Evaluating missingness assumptions for items in a frailty index

Louisa Smith

SER: June 14, 2023

Department of Health Sciences
The Roux Institute
Northeastern University

l.smith@northeastern.edu

Context: deficit-accumulation frailty index

Frailty is a syndrome of vulnerability more common in older adults

A frailty index is a quantitative measure of the aggregate burden of age-related health deficits

FI = # of deficits / # of possible deficits

<0.15 robust; 0.15-0.25 pre-frail; >0.25 frail

All of Us RESEARCH PROGRAM

- Large-scale NIH study to gather health data from 1 million+ Americans
- Focus on those underrepresented in biomedical research
- Multimodal data collection includes surveys, electronic health records, biospecimens, and more

AoU-FI

- 33 deficits based on items from multiple surveys
- Cover multiple domains, including comorbidities, function, cognition, mental health, and geriatric syndromes
- Cannot be weighted to heavily toward one domain (or it would be, e.g., a comorbidity index)

9.8% of 200,000+ participants had complete data

38% had data for >80% of deficits (>27/33)

Wong et al. (2023)

Options for missing items in an index/scale

Complete-case
Exclude those with any missing items

Proration
Adjust denominator (person-mean imputation)

Multiple imputation
Of individual items / total score

Options for missing items in an index/scale

Complete-case
Exclude those with any missing items

Throwing away *a lot* of data, strong assumptions

Proration
Adjust denominator (person-mean imputation)

Different weighting across domains

Multiple imputation
Of individual items / total score

Computationally intensive, not valid in general under MNAR

Missing data assumptions

MCAR	MAR	MNAR
missing completely at random	missing at random	missing not at random
missingness does not depend on the observed or missing data	conditional on the observed data, missingness does not depend on the missing data	missingness depends on the missing data, even conditional on the observed data

Understanding assumptions

Let's consider the joint distribution of the missing and observed data and the missingness pattern, all conditional on the fully observed variables:

$$f(Y_{mis}, Y_{obs}, R \mid X)$$

In our setting, Y_{mis} and Y_{obs} are missing and observed components of the frailty variables, R is a matrix with indicators of missingness for each observation/variable, and X are fully observed variables like age and gender

Factorization

$$f(Y_{obs}, Y_{mis}, R \mid X) =$$

$$f(Y_{mis} \mid Y_{obs}, R, X) f(Y_{obs} \mid R, X) f(R \mid X)$$

distribution of missing data conditional on whats observed, and on missingness patterns

distribution of observed data conditional on a given missingness pattern

probability of a given missingness pattern

Under **MAR**, $f(Y_{mis} \mid Y_{obs}, R, X) = f(Y_{mis} \mid Y_{obs}, X)$: missing data doesn't depend on missingness pattern

Pattern-mixture models

Model how the distribution of missing data depends on missingness pattern

- For example, a missingness pattern in which a given variable Y is missing may be associated with a *higher* probability that $Y = 1$
- We would never know that from the observed data, because by definition we are missing Y in that missingness pattern
- But in a sensitivity analysis, we can decide how much higher it might be

Little (1993); Rubin (1987)

Sensitivity analysis via delta adjustment

For a single variable with missingness:

$$E[Y \mid R, X] = \beta_0 + \beta_1 X + \delta I(R = r_0)$$

where δ parameterizes how much different the distribution of Y is in observations with missing data patterns where it is missing (r_0)

Multiple imputation

The delta adjustment approach can be done in the context of multiple imputation, e.g., with MICE

- Fit a model for Y as usual
- Add δ to the imputed values
- Analyze multiple datasets as usual

Van Buuren and Groothuis-Oudshoorn (2011); Buuren (2012)

Complications

With multiple missing variables, interpretation of sensitivity parameter δ is different

- conditional on the missingness pattern of the other variables
- D. M. Tompsett et al. (2018) proposed a solution which involves eliciting mode interpretable delta-like parameters and searching the solution space for the δs they correspond to
- computationally infeasible with 33 missing items without further assumptions

Missingness patterns

For a given item $\mathbf{Y}$, we collapsed missingness patterns into:

- data on $\mathbf{Y}$ and all surveys completed (group A)
- data on $\mathbf{Y}$ but missing some surveys (group B)
- missing data on $\mathbf{Y}$ but completed survey (group C)
- missing survey on which $\mathbf{Y}$ is collected (group D)[1]

