

# Analyzing Summary Variables in the Presence of Partially Missing Longitudinal Data

Jennifer L Thompson, MPH  
Rameela Chandrasekhar, PhD  
Vanderbilt University Medical Center

[bit.ly/jlt-jsm2018](http://bit.ly/jlt-jsm2018)

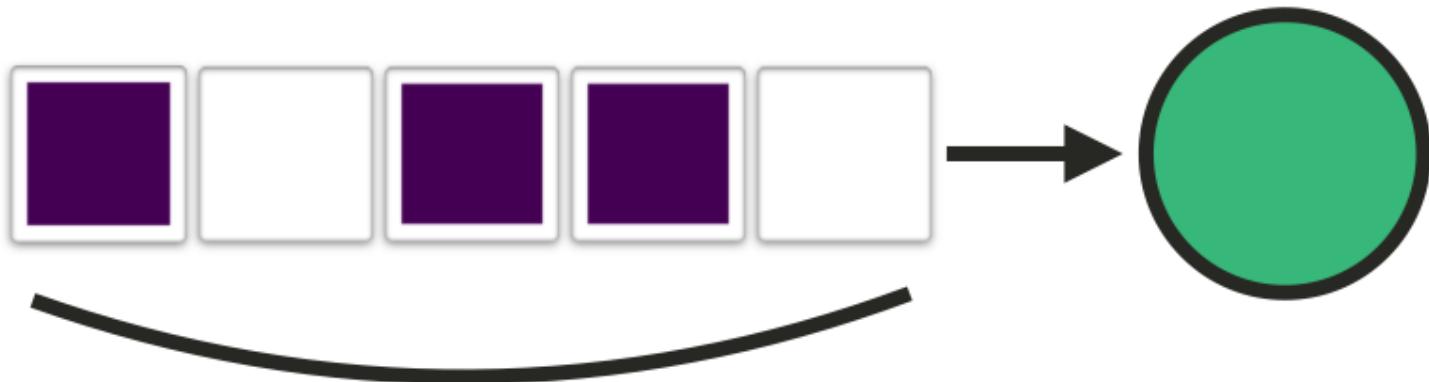
 [jenniferthompson/MissSumVars](https://github.com/jenniferthompson/MissSumVars)

 [@jent103](https://twitter.com/jent103)

# Our Motivation



# Our Motivation



# Our Motivation



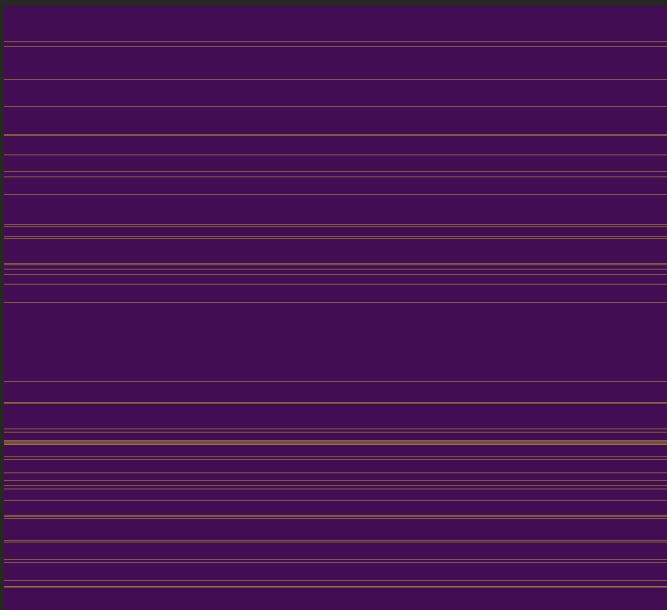
Is more delirium (*in-hospital brain dysfunction*) associated with scores on a cognitive test after hospital discharge?



# The Problem

## Prospective cohort

status (2.88%)

