JENNIFER THOMPSON, MPH @jent103 ◆ jennifer.l.thompson@vanderbilt.edu
RAMEELA CHANDRASEKHAR, PHD
MAYUR PATEL, MD, MPH

HANDLING BIAS CAUSED BY MISSING DELIRIUM ASSESSMENTS

EXAMPLE: TRAUMATIC BRAIN INJURY COHORT

2,664

1,282 (48%) **complete** mental status data

1,383 (52%) **at least one day** with no mental status assessment

1198 (45%) have **some** assessments

185 (7%) have **no** assessments

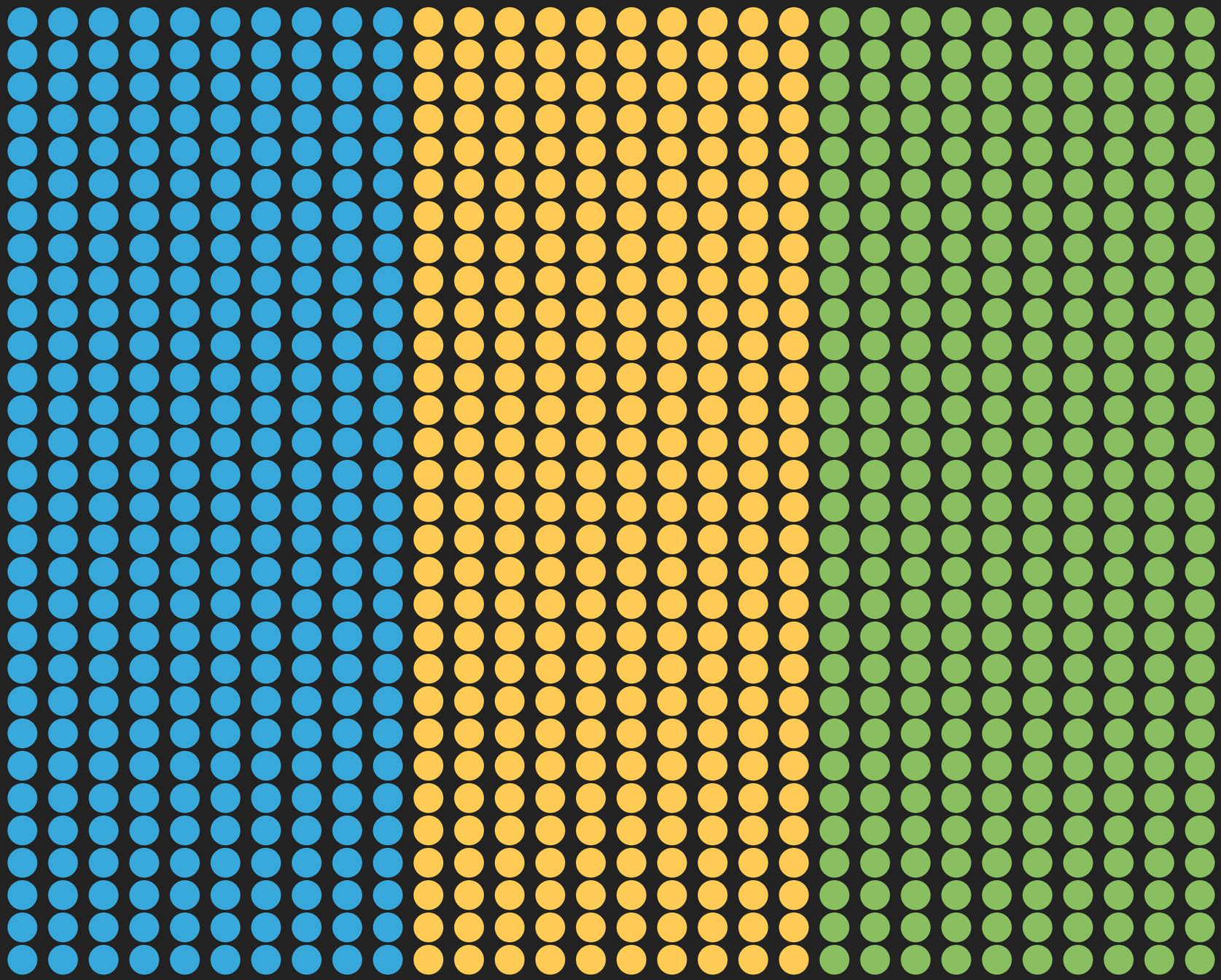
Compared to patients with complete data, patients with no assessments are:

- older
- less “injured”
- less time on MV
- more likely to be discharged home; less likely to go to rehab

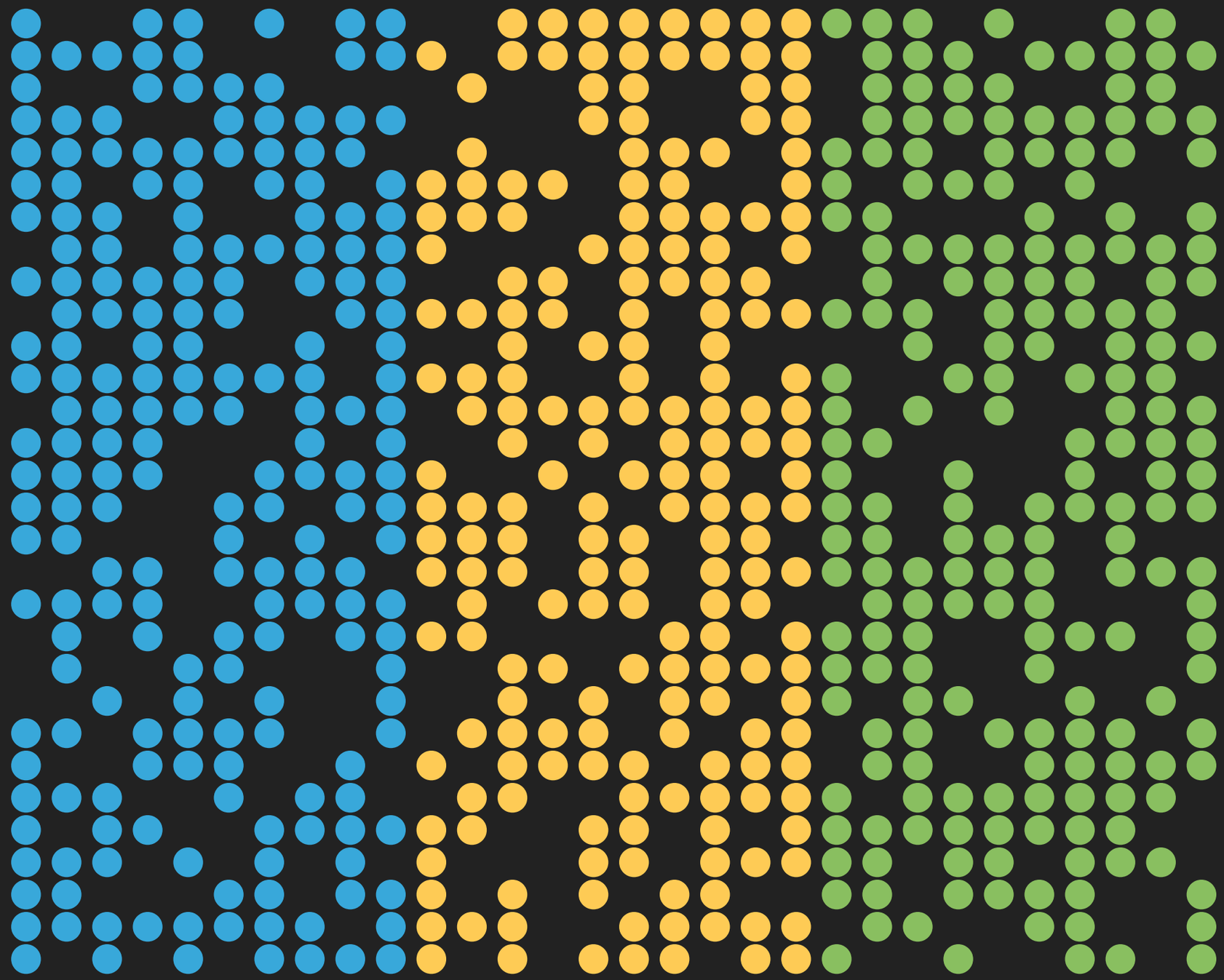
EXAMPLE: TRAUMATIC BRAIN INJURY COHORT

- Assume main exposure is duration of delirium during index hospitalization
- Value of exposure differs greatly according to how we treat hospital days with missing assessments
- Truth is probably between these extremes
- How we handle this matters: We're afraid of getting **biased results**