1. so not known whether it *would* have been observed had survey been completed

Interpretable parameters

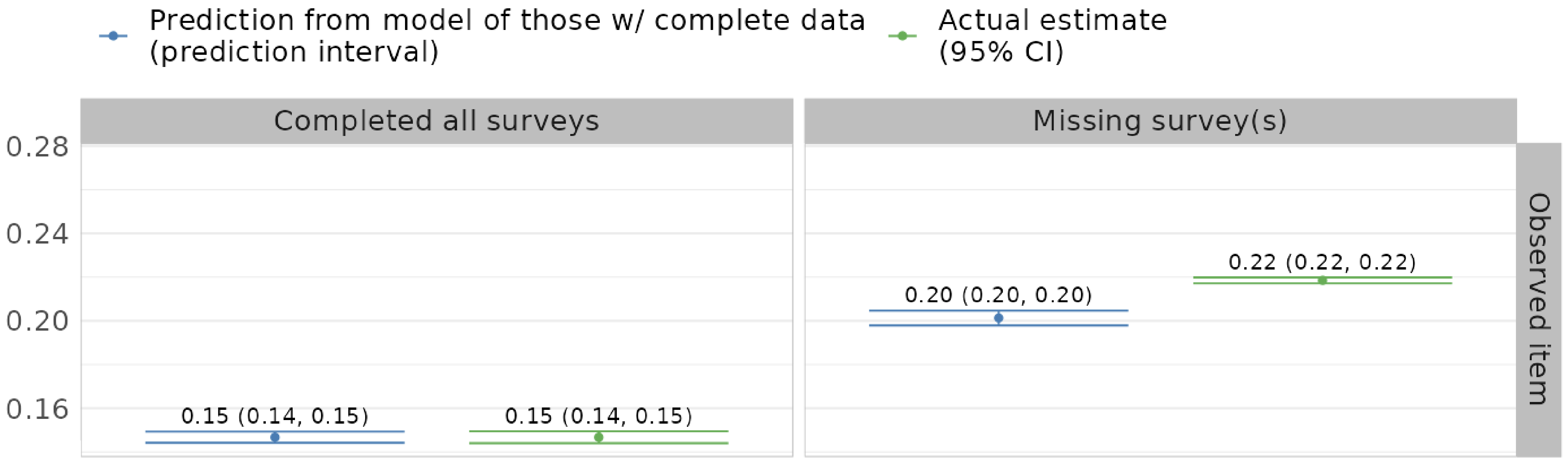
Most items are binary

- Parameters on odds ratio scale suggested in literature
  - "Non-respondents may have up to 1.3 times the odds of *item* compared to respondents who are similar in other ways"
- Even differences in means not particularly intuitive
  - "Non-respondents may have up to 10 percentage points higher prevalence of *item* compared to respondents who are similar in other ways"

Standardized means seem more interpretable

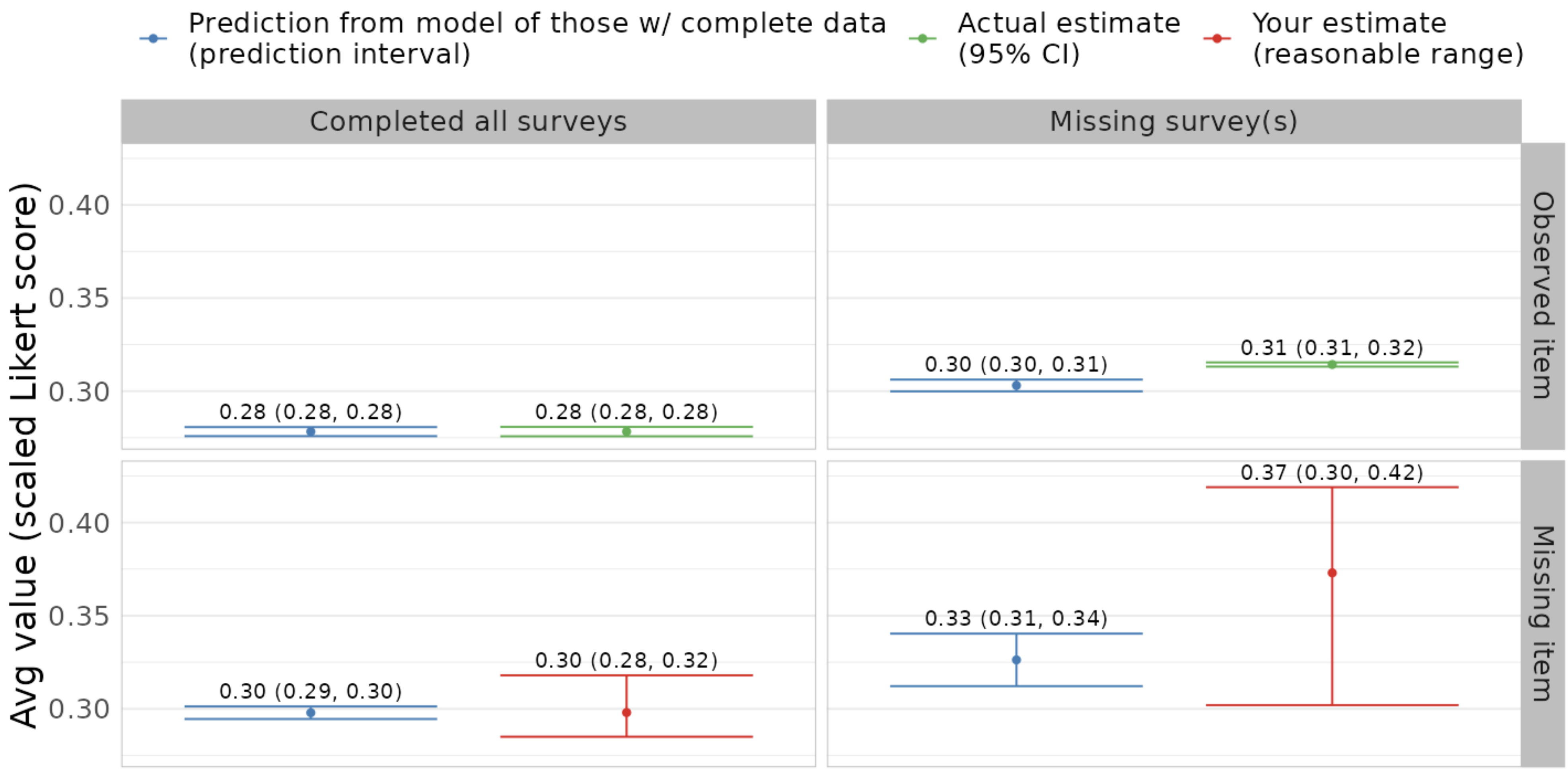
Standardized means

- Fit a model for item among participants with complete data (group A), conditional on demographics, etc.
- Predict item prevalence among participants with other missing surveys, but complete item of interest (group B)
- Compare observed and predicted item prevalence in group B: differences are not accounted by demographics, instead by missing data pattern



