Longitudinal And Survival Models In Dementia Research

Presented by Sharon Xie,
Department of Biostatistics and Epidemiology,
University of Pennsylvania Perelman School of Medicine

Joint work with Jarcy Zee, Ph.D.

July 23, 2016

Challenges in longitudinal observational studies

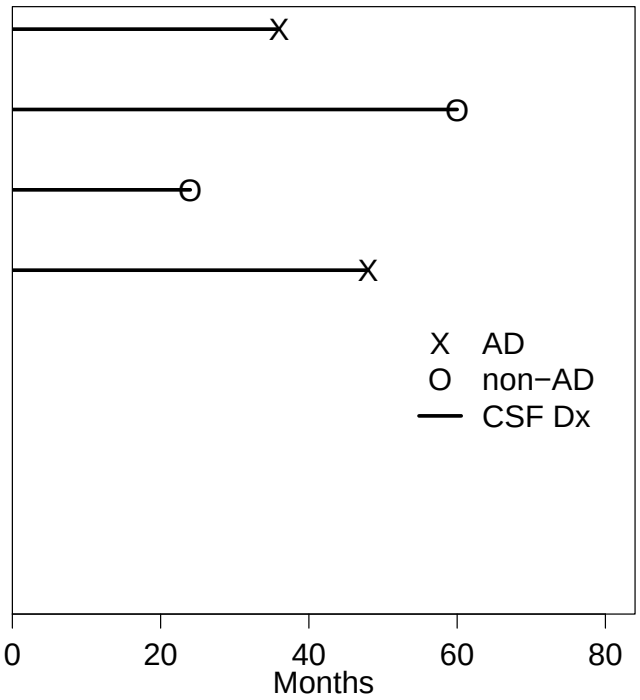
- Unmeasured confounding
- Non-linear changes
- Left-truncation
- Measurement error
- Missing data

ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (ADNI)

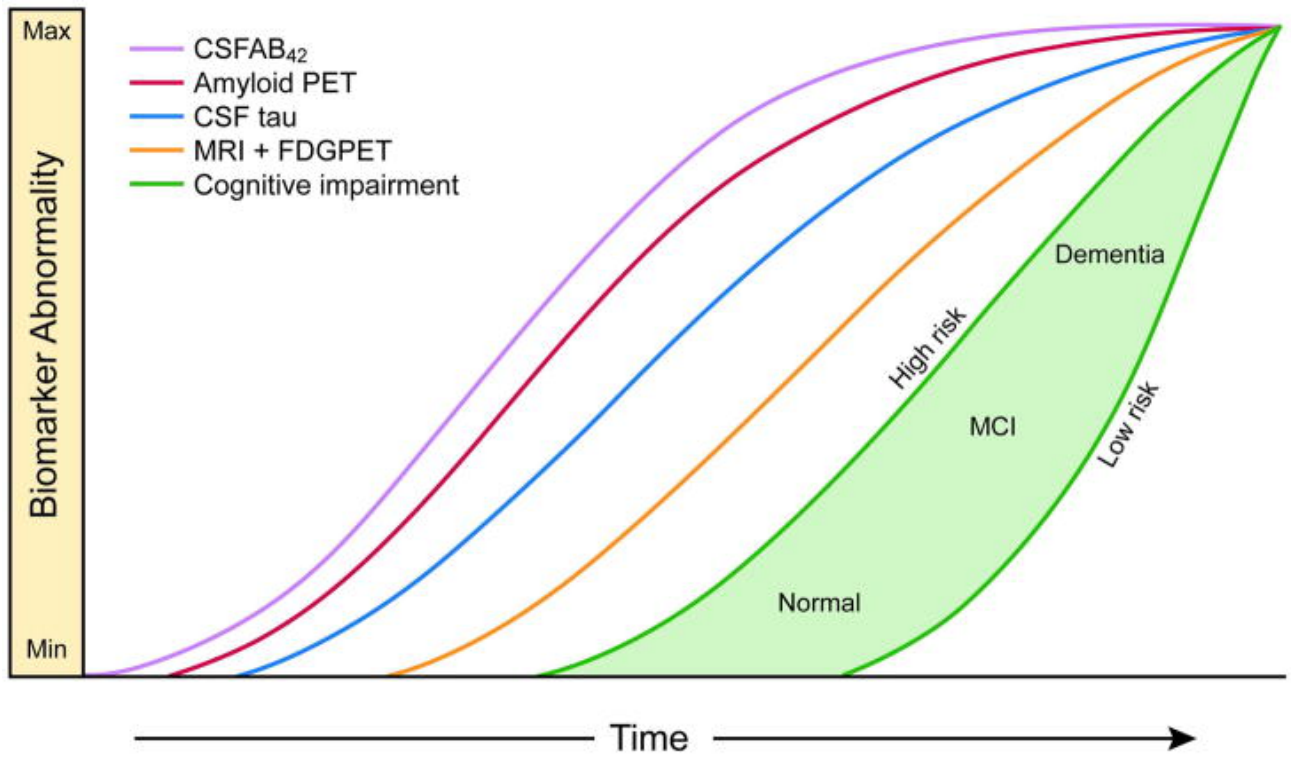
- Multi-center longitudinal study launched in 2003 (PI: Michael Weiner, UC - San Francisco)
- Participants seen every 6 months until end of 2 years, then annually thereafter
- Track longitudinal changes of neuroimaging, cerebral spinal fluid (CSF) biomarkers (Aβ_{1-42} and Tau), as well as clinical and neuropsychological assessment
- Provide a pathway to measure the conversion to AD among normal and mild cognitive impairment participants