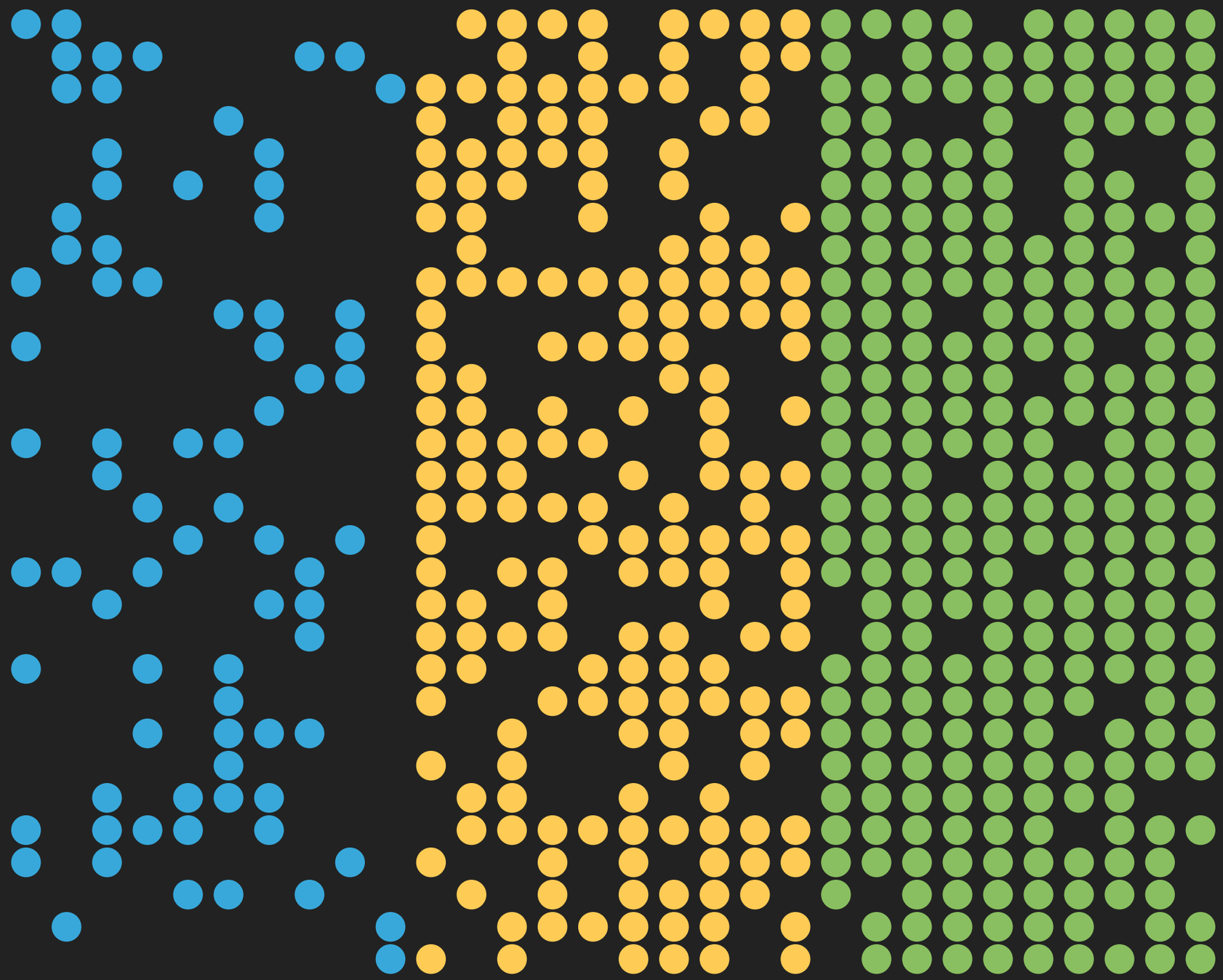
TYPES OF MISSING DATA