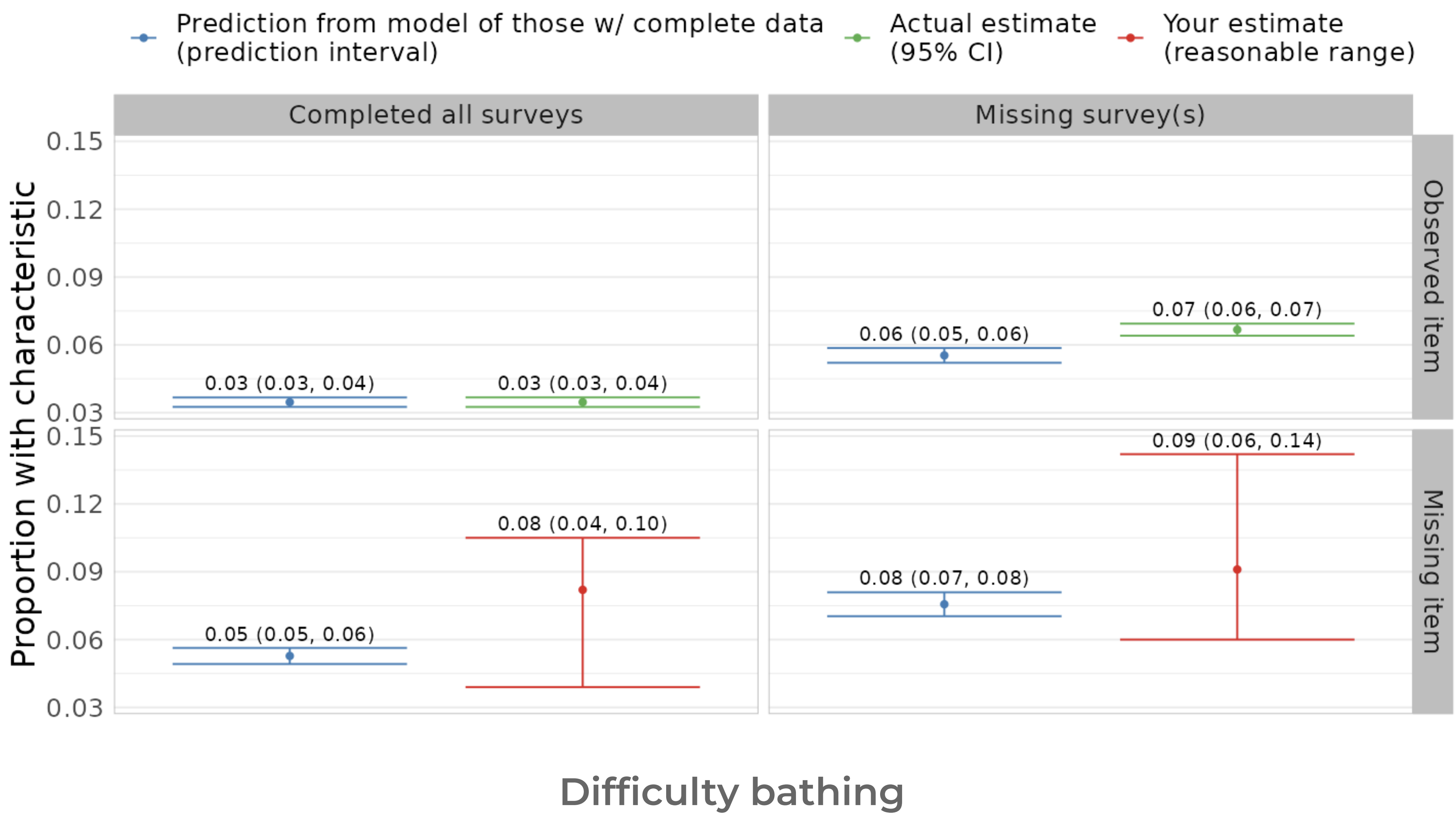
Difficulty with everyday activities

This comparison makes specifying the sensitivity parameters more concrete



Severe fatigue

Experts in this population can combine with their knowledge

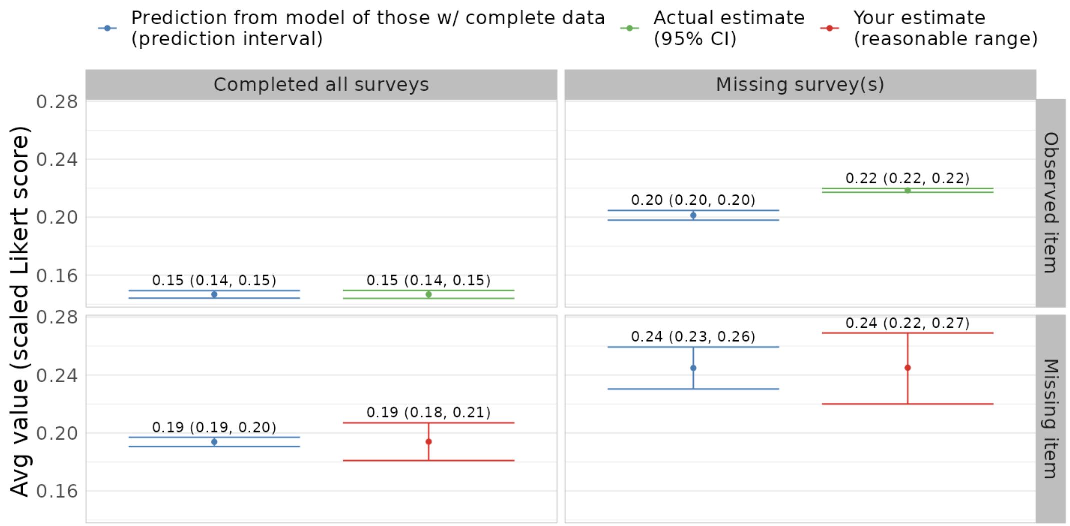


Shiny app

Sensitivity values for missing items in AoU frailty index

Progress 3%

Can't complete everyday activities



Your estimated range for participants skipping this question:

Range: 0.181 — 0.207

Best guess: 0.194

Your estimated range for participants missing entire surveys:

Range: 0.22 — 0.269

Best guess: 0.245

Next

Table | Help!