ESTIMATION OF CONVERSION RATE FROM NON-AD TO AD

- Goal to study significant pathological changes characterized by abnormal changes of CSF Aβ_{1-42} protein
- Outcome of interest is time to abnormality of CSF Aβ_{1-42} protein
- Predetermined time points of follow-up (years)
- Small subset in ADNI have CSF assay annually



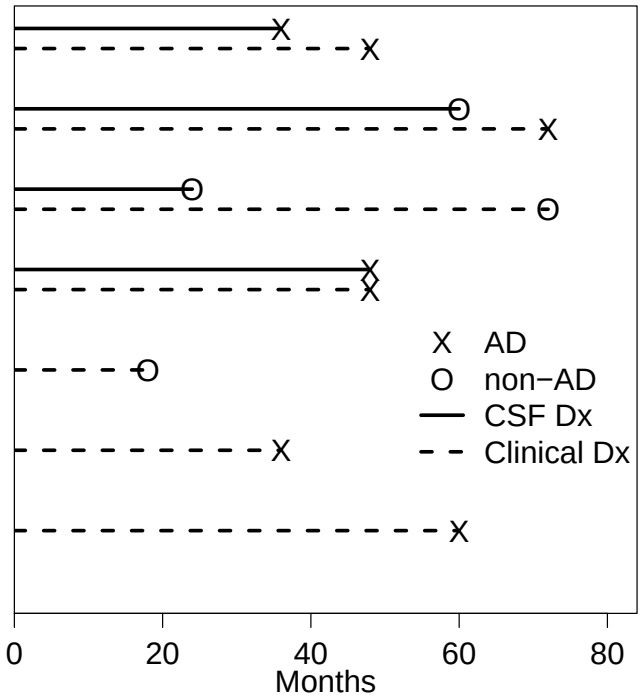
CLINICAL DIAGNOSIS - NOT THE SAME AS THE CSF CLASSIFICATION