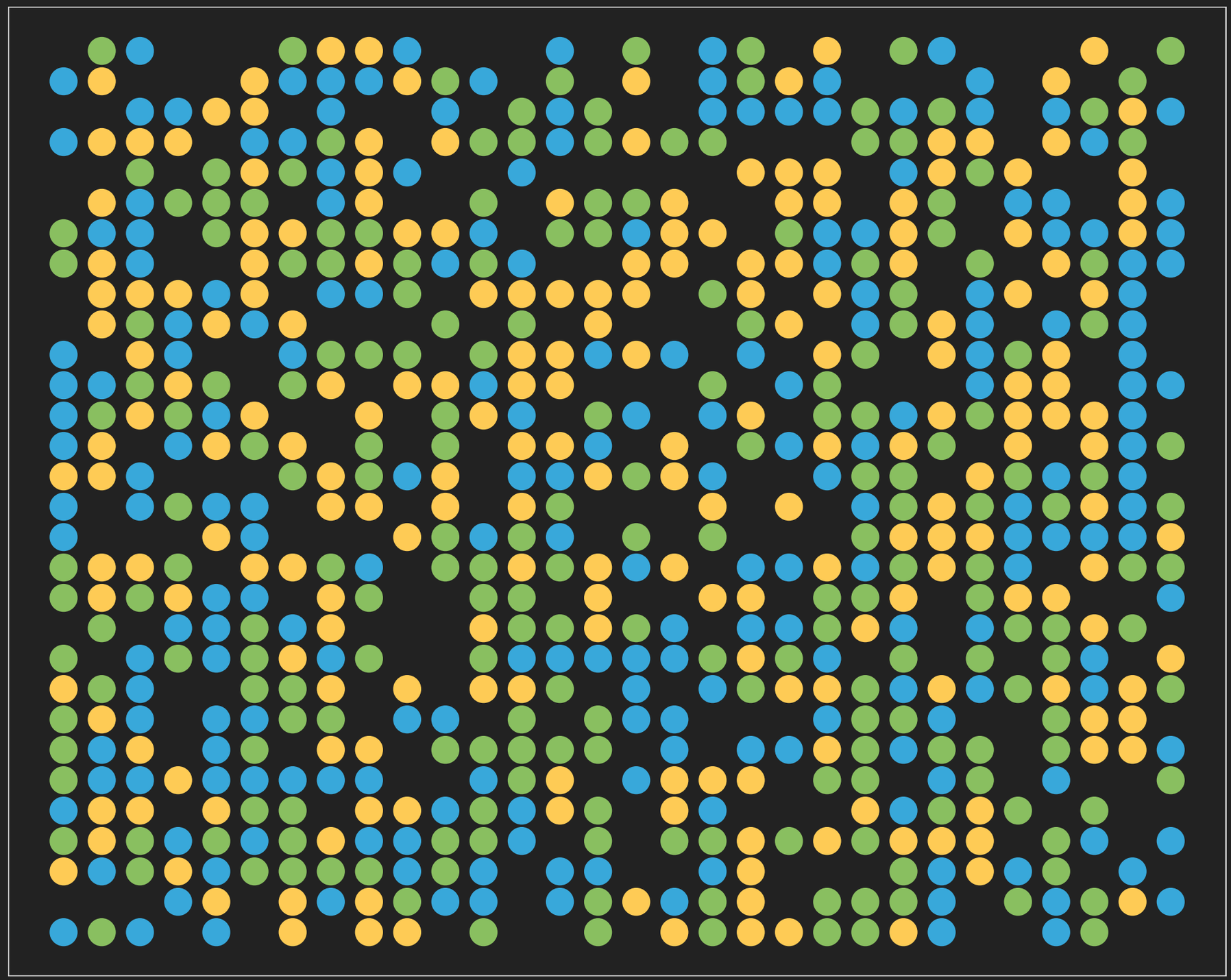
MISSING COMPLETELY AT RANDOM (MCAR)



