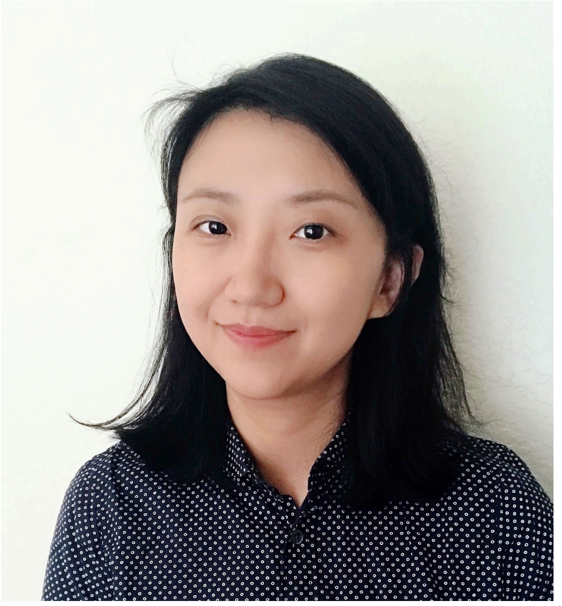
When Will Change?

A Nuanced Understanding of Turning Points through Bayesian Piecewise Growth Models

Ihnwhi Heo, M.Sc.
Quantitative Methods, Measurement, and Statistics
University of California, Merced
https://ihnwhiheo.github.io

On the intricacy of modeling when to change


Big picture question

- Can we accurately recover the true knot locations in longitudinal data?

Publication note

- Heo, I., Jia, F., & Depaoli, S. (in press). Recovering knot placements in Bayesian piecewise growth models with missing data. *Behavior Research Methods*.

Introduction

Studying change in the social and behavioral sciences

- Many research questions focus on how individuals or other units of analysis change over time
- The latent growth (curve) model is a versatile statistical modeling framework:
  - Analyzes intraindividual change over time
  - Analyzes interindividual differences in patterns of change
- Traditional latent growth models assume overall uninterrupted and smooth growth trajectories

Modeling nonlinear change

- It is not uncommon to expect growth trajectories that involve distinct developmental phases
  - How does self-esteem change from childhood to adulthood?
  - How does retention in first grade influence children's math and reading achievement?
- **Piecewise growth models (PGMs)** are useful tools to address such questions

Introduction

Bayesian estimation of piecewise trends

- Improves estimation accuracy by incorporating prior knowledge
- Addresses nonconvergence or inadmissible parameter estimates
- Enhances estimation performance with small sample sizes
- Efficiently manages computational challenges
- Enables modeling of advanced forms of piecewise models (e.g., mixed-effects models, mixture models)

An important parameter: Knot location

- The time point at which transitions occur between developmental stages
- A single knot or multiple knots can be specified a priori based on theory or study design
- More commonly, knot locations are unknown and are therefore freely estimated

Introduction

Bayesian estimation of knot locations

- In Bayesian inference, information from priors is combined with the data likelihood to form posteriors
- Different prior specifications can influence study results
- Depending on the choice of priors for knot locations, PGMs may yield varied outcomes
- **So, how do these prior settings influence the recovery of knots?**

Missing data in longitudinal research

- Missing data are a ubiquitous challenge in longitudinal research
- The presence of missing data can bias parameter estimates and harm generalizability
- Attrition is one of the most common sources of missing data
- **Bayesian estimation via data augmentation can handle missingness — but will it work for PGMs?**

Introduction

Novel contributions

- Compare prior specifications to evaluate sensitivity in estimating knot locations
- Provide the first investigation of how missing data affect knot recovery
- Isolate the impact of missing data in a simplified latent growth curve framework