Descriptive statistics for those with complete data and observed 'can't complete everyday activities' item

Characteristic	**N = 29,054**
Age	
50-59	8,539 (29%)
60-69	10,891 (37%)
70-79	8,112 (28%)
80+	1,512 (5.2%)
Gender	
Female	16,564 (58%)
Male	12,109 (42%)
Other	79 (0.3%)
Missing	302
Race/ethnicity	
Non-Hispanic White	22,539 (80%)
Non-Hispanic Black	2,413 (8.6%)
Hispanic/Latino	1,935 (6.9%)
Asian	572 (2.0%)
Other	327 (1.2%)
Mixed	319 (1.1%)
Missing	949
Income	
< 50k	7,372 (29%)
50-100k	4,364 (17%)
> 100k	13,409 (53%)
Missing	3,909
Education	
Less than high school	898 (3.1%)
High school graduate	10,044 (35%)
College graduate	17,748 (62%)
Missing	364
Smoking	
No	26,555 (93%)
Yes	1,951 (6.8%)
Missing	548
Alcohol	
Never	5,168 (19%)
Monthly Or Less	7,639 (28%)
Weekly Or More	14,498 (53%)
Missing	1,749
Frailty index (if <20% missing)	
Robust	13,472 (49%)
Pre-frail	9,068 (33%)
Frail	5,035 (18%)

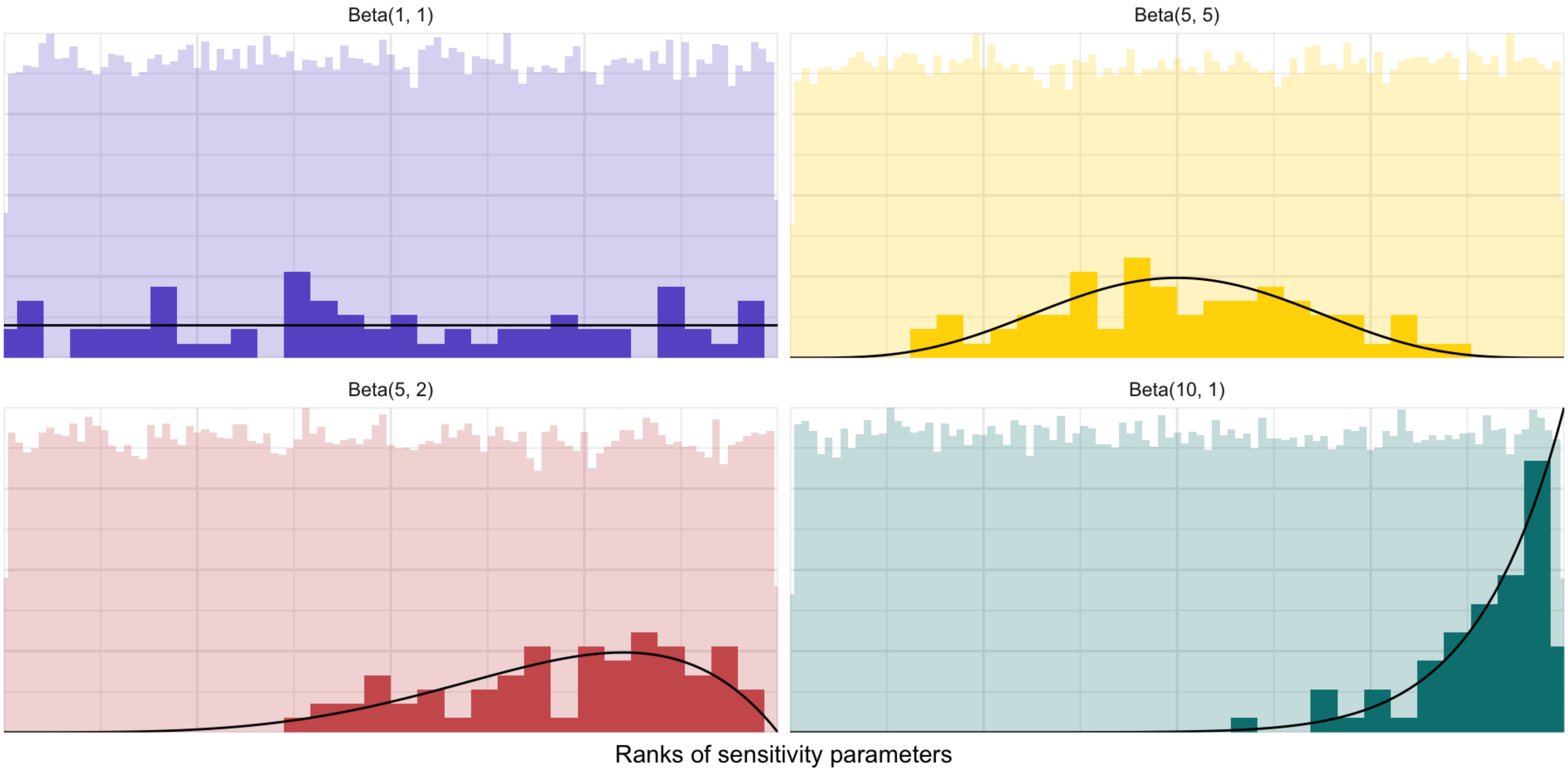
Mason et al. (2017); D. Tompsett et al. (2020)

