimation approach built on four model components:

Cure incidence $D|Z$ → logistic regression model

Time-to-event $T|D=1,Z$ → Weibull PH regression model for uncured individuals

Outcome $V|D,T,R,Z$ → Ordinal outcome regression model

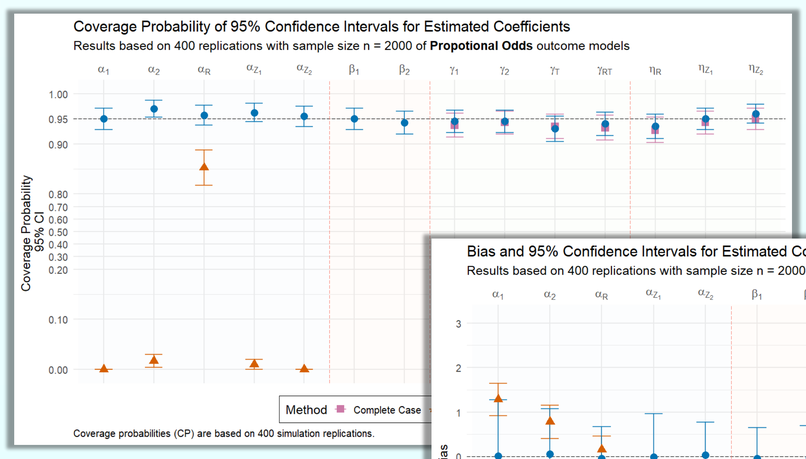
Our key assumption is $\boldsymbol{R \perp\!\!\!\perp (T,D)|Z}$

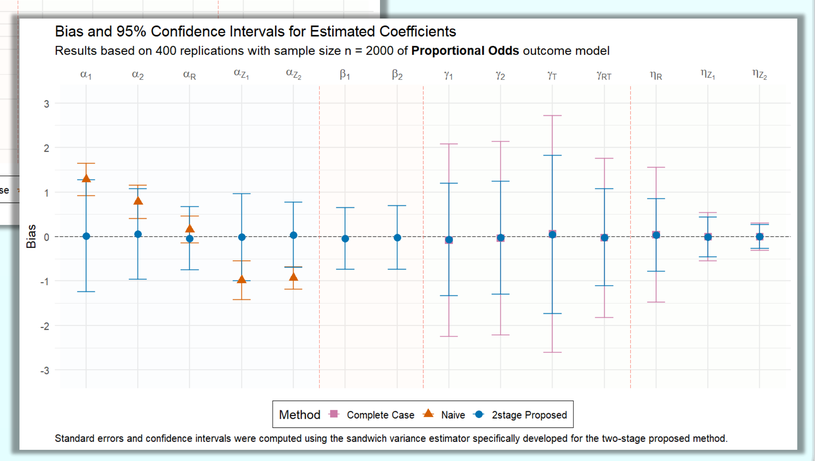
Simulations

The time-varying ordinal outcome V takes **three** ordered categories. Based on 400 replications with sample size $n = 2000$, we specify separate outcome models for each subpopulation:

- **Cured ($D = 0$)**: effects of two baseline covariates (Z_1, Z_2) and age at sampling (R); with intercepts and effects denoted $\boldsymbol{\alpha}$.
- **Uncured with observed event ($D = 1$, $T < R$)**: time-to-event (T) and their interaction $(R \times T)$; with intercepts and effects denoted $\boldsymbol{\gamma}$.
- **Uncured with censored event ($D = 1$, $T > R$)**: only intercepts are denoted by $\boldsymbol{\beta}$.
- **Uncured ($D = 1$)**: **Common effects** of R and two baseline covariates (Z_1, Z_2) before and after the event, denoted by $\boldsymbol{\eta}$.

We compare our proposed two-stage method with both **naïve** and **complete case analysis**, where we fit a proportional odds regressions in two groups defined by $\Delta = 0$ and $\Delta = 1$, respectively.





Conclusions

- **Naïve analysis completely misleading** – large bias for the cured group $\boldsymbol{\alpha}$ and coverage probabilities are **near zero**, far below 95%.
- **Our approach and complete-case are both unbiased** for $\boldsymbol{\gamma}$, but our approach is **more efficient**.
- **No existing method could estimate** the cured group effects $\boldsymbol{\alpha}$ or the uncured group effects $\boldsymbol{\beta}$ **unbiasedly with valid 95% confidence intervals**.
- **Our R library that implements our approach is at** https://github.com/Sahar-Zi/RCuredOrdinal

Acknowledgments

We want to thank M. B. Millis, H. K. W. Kim and the International Perthes' Study Group from Scottish Rite for Children, Dallas, Texas, USA, for launching the adult Perthes' study, and the study participants for their contribution to the research. This research was supported by the Israel Science Foundation (grant No. 1219/23).