Research goals

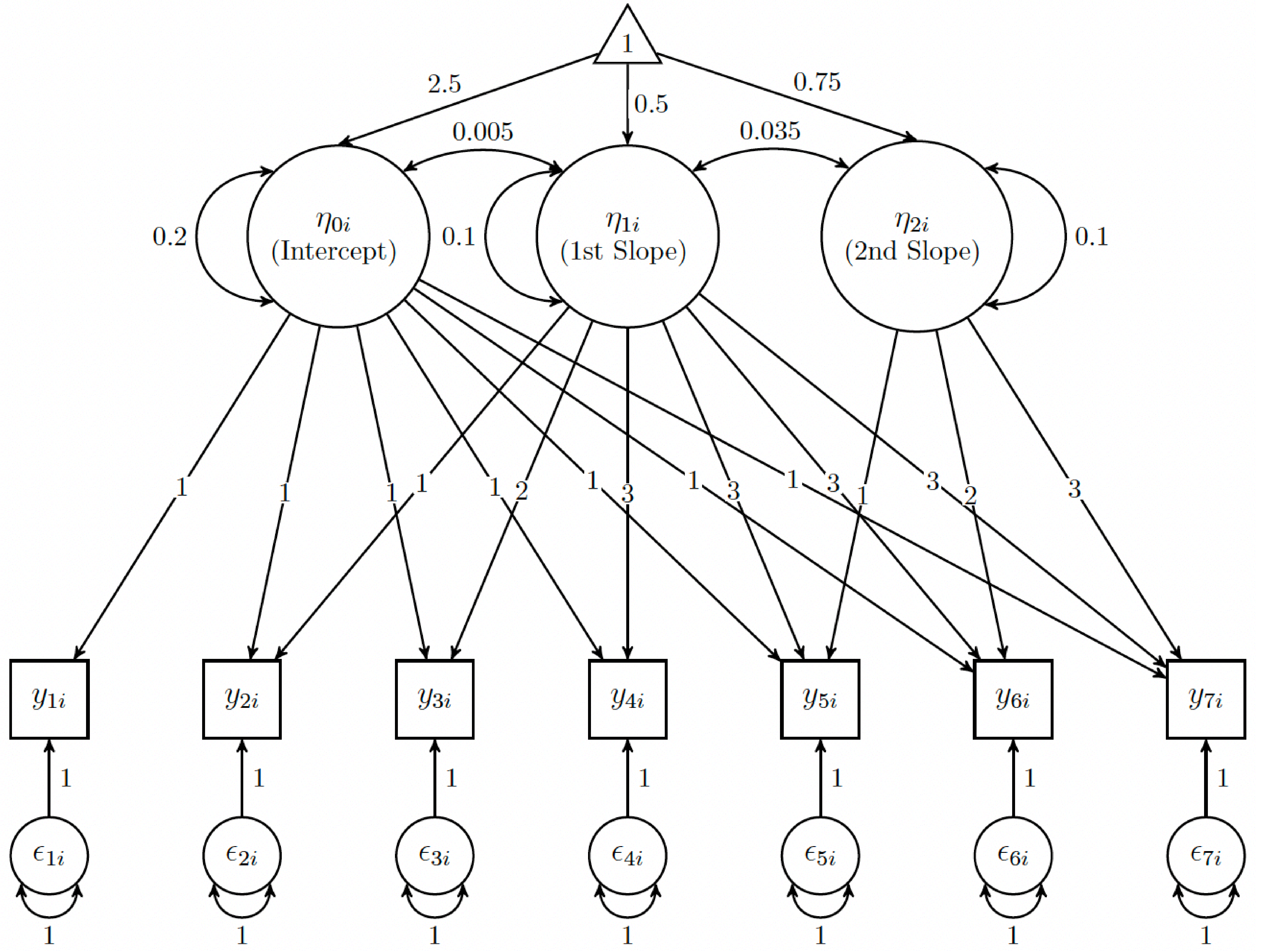
1. Examine the impact of different prior specifications on the recovery of knot locations
2. Examine the impact of missing data on the recovery of knot locations
3. Demonstrate practical applications through illustrative examples using real data

Methods: Simulation

Design: Population model

Figure 1

A Path Diagram For The Population Model.