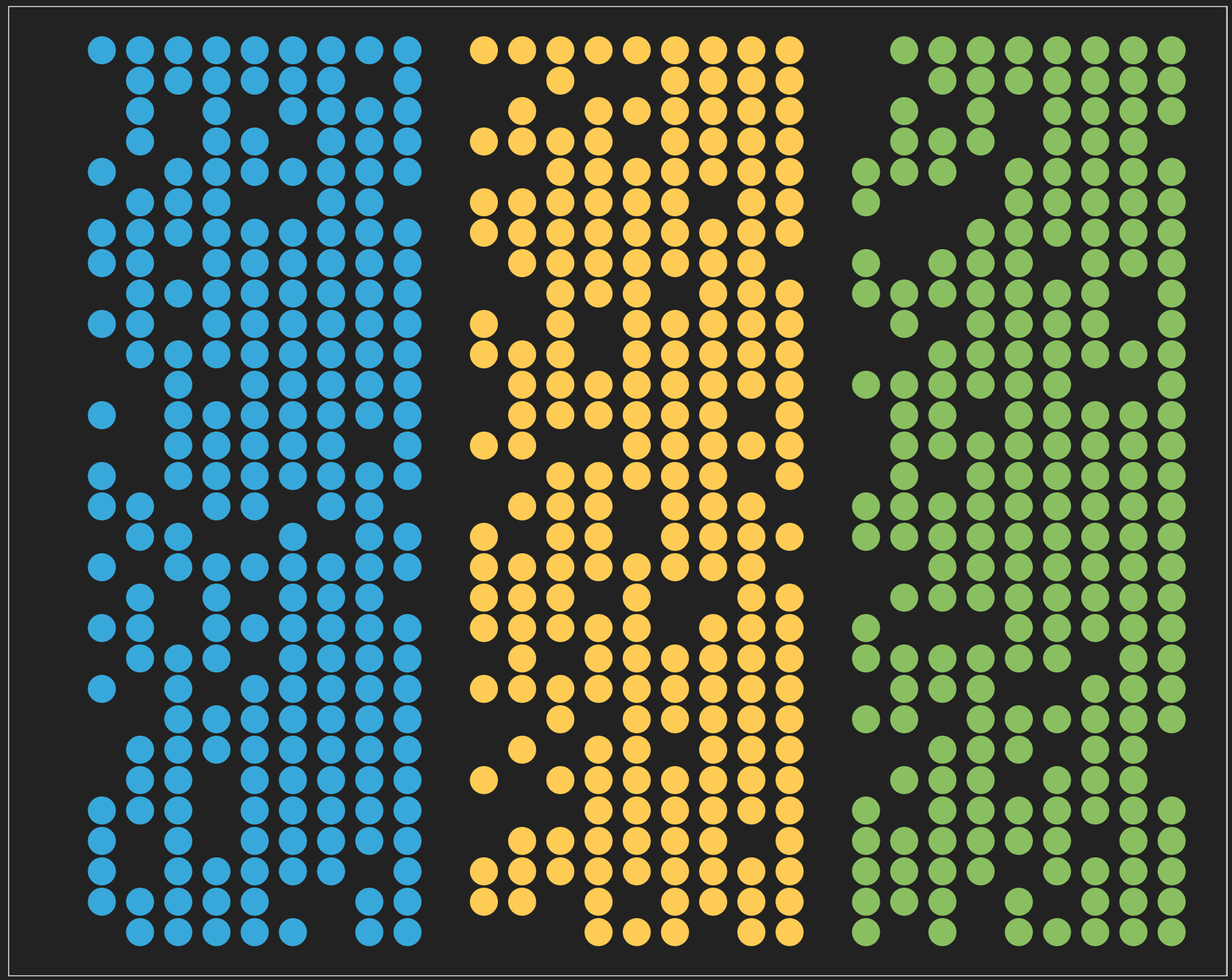
MISSING AT RANDOM (MAR)



MISSING NOT AT RANDOM (MNAR)



MISSING NOT AT RANDOM (MNAR)