Analysis: FI distribution

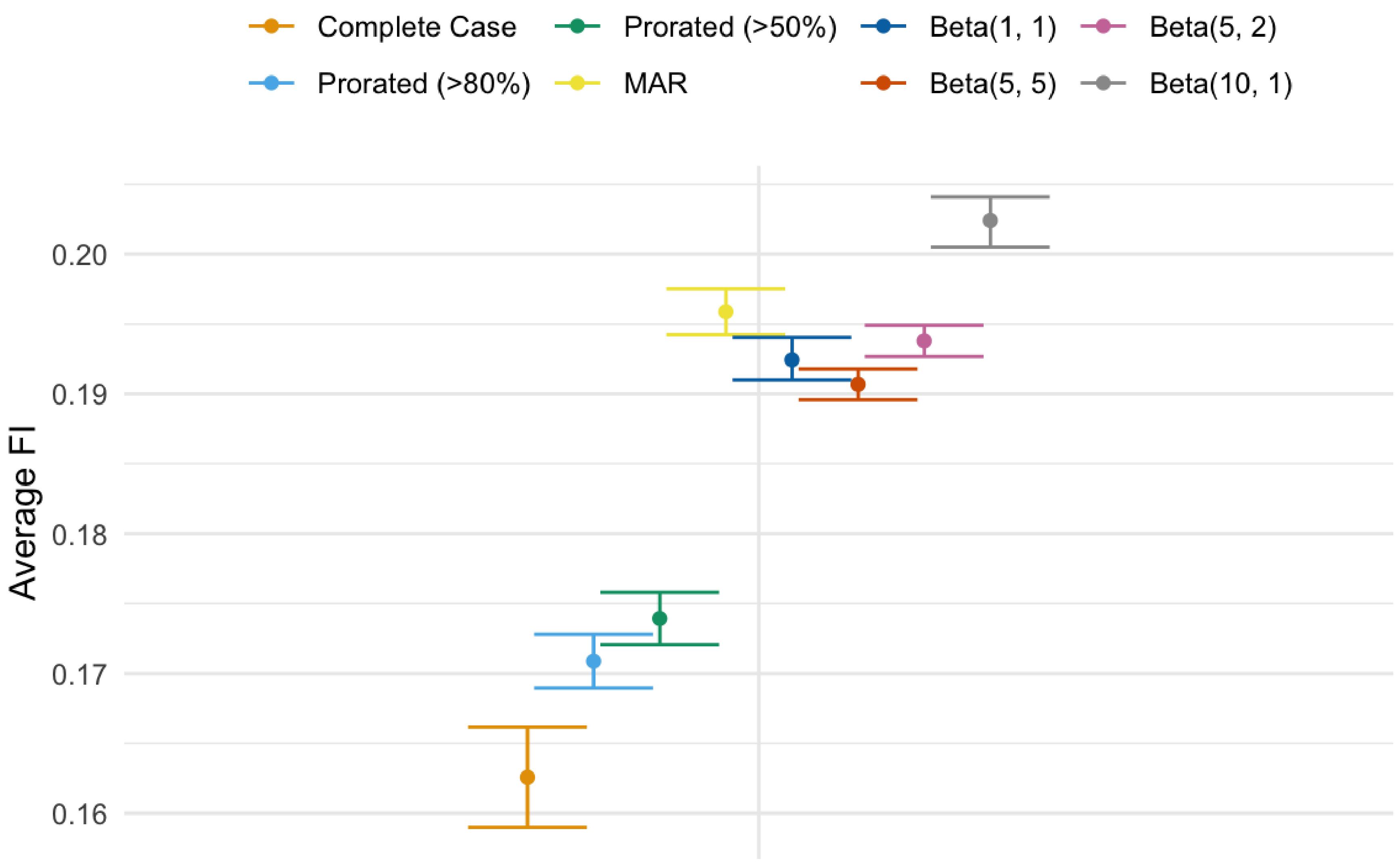
Synthetic AoU dataset

- complete case
- proration > 50% complete
- proration > 80% complete
- MAR (MICE with no delta-adjustment)
- MNAR, drawing sensitivity parameters from various distributions taking in account possible correlations
  - draw from triangle distribution, individually
  - compute rank within all draws
  - draw across all items by rank to allow for correlation