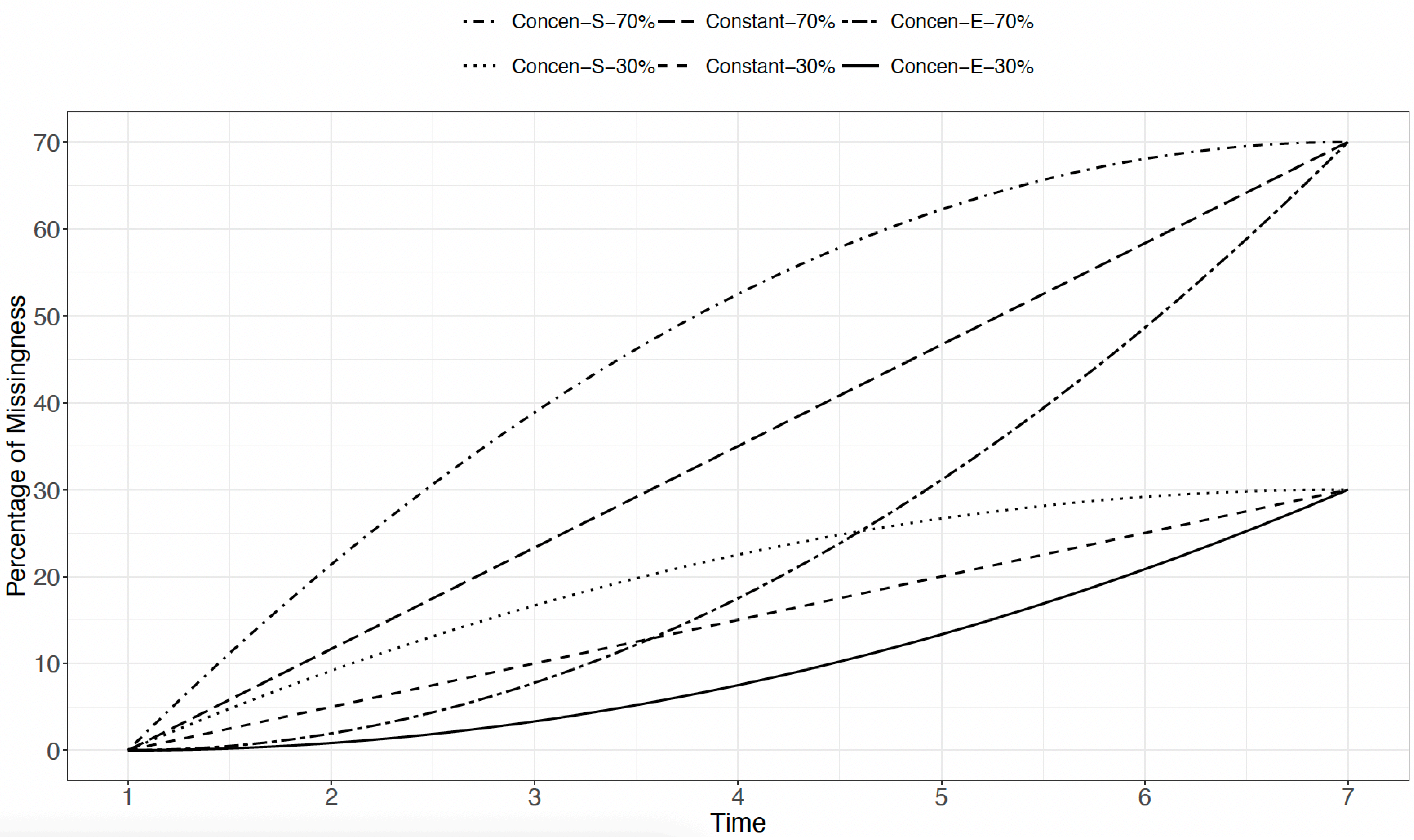


Methods: Simulation

Design: Missing data

Figure 3

Missing Data Patterns.