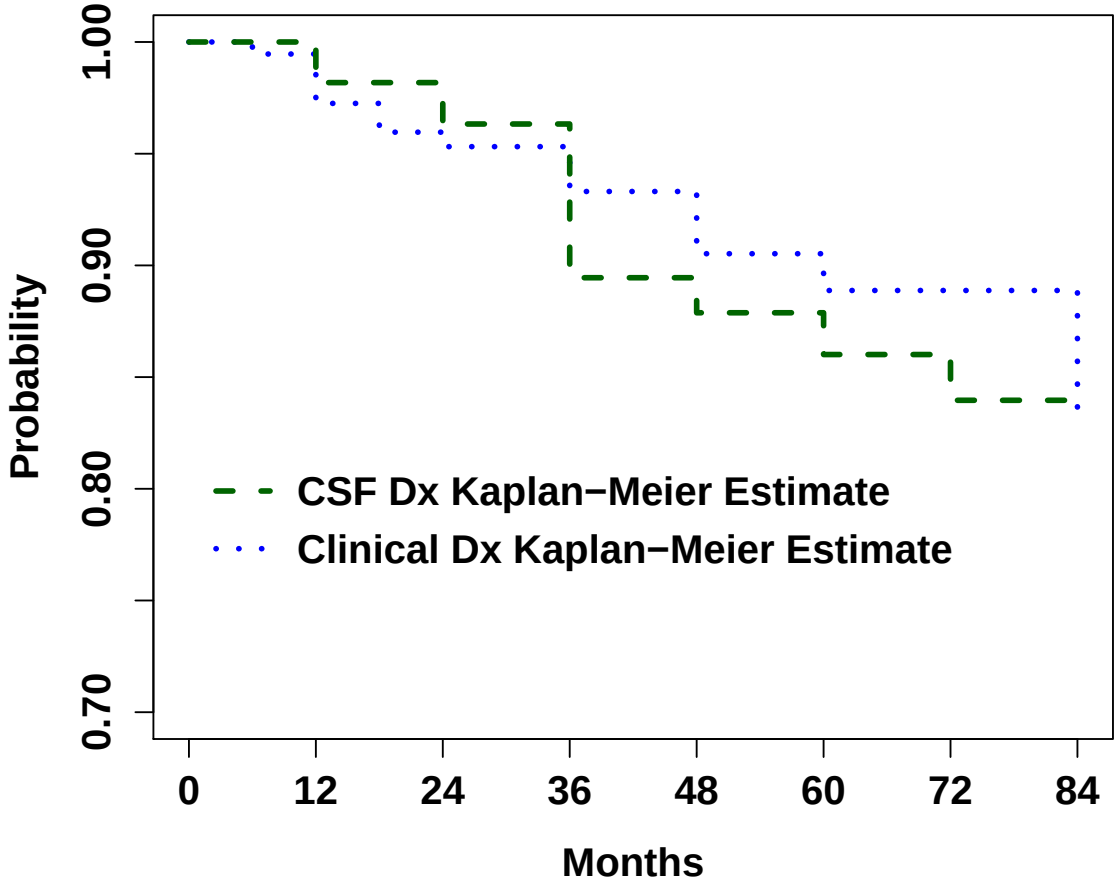
Jack et al., 2013

True outcome and uncertain Outcome

- True outcome: time to CSF abnormality
- Uncertain outcome: time to clinical diagnosis, widely available, based on neuropsychological testing
- Both outcomes can be censored
- Those “non-AD” at baseline by both outcomes are included (*n*=186)



SURVIVAL FUNCTION ESTIMATES



New approach [Zee J, Xie SX. (2015), Biometrics]: nonparametric survival function estimate using both true and uncertain endpoints.