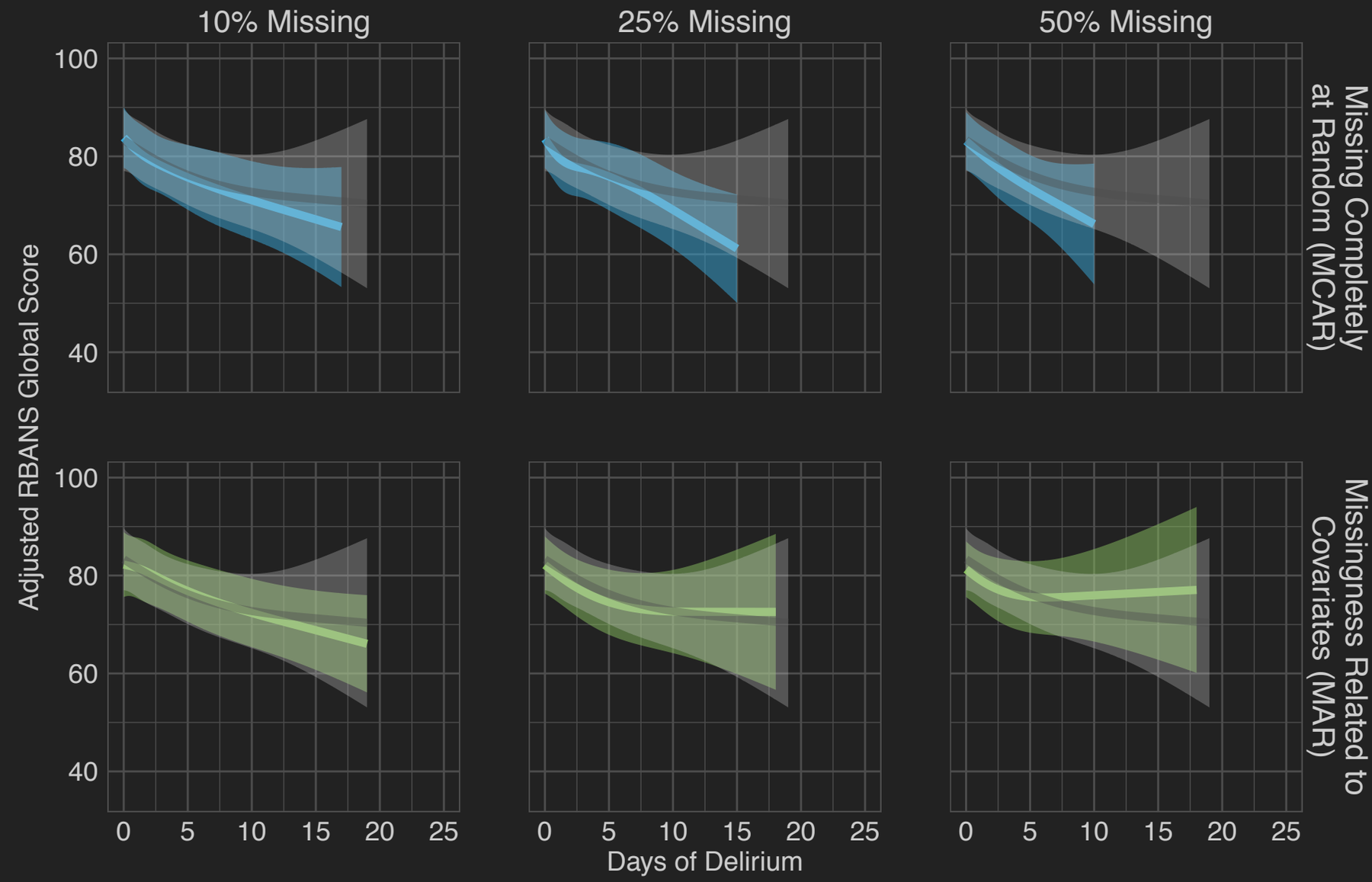


EXAMPLE OF BIAS: MODIFYING BRAIN–ICU

Link: Full BRAIN-ICU manuscript, NEJM

- Study goal: association of delirium duration in hospital & long-term cognition
- Original data had very little missingness; we assume study results are Truth
- What might have been, if we had had more missing assessments:

RESULTS USING DATA WITH FORCED MISSINGNESS



Grayscale = original study results; color = modified as noted

THAT'S NOT GREAT... SO WHAT ARE MY OPTIONS?

image source: Benjamin Reay via Flickr

1. IGNORE IT

i.e., complete case analysis

- PROS
  - Easy
  - Probably OK if you have very little missing data (<3-5%)
- CONS
  - Can lead to biased, misleading results
  - Reduces sample size and therefore power (see TBI example)

2. SINGLE IMPUTATION: MAKE AN ASSUMPTION

Make assumption(s) about values for missing assessments, then proceed with analysis.

Classic examples: imputing the mean; last observation carried forward (LOCF)

- PROS
  - Still pretty easy; probably OK with very little missing data (<3-5%)
- CONS
  - If we don't have very strong rationale, and/or have lots of missingness, can still lead to bias - this is especially problematic with quickly-changing mental status
  - Doesn't account for uncertainty surrounding our assumptions

Possible (But Not Necessarily Advised) Techniques:

- Assume all missing assessments are normal/delirious/coma
- Use a little clinical info:
  - LOCF
  - Missing among survivors = normal; missing among deceased = coma
  - BRAIN-ICU approach: impute status based on the days before & after missing assessment

3. MULTIPLE IMPUTATION: USE WHAT WE'VE GOT

Use data on all covariates to make **multiple** predictions for missing assessments.
Account for uncertainty during analysis.

- PROS
  - Uses all available information
  - Does the most to reduce bias
  - Accounts for uncertainty in predictions
- CONS
  - More complicated
  - Depends on predictors being as/more reliable than delirium assessments
  - Assumes assessments are **missing at random**

3. MULTIPLE IMPUTATION: USE WHAT WE'VE GOT

Use data on all covariates to make **multiple** predictions for missing assessments.
Account for uncertainty during analysis.

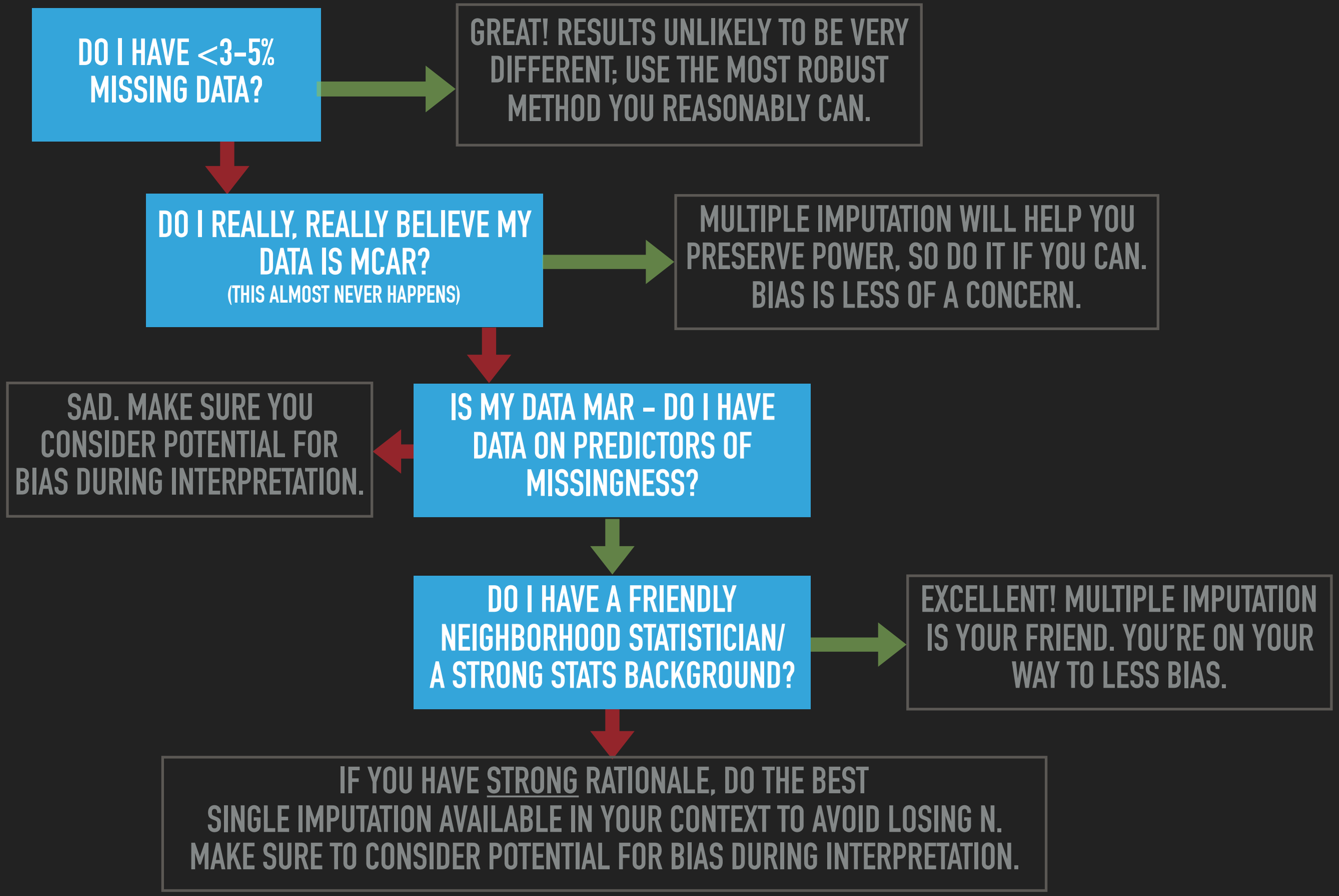
BUT WAIT. . . ISN'T IMPUTATION MAKING UP DATA?