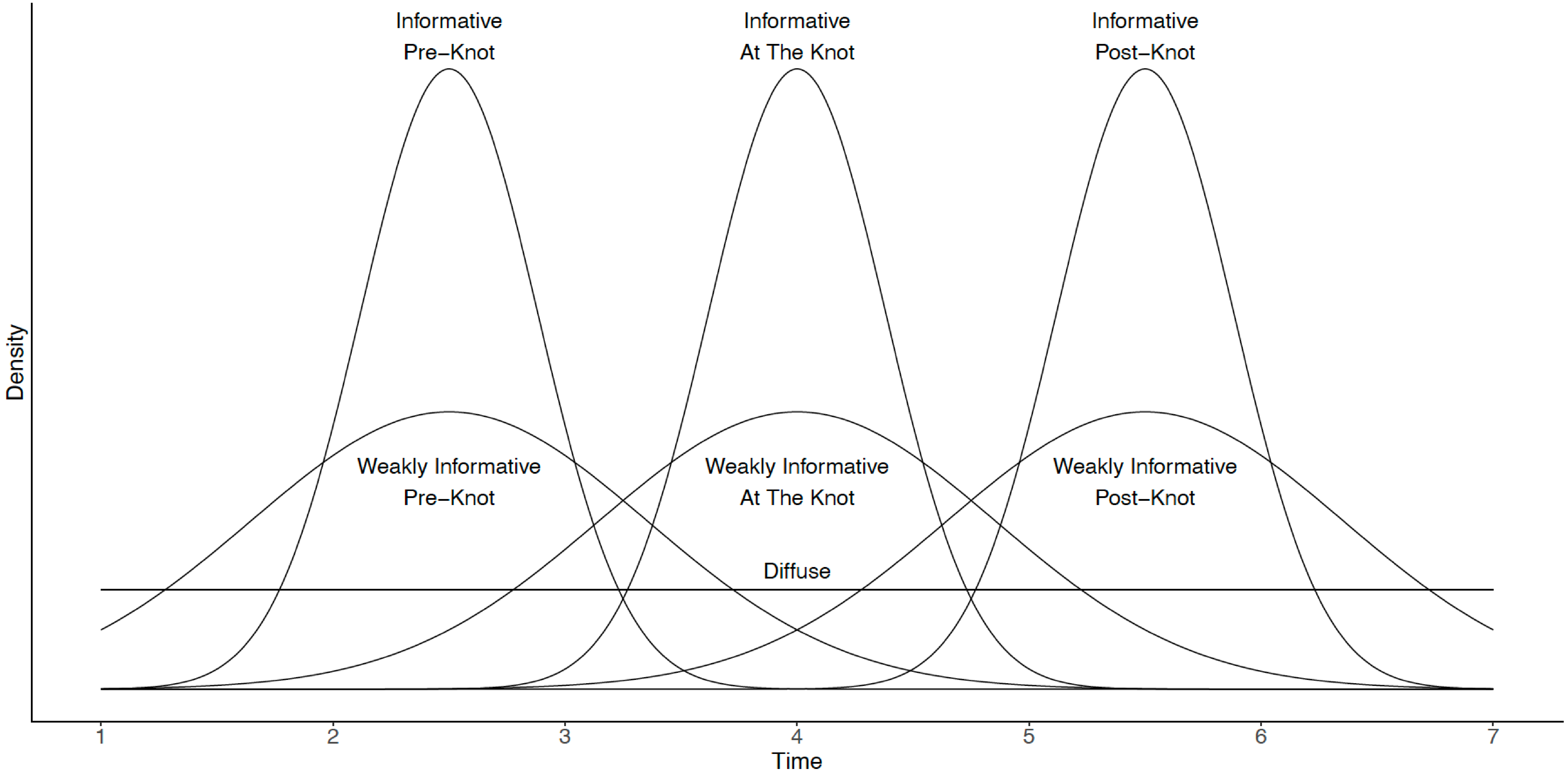


Methods: Simulation

Design: Prior specification

Figure 2

Prior Specifications.



Methods: Simulation

Analytic details

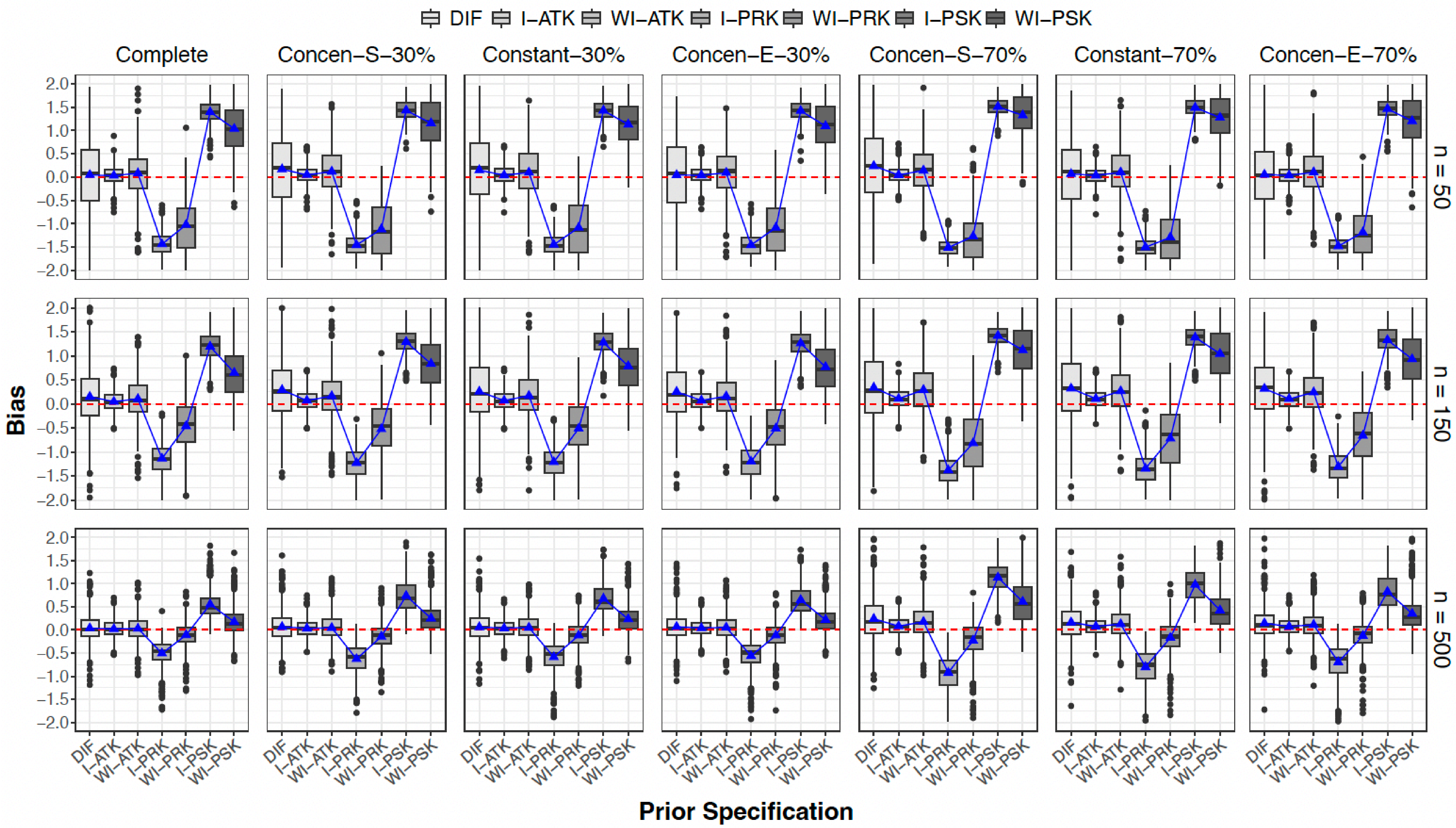
- 147 simulation cells
  - 3 sample sizes: 50, 150, 500
  - 7 missing data conditions
  - 7 prior distribution conditions
  - 500 replications per cell
- Bayesian estimation using R and JAGS
  - Gibbs sampler with 4 chains, each consisting of 25,000 iterations (first 5,000 discarded as burn-in)
  - Thinned by retaining every 10th sample
  - Convergence assessed using $\widehat{\mathrm{R}} < 1.1$
- Outcome measures: convergence, coverage, average bias, RMSE