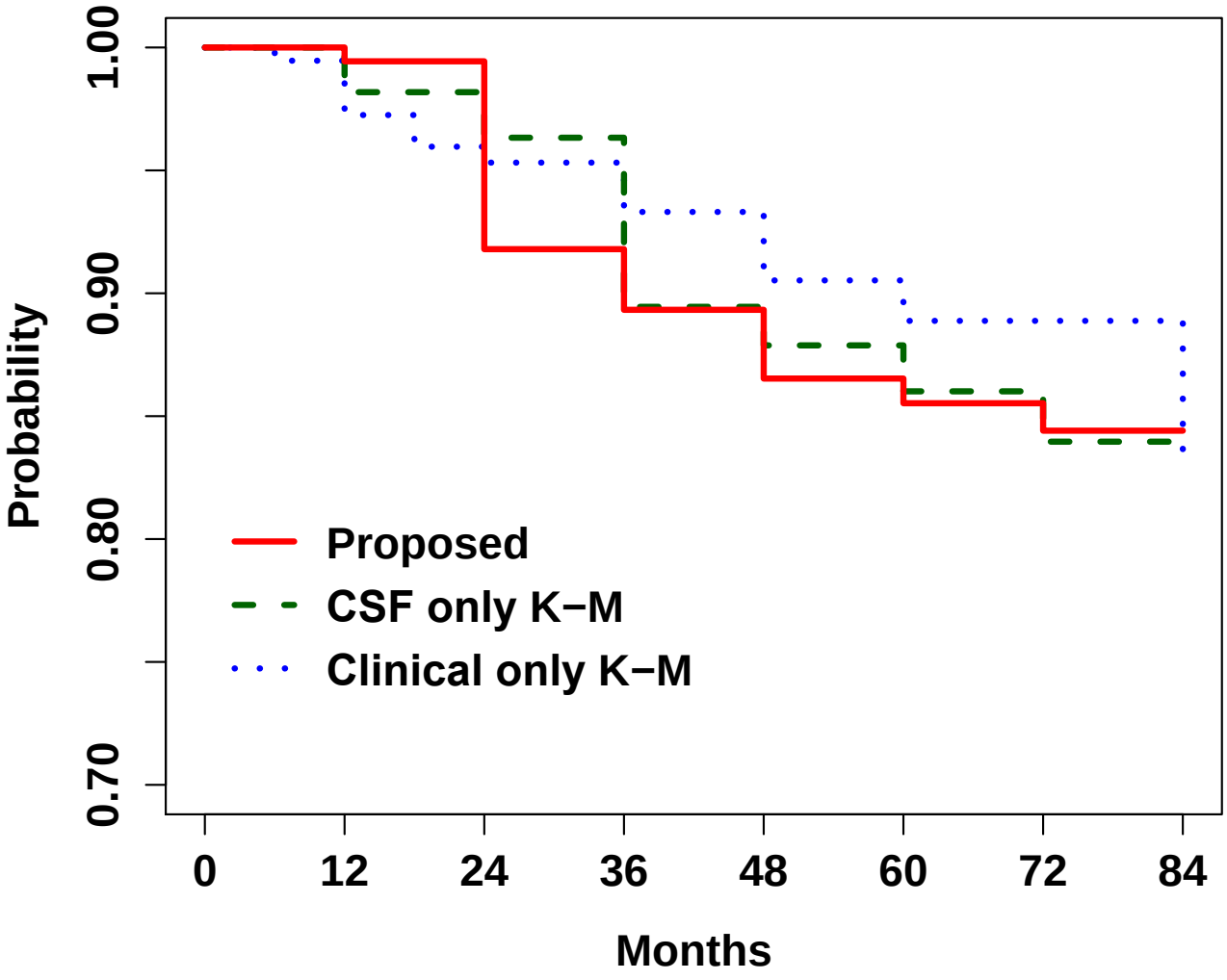
Main Idea

- Two outcomes are correlated
- Extract information from the uncertain outcome to improve the efficiency of the survival estimate of the true outcome
- Using the subset (internal validation sample) where both outcomes are available

Analysis of AD conversion rate in ADNI

- Time to clinical diagnosis of AD (uncertain outcome) available for all n =186 participants
- Time to CSF Aβ_{1-42} abnormality (true outcome) available for subset of participants, n_V=110 (40.8% missing)
- Correlation between true and uncertain outcomes was 0.363

SURVIVAL ESTIMATES



Standard Error Estimates

Month	Proposed Estimator	CSF only Kaplan-Meier	Clinical only Kaplan-Meier
6	0.000	0.000	0.005
12	0.008	0.013	0.012
18	0.008	0.013	0.016
24	0.022	0.019	0.017
36	0.036	0.036	0.023
48	0.038	0.040	0.031
60	0.040	0.045	0.036
72	0.046	0.051	0.036
84	0.046	0.051	0.074

SUMMARY

- The proposed estimator can handle different censoring mechanisms
- The proposed estimator is fully nonparametric, allows real-time validation and allows any participant to be validated
- Using the internal validation subsample can reduce the bias of survival estimates compared to using only uncertain outcomes
- Using uncertain endpoints in the non-validation subsample can improve efficiency compared to using only true outcomes
- Efficiency gains seen with 50% or less missingness