In a word: no.

"The goal of imputation is to preserve the information and meaning of the non-missing data." - Frank Harrell

Ignoring missing data is more of a disservice to patients and to science than addressing the resulting bias head-on.

SUMMARY


DO I HAVE <3-5% MISSING DATA?
GREAT! RESULTS UNLIKELY TO BE VERY DIFFERENT; USE THE MOST ROBUST METHOD YOU REASONABLY CAN.
DO I REALLY, REALLY BELIEVE MY DATA IS MCAR?
(THIS ALMOST NEVER HAPPENS)
MULTIPLE IMPUTATION WILL HELP YOU PRESERVE POWER, SO DO IT IF YOU CAN. BIAS IS LESS OF A CONCERN.
SAD. MAKE SURE YOU CONSIDER POTENTIAL FOR BIAS DURING INTERPRETATION.
IS MY DATA MAR - DO I HAVE DATA ON PREDICTORS OF MISSINGNESS?
DO I HAVE A FRIENDLY NEIGHBORHOOD STATISTICIAN/ A STRONG STATS BACKGROUND?
EXCELLENT! MULTIPLE IMPUTATION IS YOUR FRIEND. YOU'RE ON YOUR WAY TO LESS BIAS.
IF YOU HAVE STRONG RATIONALE, DO THE BEST SINGLE IMPUTATION AVAILABLE IN YOUR CONTEXT TO AVOID LOSING N. MAKE SURE TO CONSIDER POTENTIAL FOR BIAS DURING INTERPRETATION.

RESOURCES

- Janssen et al. "Missing covariate data in medical research: To impute is better than to ignore." *Journal of Clinical Epidemiology,* 2010 Jul;63(7):721-7. DOI: 10.1016/j.jclinepi.2009.12.008
- Donders et al. "Review: A gentle introduction to the imputation of missing values." *Journal of Clinical Epidemiology,* 2006 Oct;59(10):1087-91. DOI: 10.1016/j.jclinepi.2006.01.014
- Sterne et al. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." *The BMJ*, 2009;338:b2393. DOI: 10.1136/bmj.b2393

- Multiple imputation software options:
  - R: mice package by Stef van Buuren (package manual)
  - R: Hmisc::aregImpute() by Frank Harrell (documentation), usually used in conjunction with rms::fit.mult.impute()
  - Stata: mi (feature page)
  - SAS: PROC MI (user guide)
- Slides (and R code for visuals, if you're really into it): github.com/jenniferthompson/ADS2017