Results: Simulation

Boxplots of average bias

Figure 4

Boxplots of Bias for Knot Location Estimates Across Simulation Conditions.

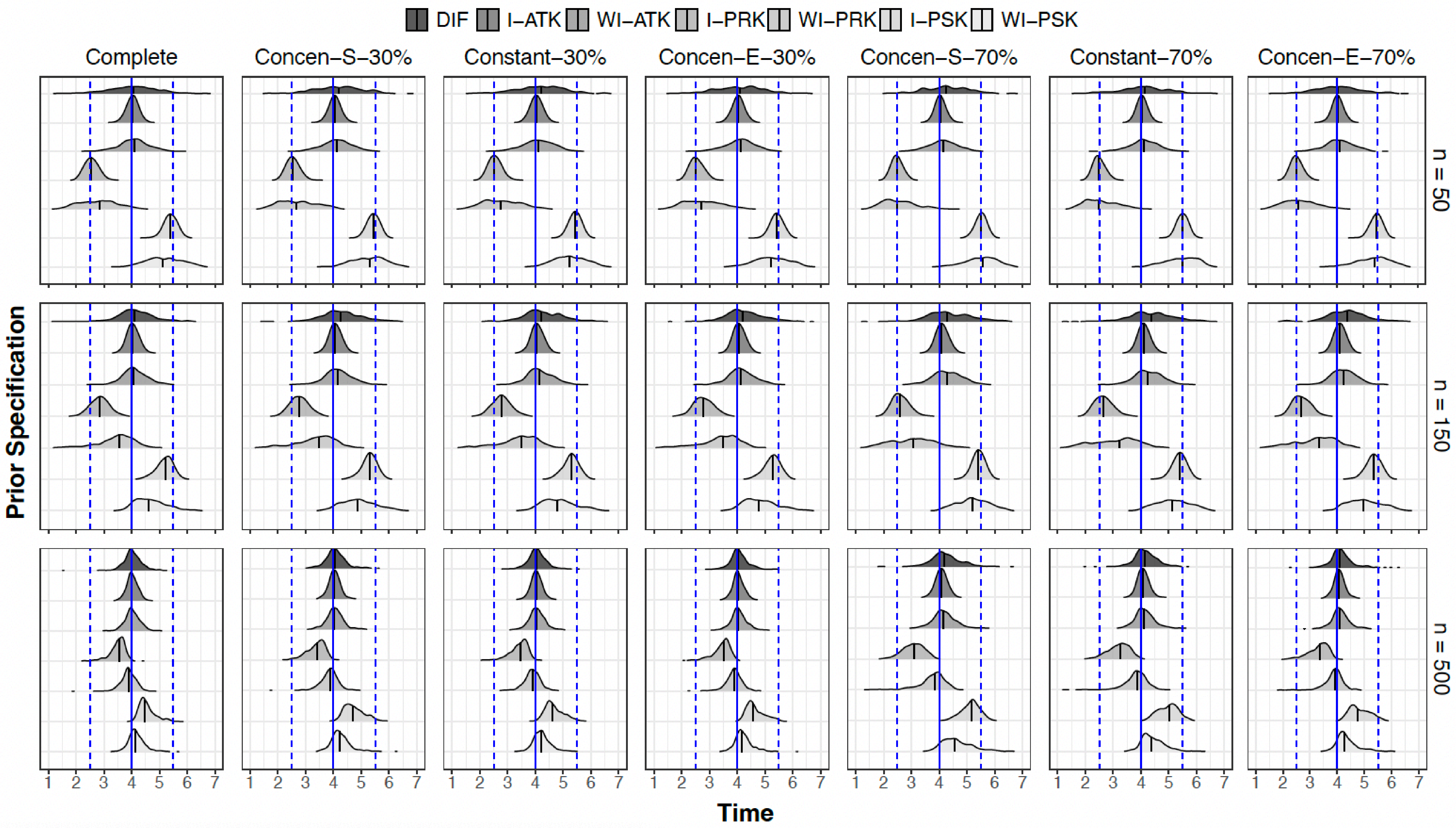


Results: Simulation

Distributions of posterior means for knot location

Figure 5

Distributions of Posterior Means for Knot Location Across Simulation Conditions.



Methods: Illustrative Example

- Goal: Evaluate changepoint estimation under realistic missing data and different prior settings

Data: ECLS-K

- Nationally representative sample of ~21,000 U.S. kindergarteners (1998–1999 cohort)