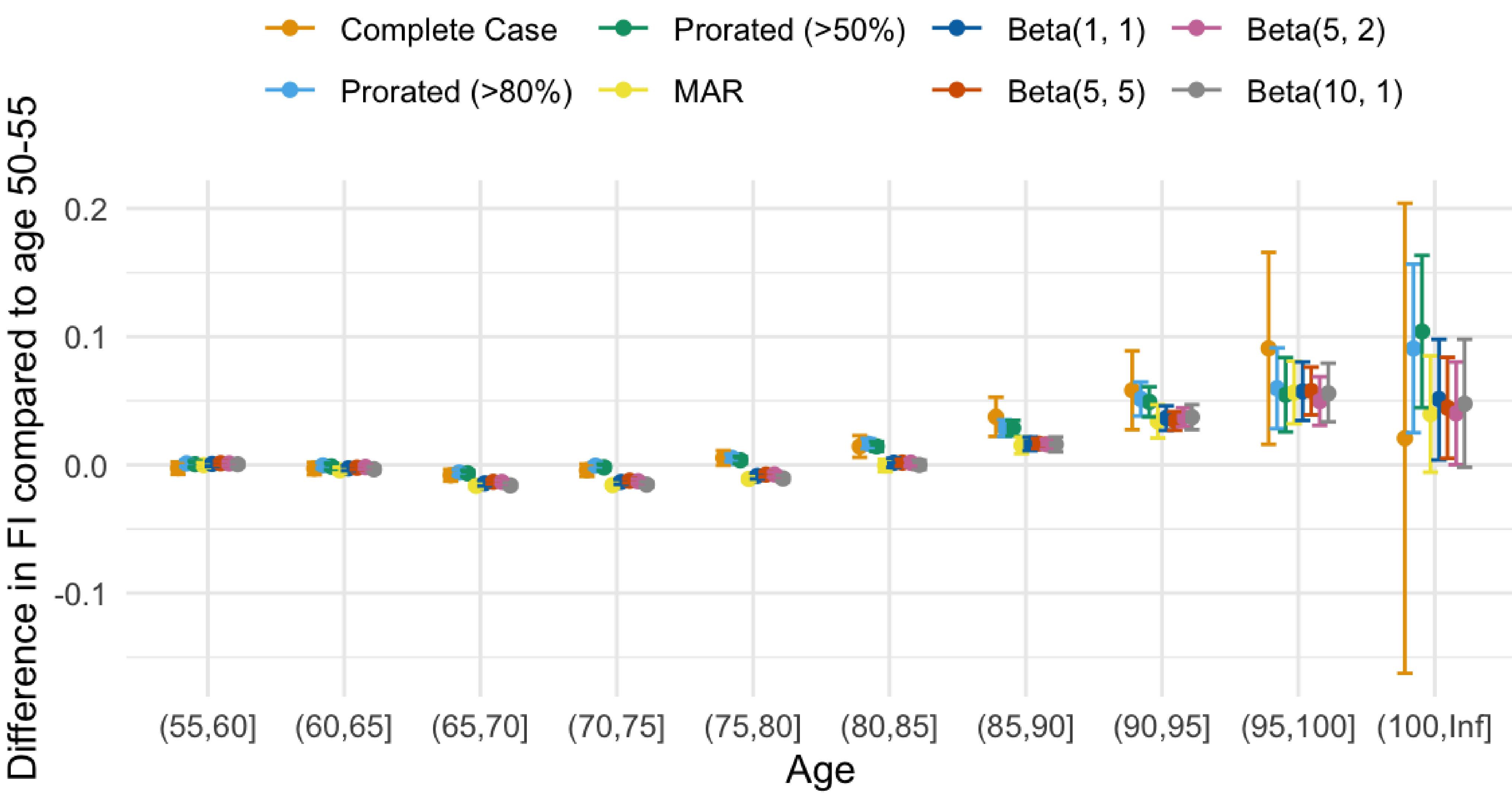
Distributions of sensitivity parameters

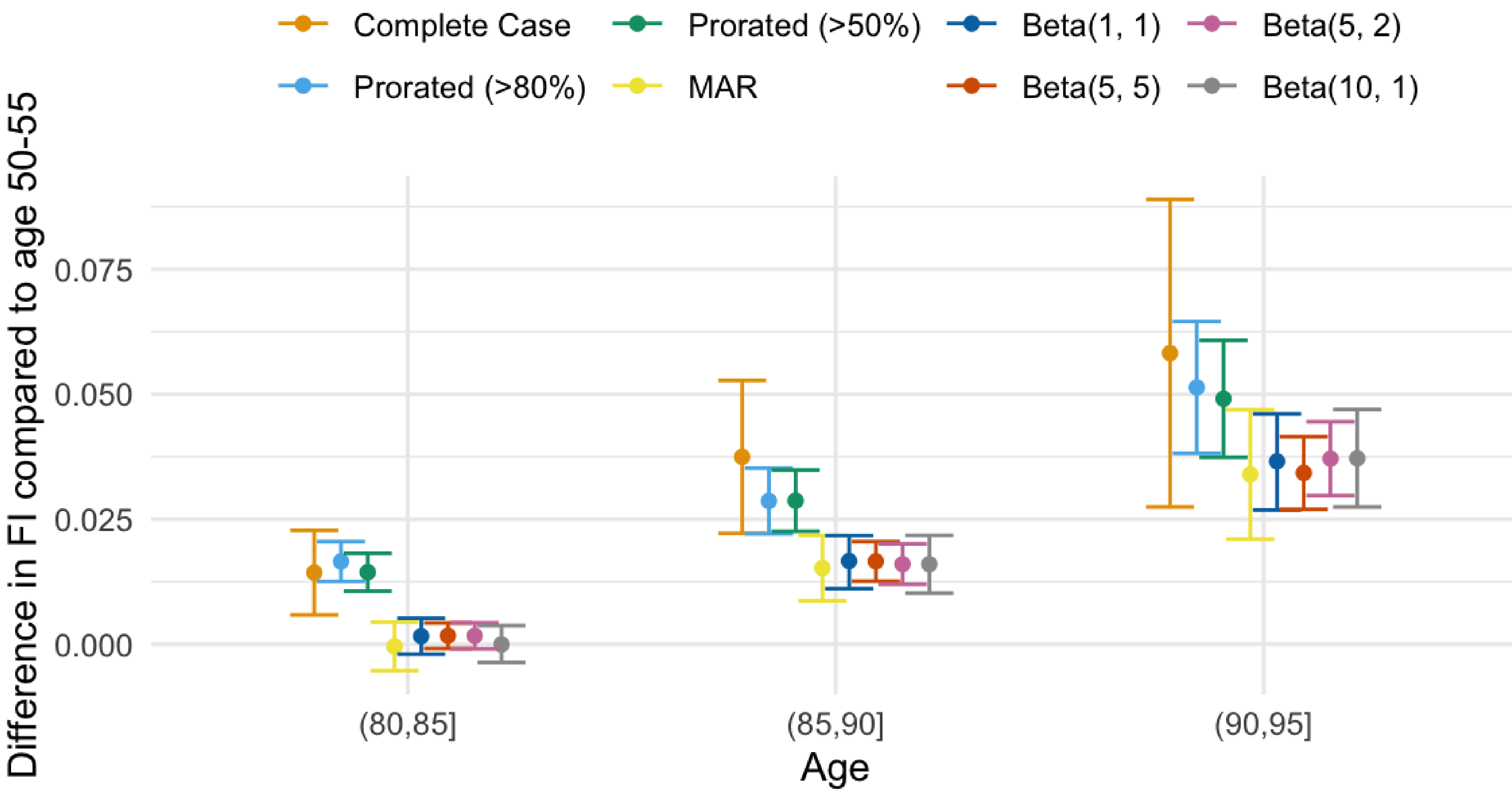


Average FI age 50-55

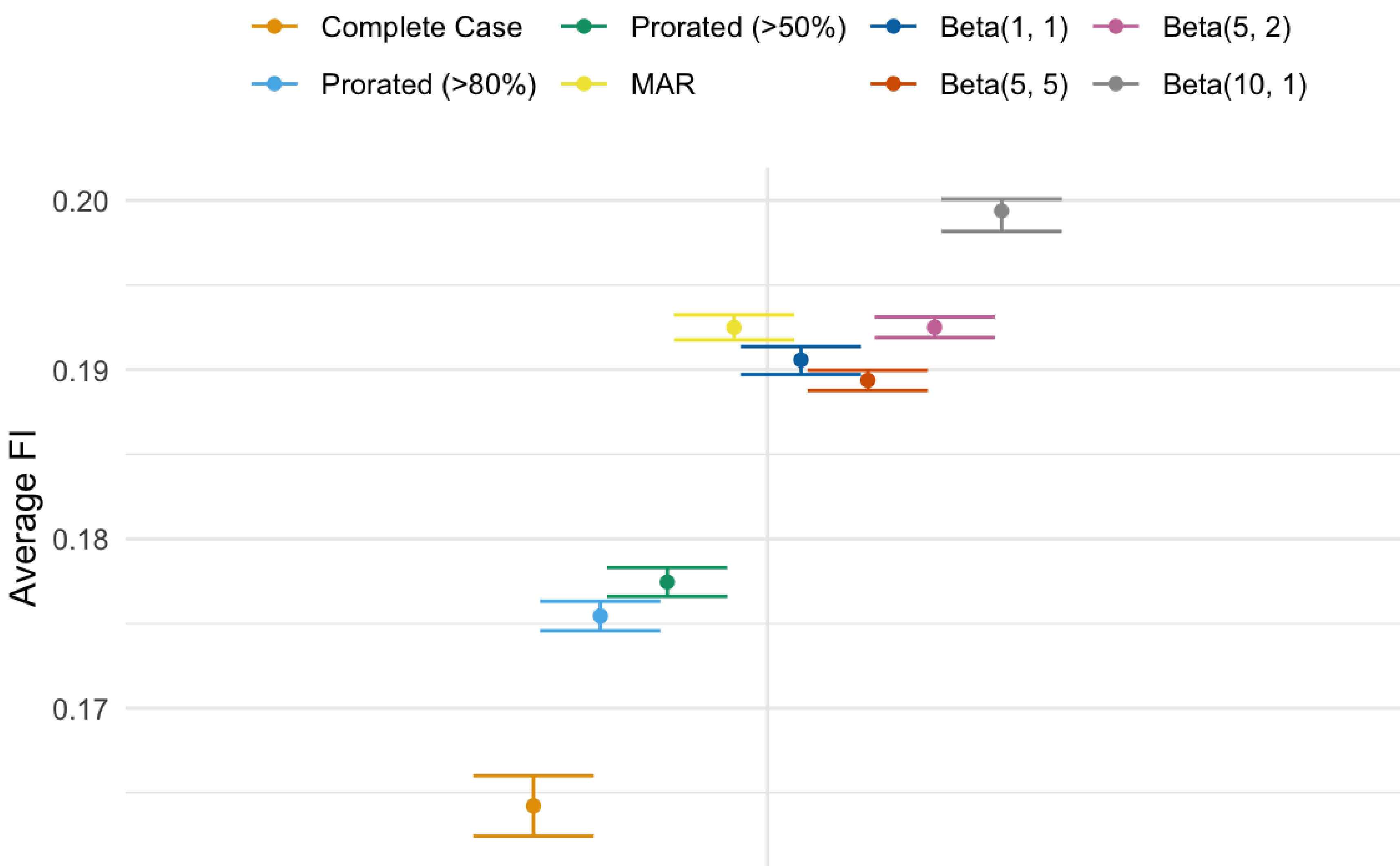


Age differences in FI

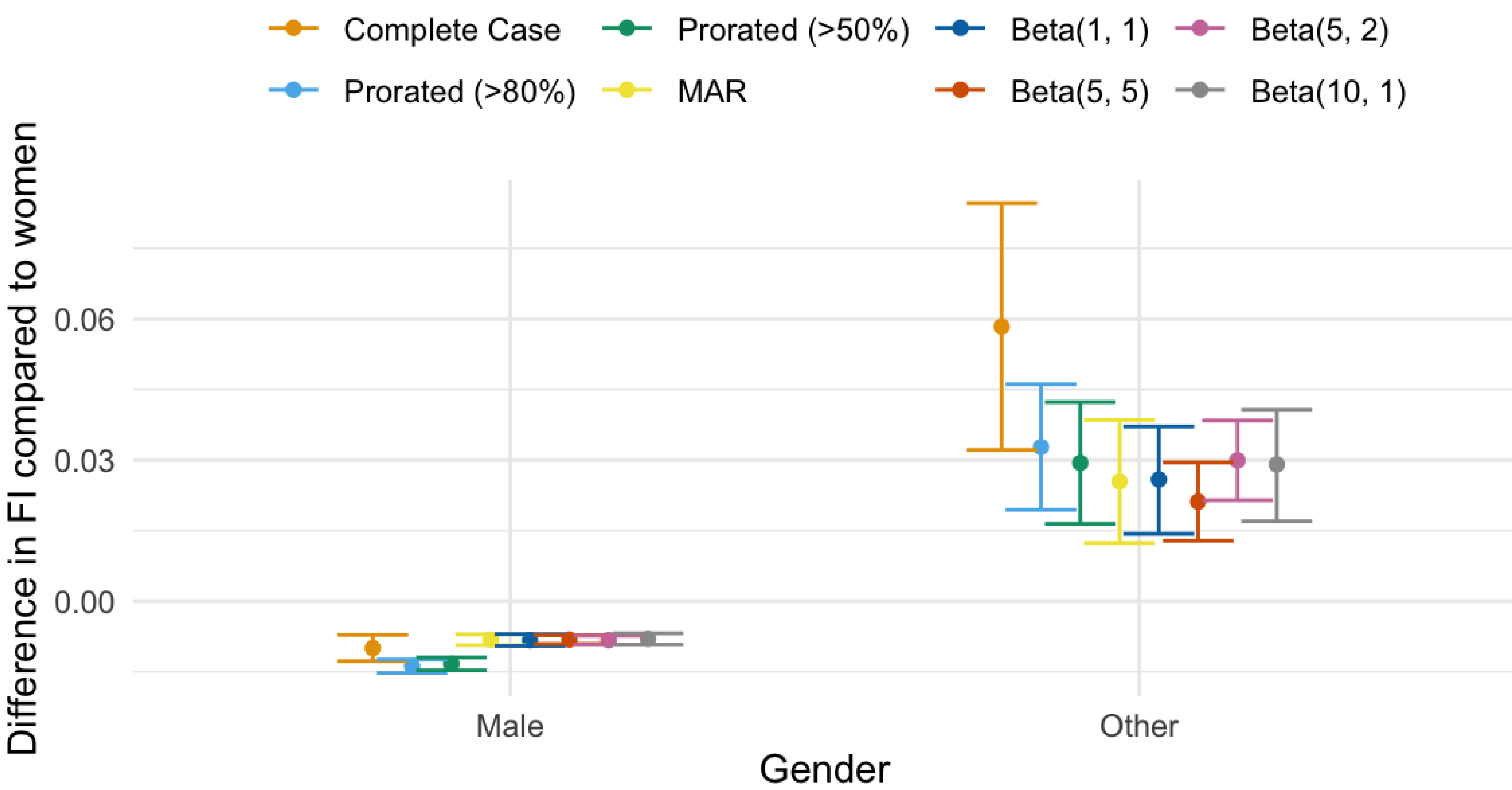



Complete Case
Prorated (>80%)
Prorated (>50%)
MAR
Beta(1, 1)
Beta(5, 5)
Beta(5, 2)
Beta(10, 1)
Difference in FI compared to age 50-55
0.075
0.050
0.025
0.000
(80,85]
(85,90]
(90,95]
Age

Average FI among women



Age-adjusted gender differences



Conclusions and future directions

Observations with missing data are quite different, but it's not clear that reasonable non-random missingness makes any difference

- Deal with computational challenges
  - Is it necessary to recompute frailty index in between every item?
- At what point is this necessary?
  - "Tipping point" analysis