References

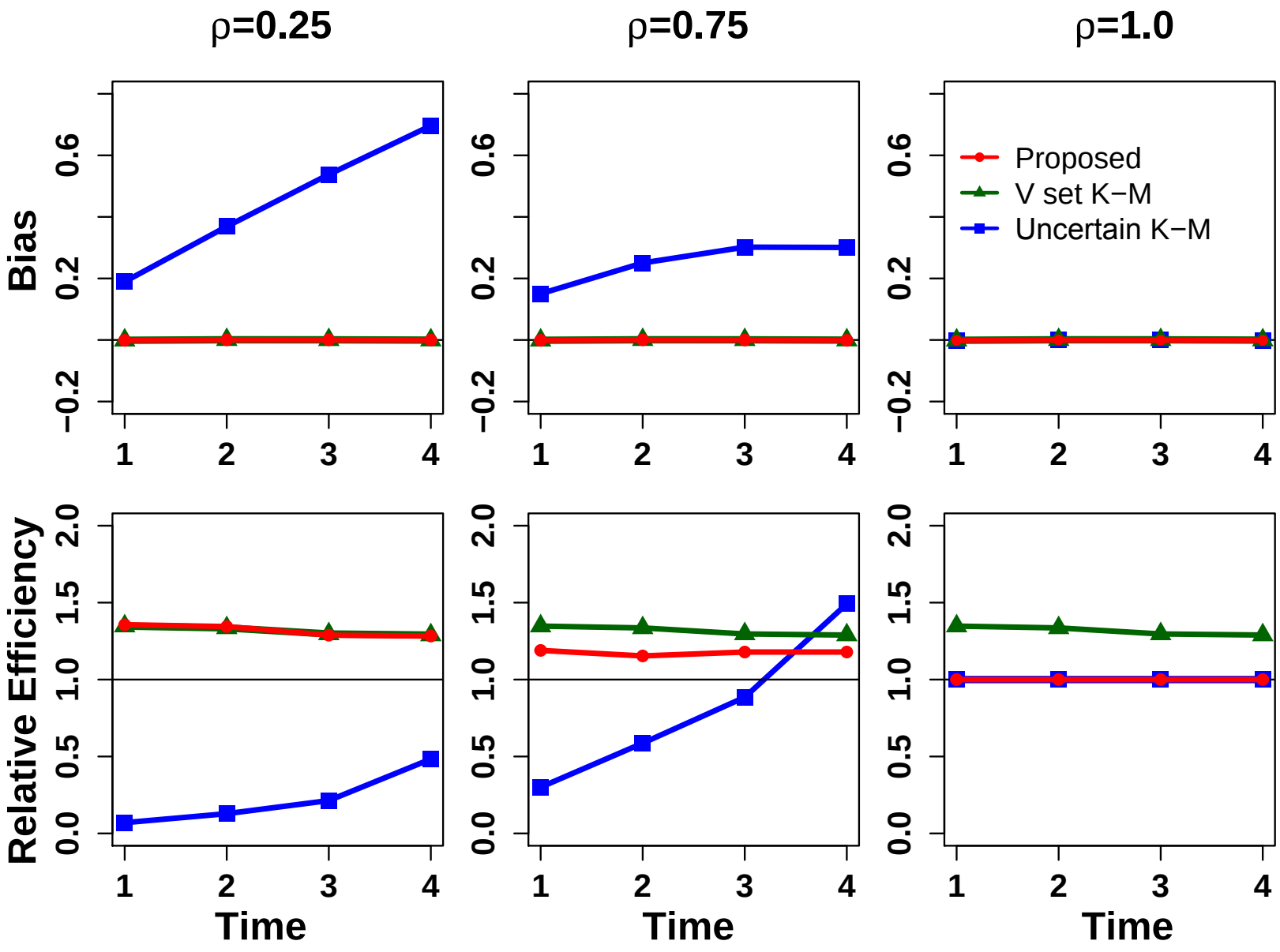
- Zee J, Xie SX. (2015). Nonparametric discrete survival function estimation with uncertain endpoints using an internal validation subsample. *Biometrics* 71(3): 772-81.
- Zee J, Xie SX; for the Alzheimer's Disease Neuroimaging Initiative. (2015). Assessing treatment effects with surrogate survival outcomes using an internal validation subsample. *Clinical Trials* 12(4): 333-41.

Acknowledgments

- Jarcy Zee, Ph.D.
- Alzheimer's Disease Neuroimaging Initiative
- NIH AG-10124 (UPenn Alzheimer's Disease Core Center)
- NIH Mental Health Training Grant: T32MH065218

Thank you!