- Outcome: Math IRT scores measured at 7 time points (coded as 0, 0.5, 1, 1.5, 3.5, 5.5, 8.5)
- Subsample of 150 children to represent a medium sample size
- Attrition pattern followed the Concen-S-70% condition

Analytic Details

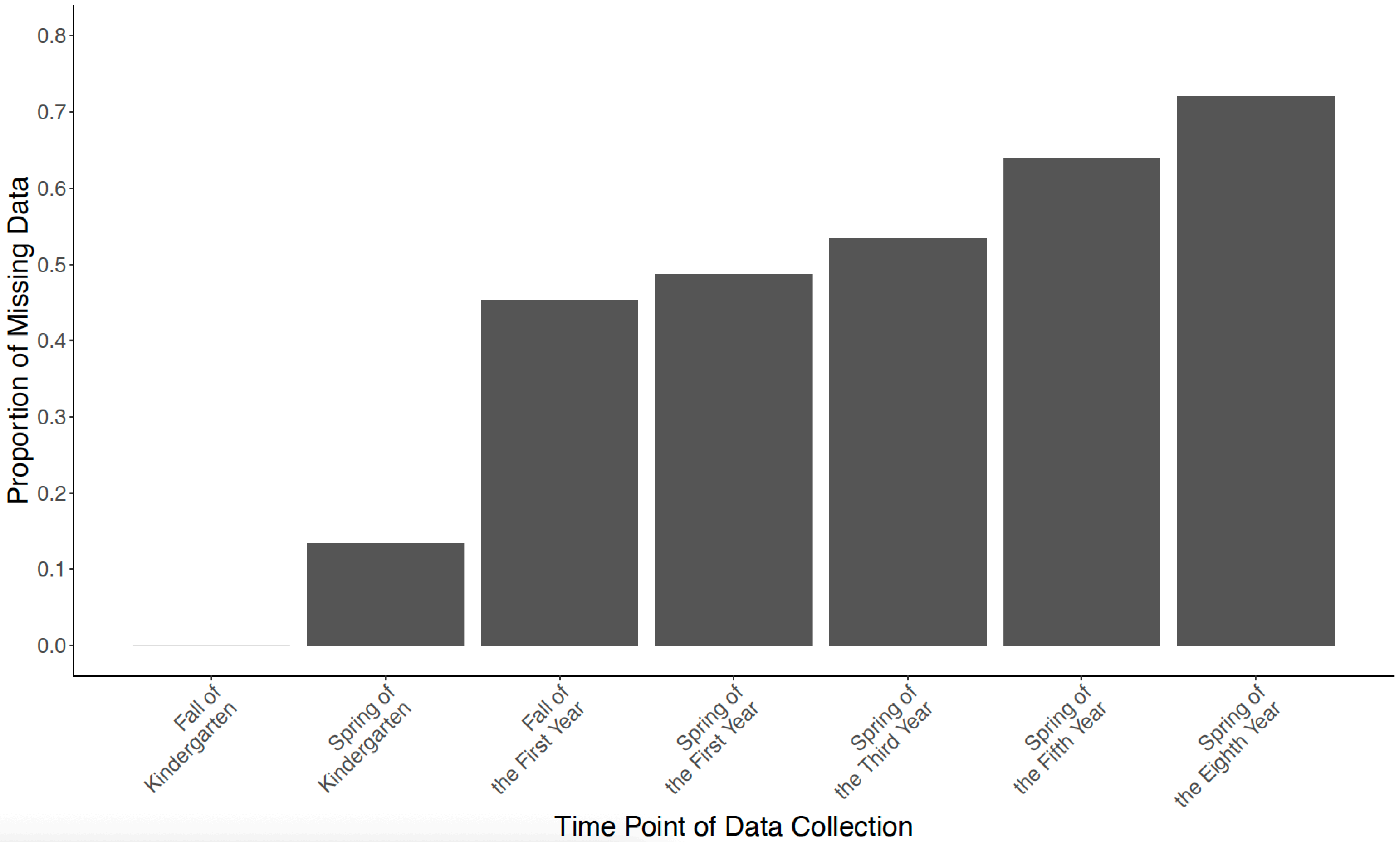
- Bayesian estimation under 7 prior distributions (same as in the simulation)
- Implemented in R and JAGS using 4 chains via Gibbs sampling
- 50,000 burn-in iterations and 50,000 additional iterations for posterior sampling
- Thinning interval of 1