Thanks to Chelsea Wong MD, Ariela Orkaby MD, Brianne Olivieri-Mui PhD

Buuren, Stef van. 2012. *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton, FL: CRC Press.

Little, Roderick J. A. 1993. "Pattern-Mixture Models for Multivariate Incomplete Data." *Journal of the American Statistical Association* 88 (421): 125–34. https://doi.org/10.2307/2290705.

Mason, Alexina J, Manuel Gomes, Richard Grieve, Pinar Ulug, Janet T Powell, and James Carpenter. 2017. "Development of a Practical Approach to Expert Elicitation for Randomised Controlled Trials with Missing Health Outcomes: Application to the IMPROVE Trial." *Clinical Trials* 14 (4): 357–67. https://doi.org/10.1177/1740774517711442.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons. https://doi.org/10.1002/9780470316696.

Tompsett, Daniel Mark, Finbarr Leacy, Margarita Moreno-Betancur, Jon Heron, and Ian R. White. 2018. "On the Use of the Not-at-Random Fully Conditional Specification (NARFCS) Procedure in Practice." *Statistics in Medicine* 37 (15): 2338–53. https://doi.org/10.1002/sim.7643.

Tompsett, Daniel, Stephen Sutton, Shaun R. Seaman, and Ian R. White. 2020. "A General Method for Elicitation, Imputation, and Sensitivity Analysis for Incomplete Repeated Binary Data." *Statistics in Medicine* 39 (22): 2921–35. https://doi.org/10.1002/sim.8584.

Van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal Of Statistical Software* 45 (3):

1–67. https://doi.org/10.1177/0962280206074463.
Wong, Chelsea, Michael P. Wilczek, Louisa H. Smith, Jordon D. Bosse, Erin L. Richard, Robert Cavanaugh, Justin Manjourides, Ariela R. Orkaby, and Brianne Olivieri-Mui. 2023. "Frailty Among Sexual and Gender Minority Older Adults: The All of Us Database." *Journal of Gerontology: Medical Sciences*, in press.