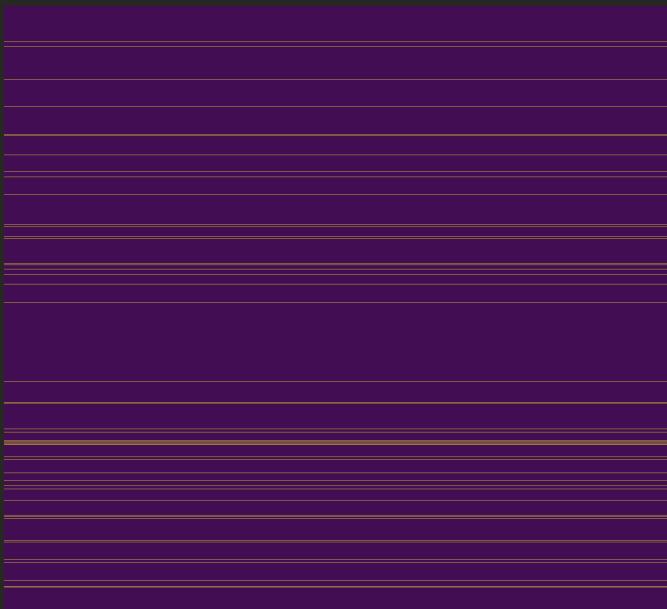


# The Problem

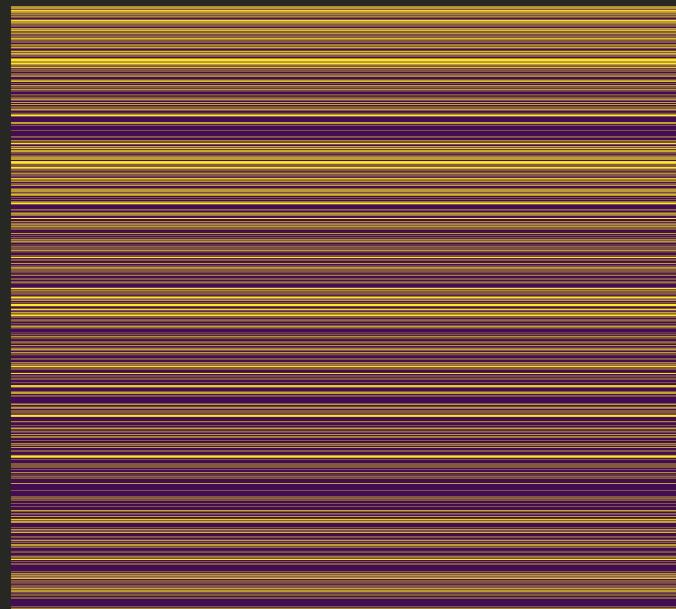
Prospective cohort

Retrospective EMR

status (2.88%)



status (32.19%)

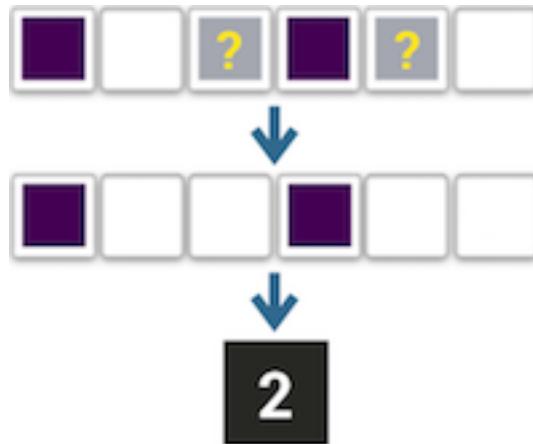


# Strategies: Simple

## Assume the "best"

*NA = unexposed:*

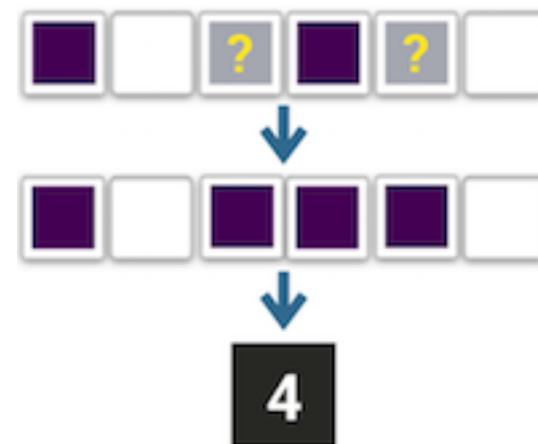
*Only count the exposure we know about*



## Assume the "worst"

*NA = exposed:*

*All missing time points get exposure*



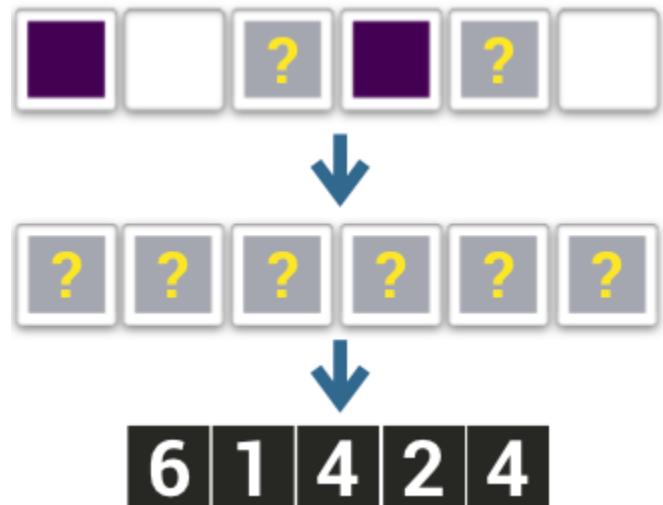
**Pros:** Straightforward to implement; plausible, if we know a lot about data collection

**Cons:** Prone to bias

# Strategies: Imputation

## Assume nothing

1. If subject missing any time point,  
entire summary value = NA
2. Multiply impute missing *summary values*  
before modeling



### Pros:

- Fairly simple to implement
- Acknowledges uncertainty

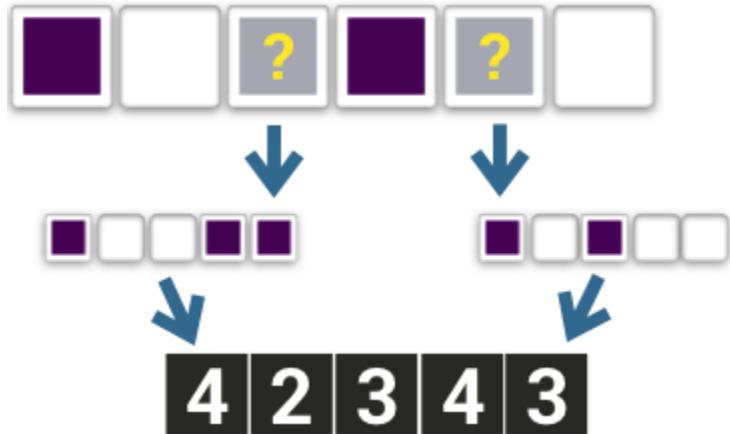
### Cons:

- Ignores the data we *do* have
- Likely to *overestimate* uncertainty

# Strategies: Imputation

## Assume the minimum

1. Multiply impute missing *time points*
2. Summarize each imputed dataset
3. Use these imputed summary datasets when modeling



### Pros:

- Maximizes use of data we have, including available covariate information

### Cons:

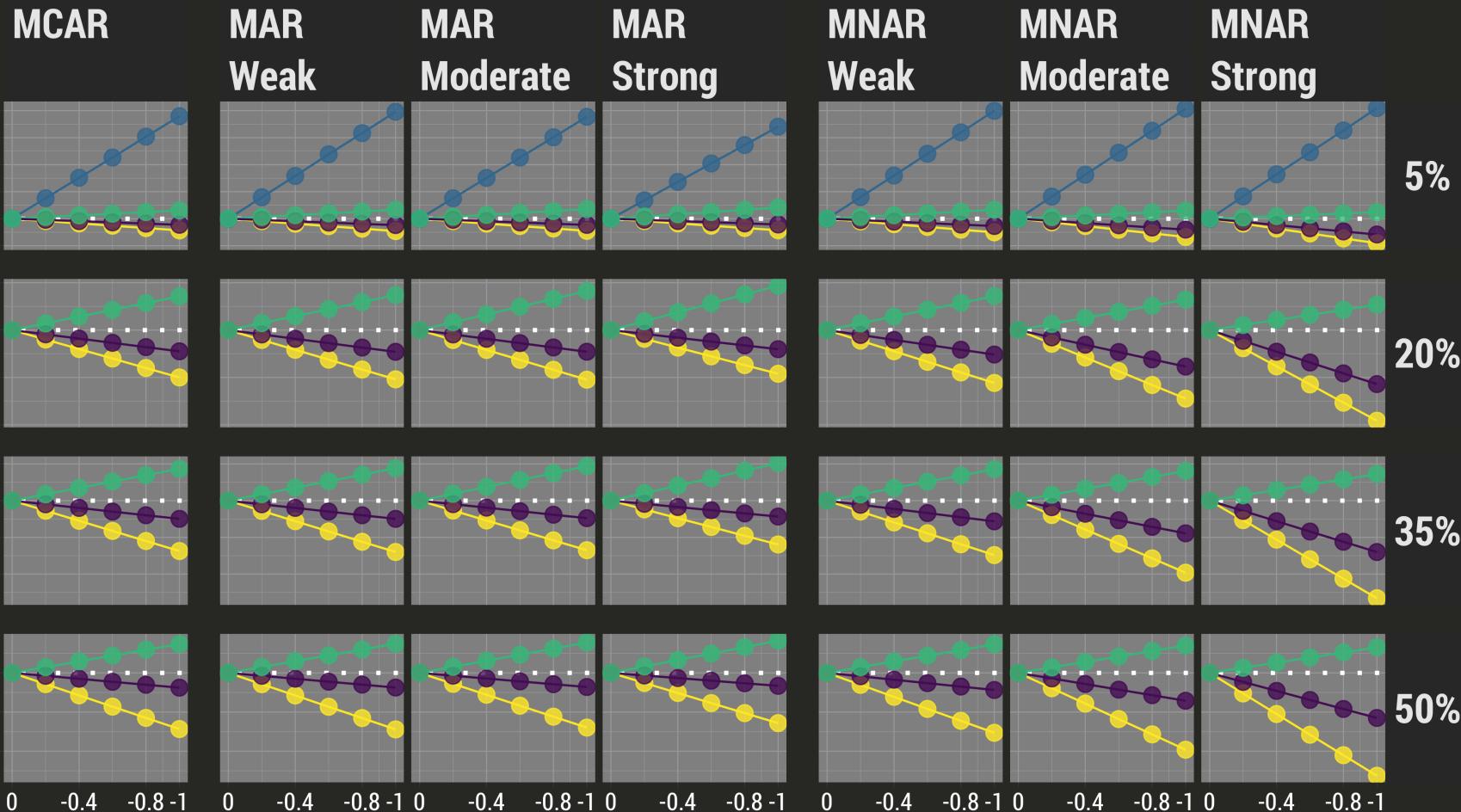
- Computation time
- Data wrangling

# Simulations

- 5%, 20%, 35%, 50% of patient-days missing exposure value
- Types of missingness:
  - MCAR
  - MAR
    - Missingness in exposure weakly, moderately, strongly associated with daily severity of illness
  - MNAR
    - Missingness in exposure weakly, moderately, strongly associated with true exposure value
- True relationship with outcome: ranges from 0 to -5
- Imputation methods incorporate severity of illness