Methods: Illustrative Example

Missing data pattern in the ECLS-K dataset

Figure 6

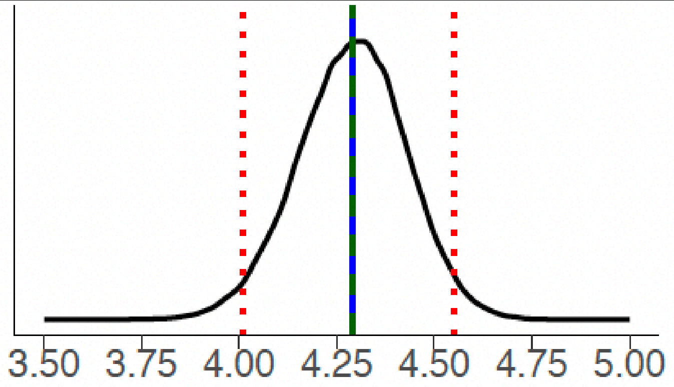
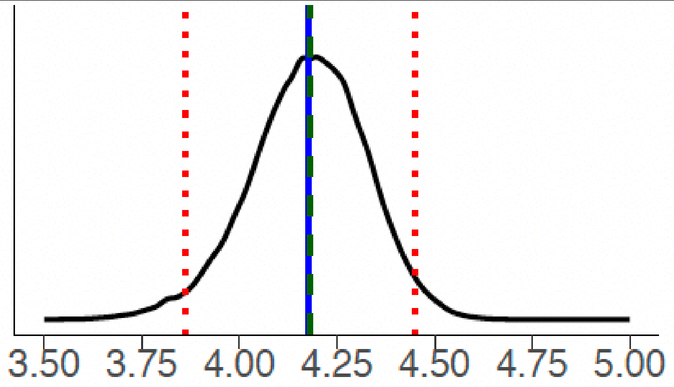
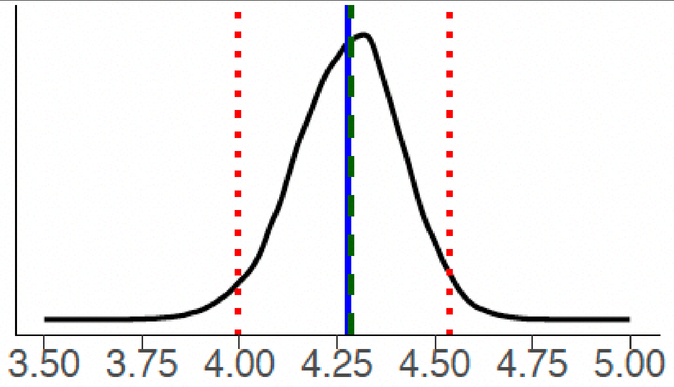
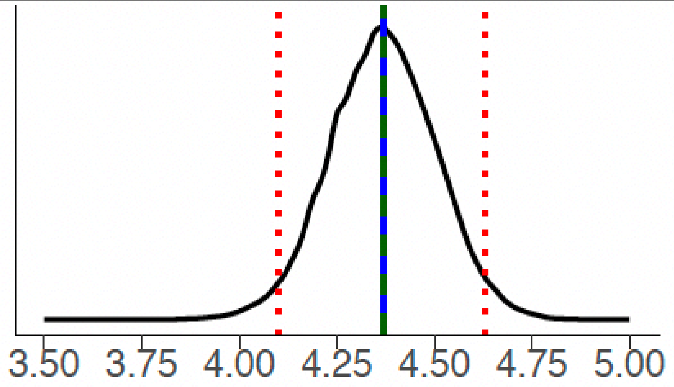
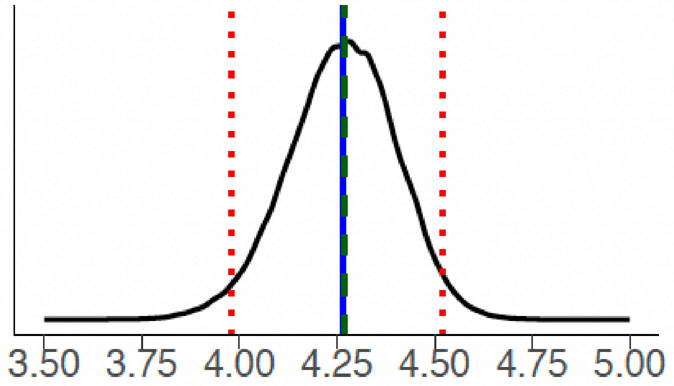
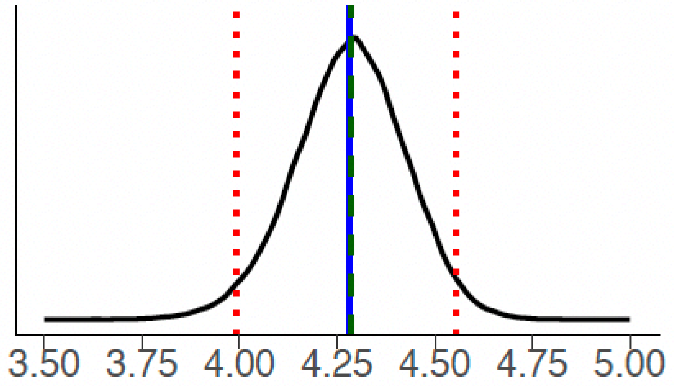
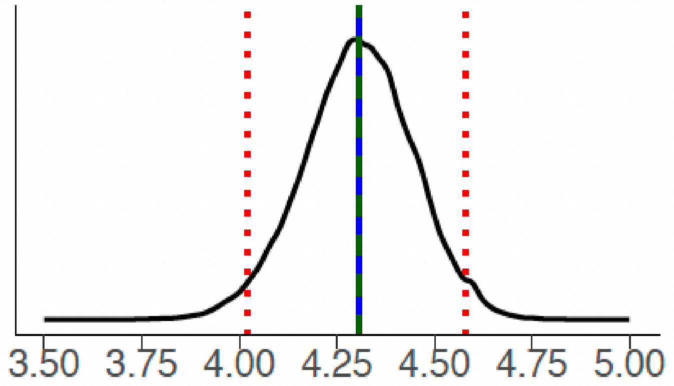
Attrition Pattern for the ECLS-K Data.