RESULTS–25% MISSING



Checking Missing Data Assumption

- Missing CSF diagnoses are missing right after baseline
- Data appear to be missing completely at random
  - Log-rank test ($p = 0.662$): comparing distributions of time to clinical dx between validation and non-validation sets
  - Fisher's exact test ($p = 1$): association between missingness and clinical event indicator

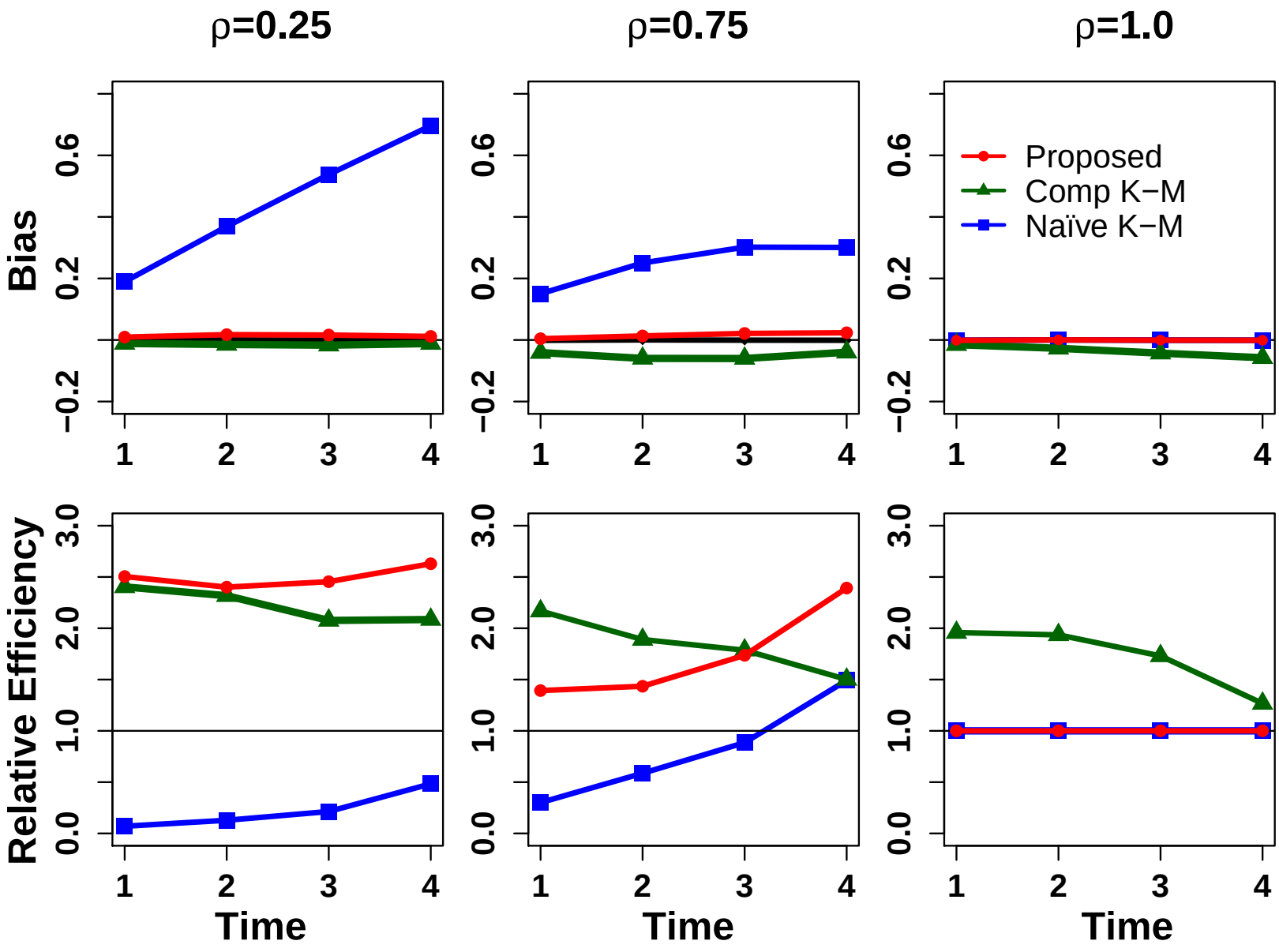
Data Missing at Random

- Those with positive uncertain endpoints may have greater chance for validation
- Simulated MAR data, for missingness indicator R,