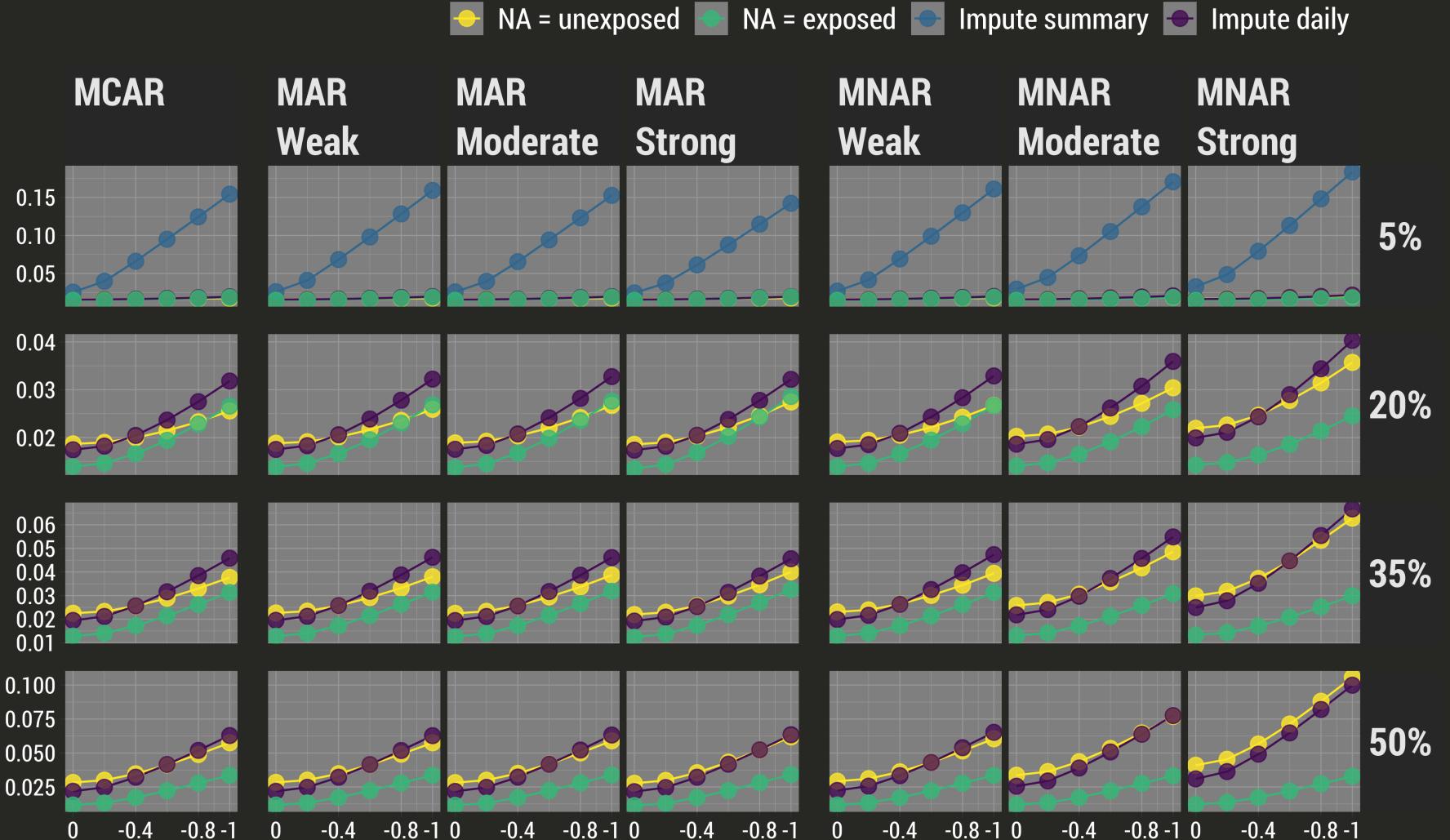
# BIAS

 NA = unexposed  NA = exposed  Impute summary  Impute daily



# True Effect Size

# STANDARD ERROR



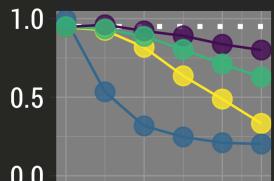
# True Effect Size



# COVERAGE

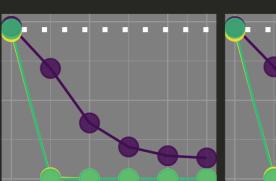
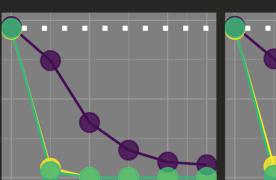
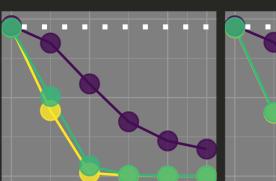
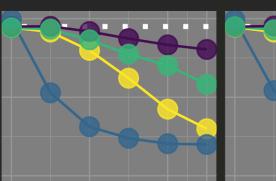
NA = unexposed NA = exposed Impute summary Impute daily

MCAR



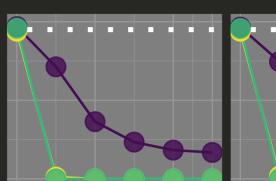
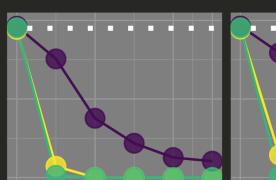
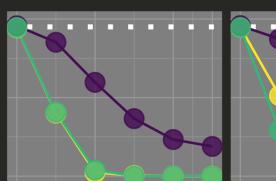
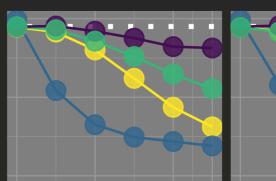
0 -0.4 -0.8 -1