Results: Illustrative example

Table 4

Summary of Posterior Distributions of Knot Locations from ECLS-K Data.

	DIF	I-PRK	I-ATK	I-PSK
				
Mean	4.288	4.176	4.279	4.368
Median	4.292	4.182	4.284	4.368
SD	0.138	0.148	0.138	0.135
95% CI	[4.010, 4.550]	[3.861, 4.449]	[3.995, 4.538]	[4.102, 4.631]
		WI-PRK	WI-ATK	WI-PSK
				
Mean		4.263	4.282	4.305
Median		4.268	4.286	4.307
SD		0.138	0.142	0.140
95% CI		[3.980, 4.520]	[3.991, 4.552]	[4.021, 4.577]

Note. DIF = diffuse prior. I-ATK = informative prior at the true knot location. WI-ATK = weakly informative prior at the true knot location. I-PRK = informative prior at the pre-knot location. WI-PRK = weakly informative prior at the pre-knot location. I-PSK = informative prior at the post-knot location. WI-PSK = weakly informative prior at the post-knot location. SD = standard deviation. CI = credible interval.

Discussion

Advice for Researchers

Which priors should we choose?

- Avoid relying solely on diffuse priors
  → greater posterior uncertainty, wider credible intervals, and increased sensitivity in small samples
- Be cautious with overly informative priors
  → if misspecified, can introduce bias, even with large samples; worse under missing data
- Always conduct prior sensitivity checks
  → assess whether results fluctuate across prior settings and revisit substantive theories on changepoints

How can we prevent missing data?

- Take proactive steps to minimize missing data and participant attrition
- Use planned missing data designs
  → reduce participant burden and lower attrition rates
- Improve efficiency by balancing the number of repeated measures and the overall sample size
- Reduce cost while increasing statistical power by recruiting more participants with fewer time points

Epilogue

Thank you. Any questions are welcome!

"Theoretical satisfaction and practical implementation are the twin ideals of coherent statistics." - Dennis Lindley

Contact

- Email: ihnwhi.heo@gmail.com
- Website: https://ihnwhiheo.github.io
- X: IhnwhiHeo
- GitHub: IhnwhiHeo