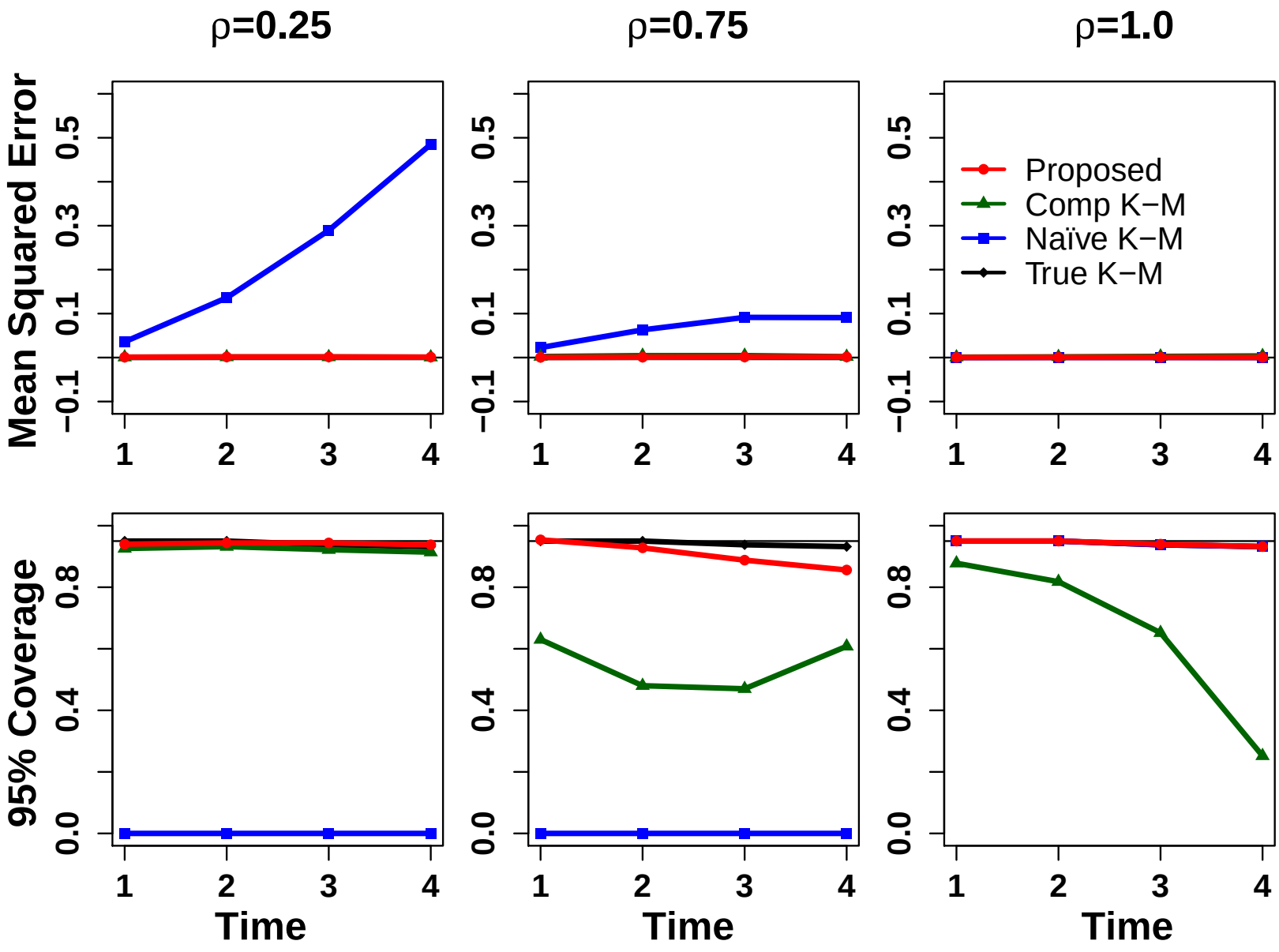
$$R|(\delta^* = 0) = \begin{cases} 1 & \text{with probability } 0.6 \\ 0 & \text{with probability } 0.4 \end{cases}$$

$$R|(\delta^* = 1) = \begin{cases} 1 & \text{with probability } 0.4 \\ 0 & \text{with probability } 0.6 \end{cases}.$$

Results–MAR Bias & RE (Type 1)