MAR  
Weak



0 -0.4 -0.8 -1

MAR  
Moderate



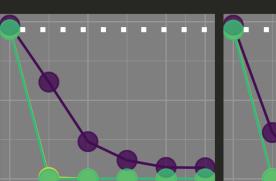
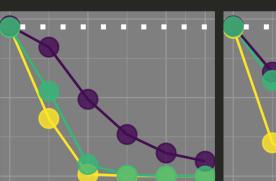
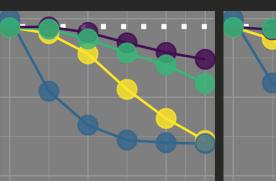
0 -0.4 -0.8 -1

MAR  
Strong



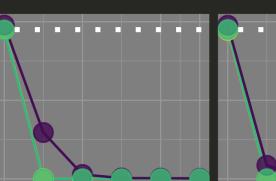
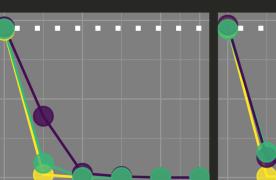
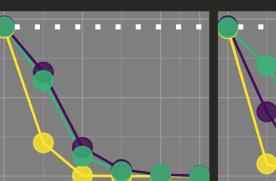
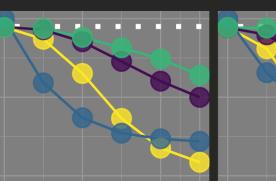
0 -0.4 -0.8 -1

MNAR  
Weak



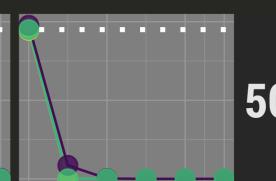
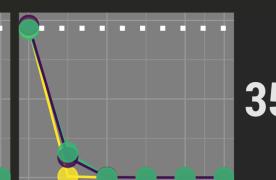
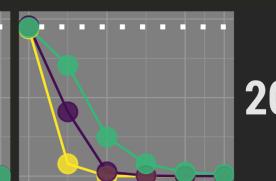
0 -0.4 -0.8 -1

MNAR  
Moderate



0 -0.4 -0.8 -1

MNAR  
Strong



0 -0.4 -0.8 -1

True Effect Size

# POWER

NA = unexposed NA = exposed Impute summary Impute daily

MCAR

MAR  
Weak

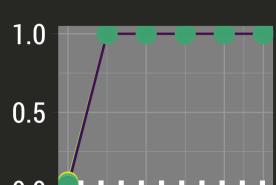
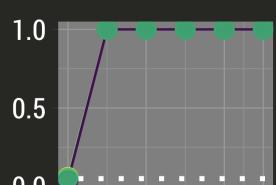
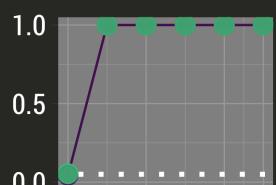
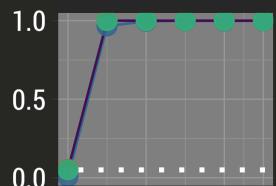
MAR  
Moderate

MAR  
Strong

MNAR  
Weak

MNAR  
Moderate

MNAR  
Strong



5%

20%

35%

50%

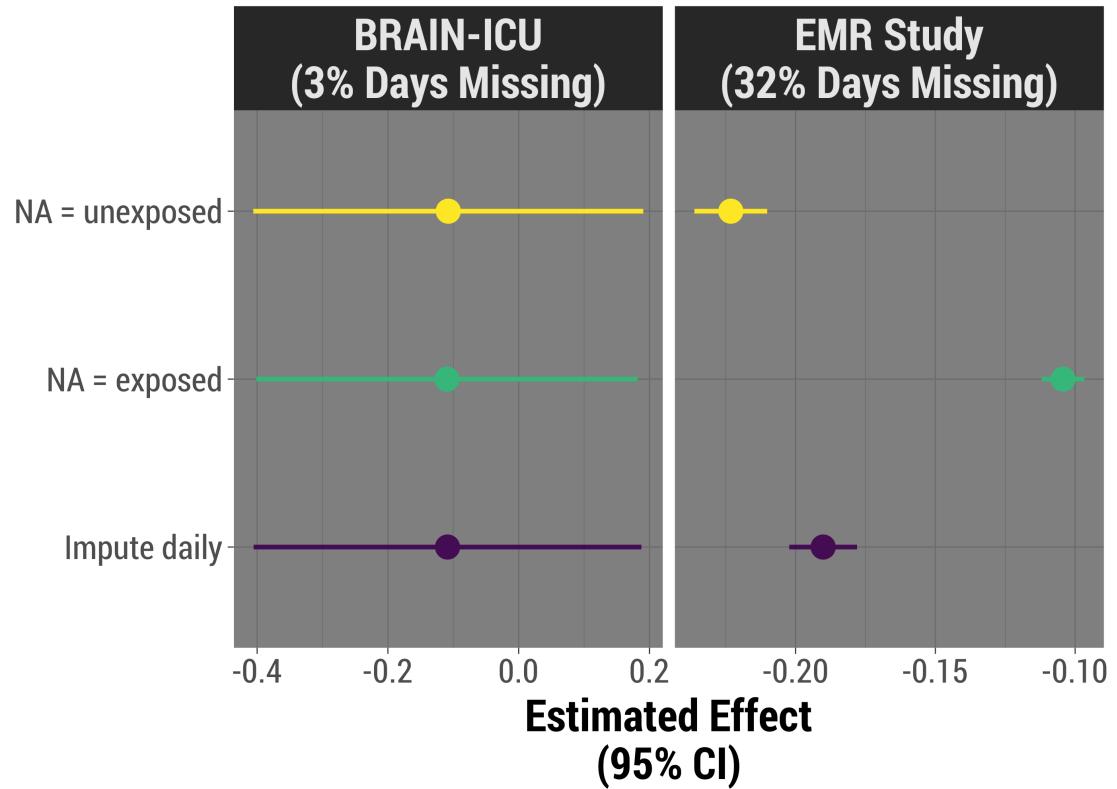
True Effect Size

# Examples with Motivating Data

- Duration of exposure vs outcome in two real-world studies, adjusting for one covariate
- Three strategies to summarize duration of exposure

# Examples with Motivating Data

- Duration of exposure vs outcome in two real-world studies, adjusting for one covariate
- Three strategies to summarize duration of exposure





# Takeaways

- Understand your data and its missingness
- Better safe than sorry: imputing at the lowest level is usually worth it
- Don't
  - Throw out data
  - Ignore your missingness

## Future Exploration

- Continuous exposure
- Complex relationships between covariates/exposure, missingness

# Acknowledgements

- Stef van Buuren for mice
- Davis Vaughn, Henrik Bengtsson for furrr and future
- Brooke Watson for BRRR
- VUMC CIBS Center PIs and research team

 [bit.ly/jlt-jsm2018](http://bit.ly/jlt-jsm2018)

 [@jent103](https://twitter.com/jent103)

 [@jenniferthompson](https://www.linkedin.com/in/jenniferthompson)

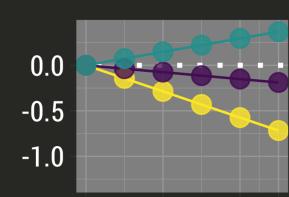
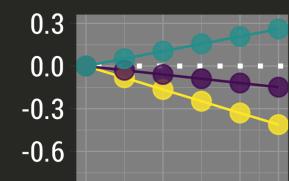
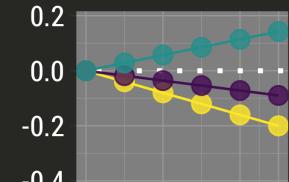
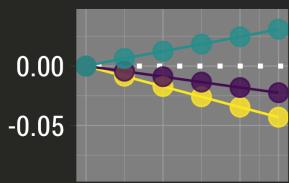
 [j.thompson@vumc.org](mailto:j.thompson@vumc.org)

 [jenthompson.me](http://jenthompson.me)