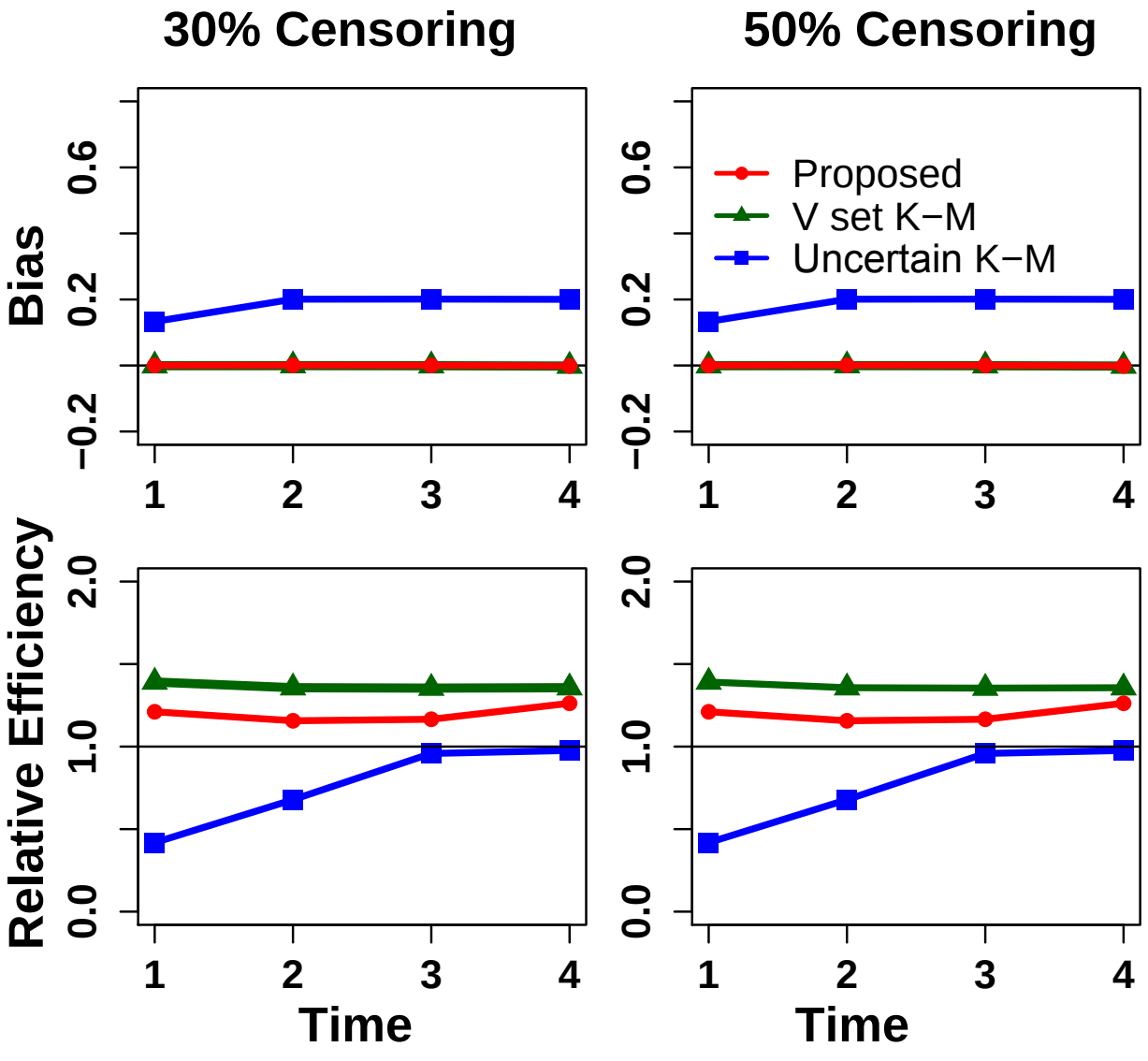
Results–MAR MSE & Coverage (Type 1)



Random Censoring

- Let $T \sim \text{Unif}[1,5]$, $T^* = T + \epsilon$, where $\epsilon \sim \text{Unif}[0,2]$
- Changed type and amount of censoring
  - Let $C \sim \text{Unif}[3,4]$, $C^* = C + \gamma$, where $\gamma \sim \text{Unif}[0,2]$
  - Let $C \sim \text{Unif}[1,4]$, $C^* = C + \gamma$, where $\gamma \sim \text{Unif}[0,2]$

Results–Random Censoring, 25% Missing



Discrete Time vs Continuous Time

True beta for Education in Cox model: -0.74

Summary statistics	Discrete Method	Continuous Method
Bias	0.0059	-0.0031
95% coverage	0.942	0.945

Details: Events measured on a monthly scale. Data discretized by rounding to the nearest year (12 month intervals).