# BIAS

NA = unexposed NA = exposed Impute daily

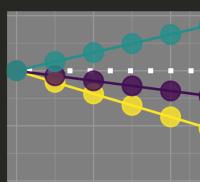
MCAR



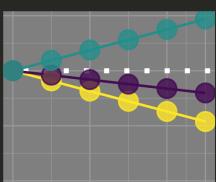
MAR  
Weak



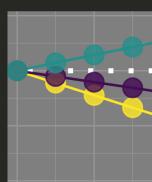
MAR  
Moderate



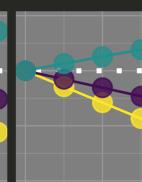
MAR  
Strong



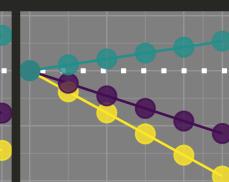
MNAR  
Weak



MNAR  
Moderate



MNAR  
Strong



5%

20%

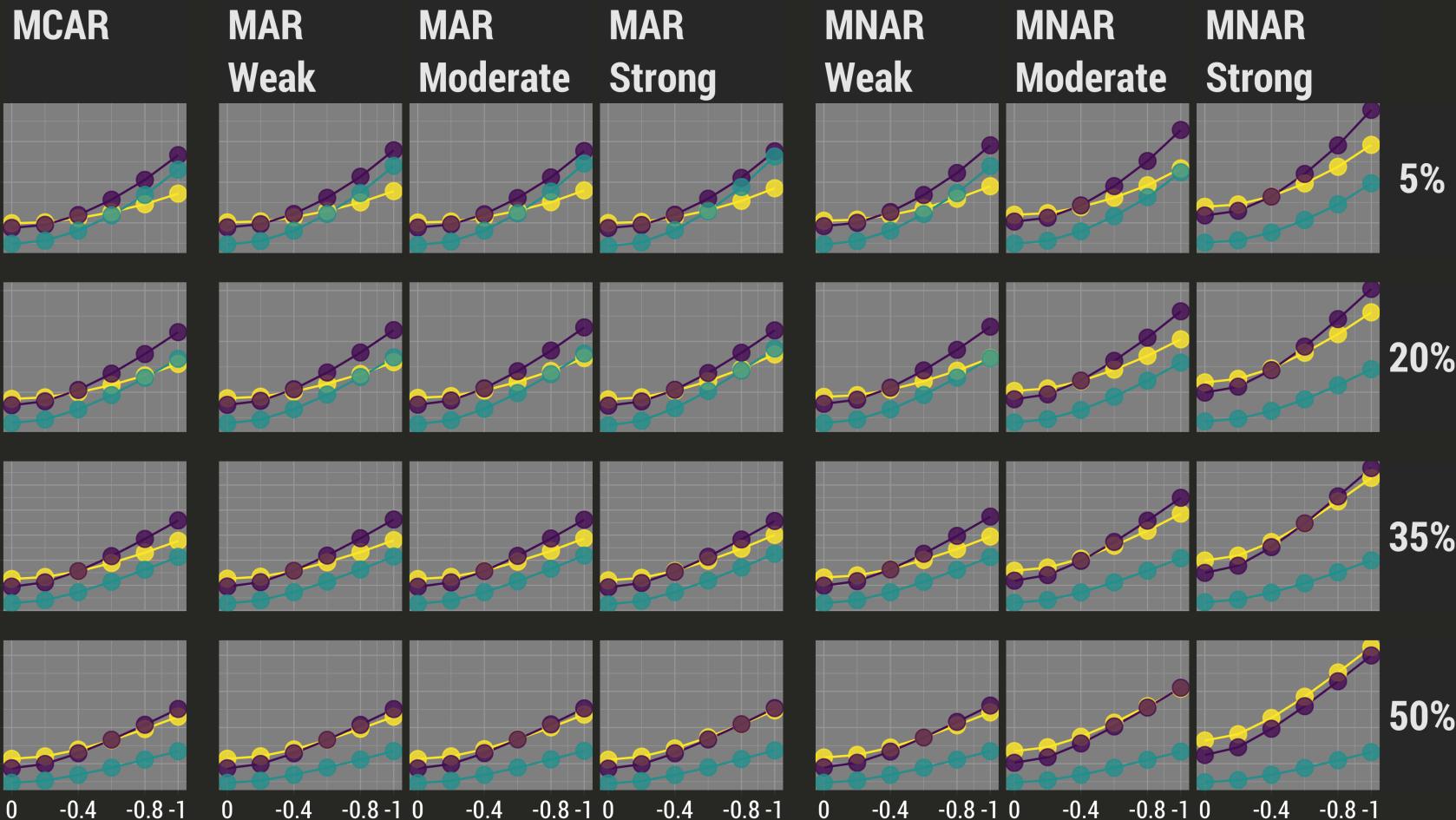
35%

50%

True Effect Size

# STANDARD ERROR

● NA = unexposed   ● NA = exposed   ● Impute daily



# True Effect Size

# CI WIDTH

NA = unexposed NA = exposed Impute daily

MCAR

MAR  
Weak

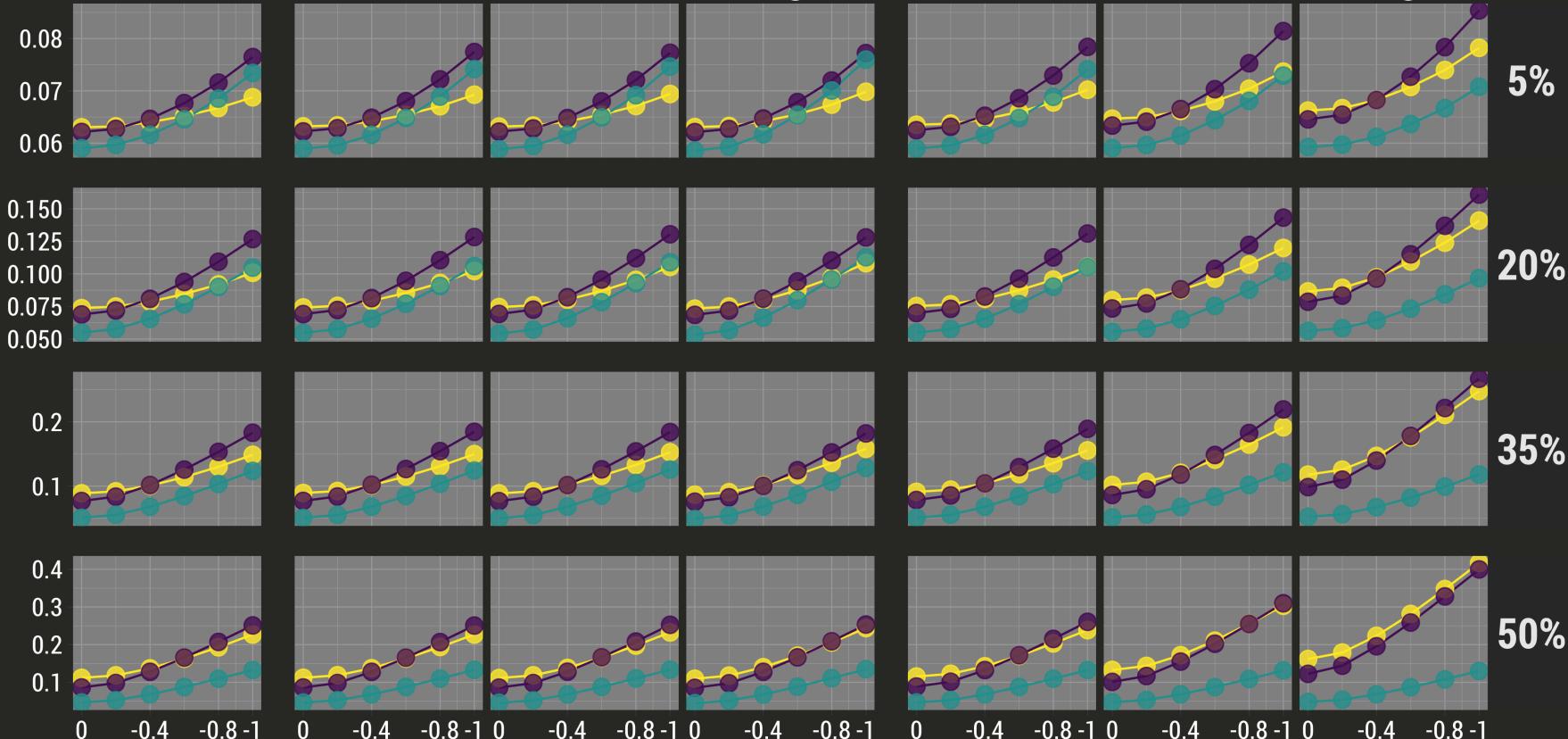
MAR  
Moderate

MAR  
Strong

MNAR  
Weak

MNAR  
Moderate

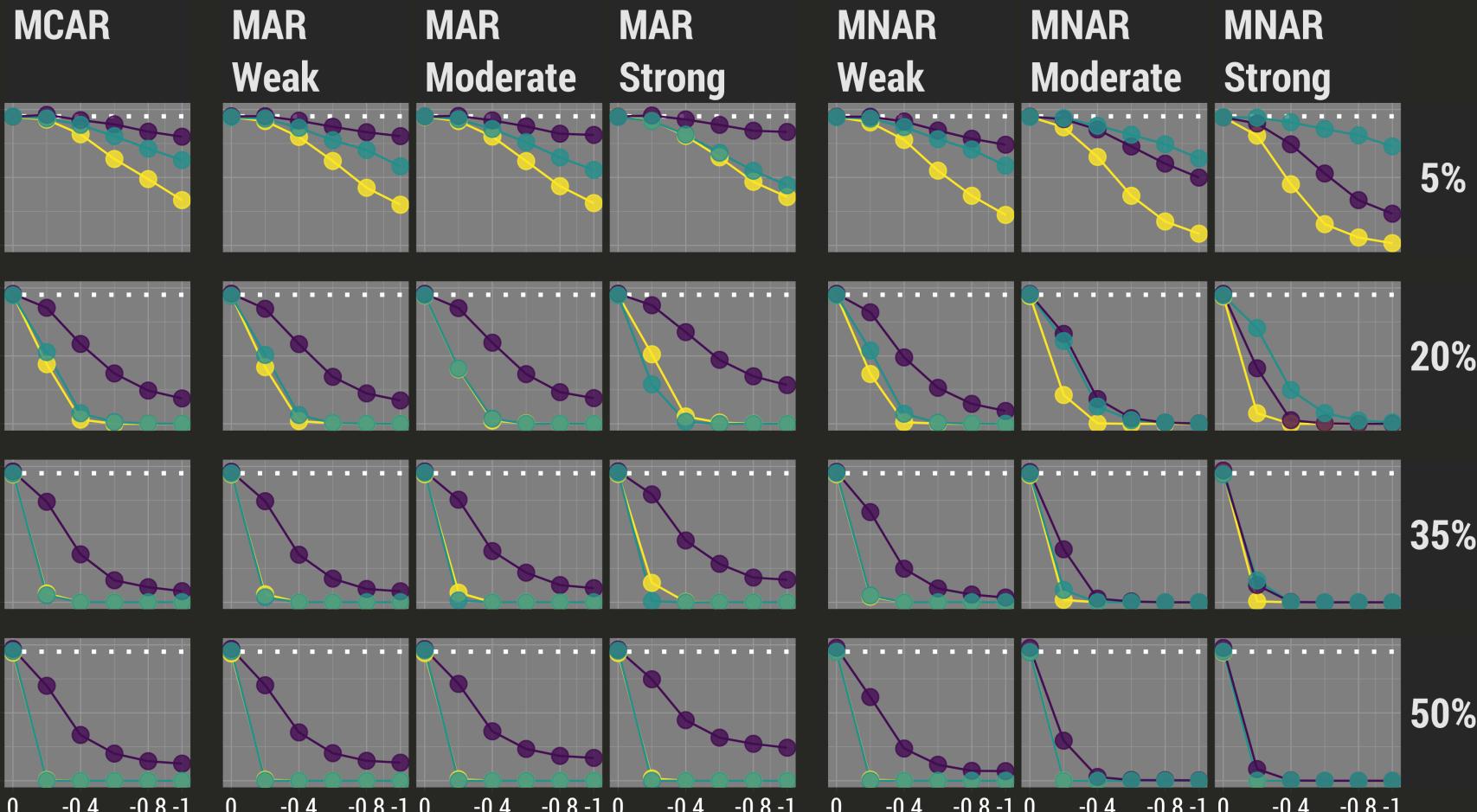
MNAR  
Strong



True Effect Size

# COVERAGE

 NA = unexposed  NA = exposed  Impute daily



## True Effect Size

# POWER

 NA = unexposed  NA = exposed  Impute daily

