MULTIVERSE ANALYSES

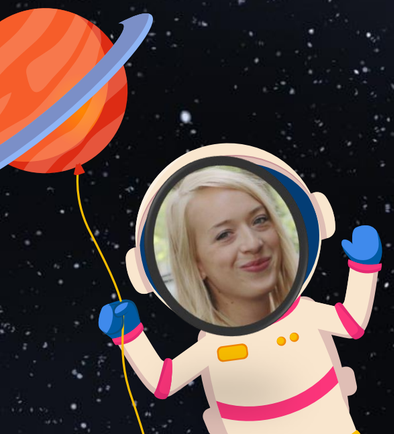
ACROSS THE RESEARCH PIPELINE

Leonie Cloos

Department of Quantitative Psychology and Individual Differences

KU LEUVEN

THE MULTIVERSE IS EVERYWHERE



Measurement, Experience Sampling, Affect Theories…
What do I have to do with multiverses?

DEGREES OF FREEDOM IN THE RESEARCH PIPELINE

STUDY DESIGN

WHAT MEASURE SHOULD I USE?

HOW OFTEN SHOULD I MEASURE?

DATA COLLECTION

SHOULD MORE DATA BE COLLECTED?

SHOULD I EXCLUDE PARTICIPANTS?

DATA PROCESSING

HOW SHOULD I TRANSFORM VARIABLES?

HOW DO I HANDLE MISSING DATA

DATA ANALYSIS

WHAT CONTROL VARIABLES TO USE?

ALLOW FOR RANDOM EFFECTS?

THE MULTIVERSE

Not to be confused with METAVERSE!

Philosophy/Physics

The idea that **multiple diverse universes exist**, beyond the observable universe.

Multiverse Analysis

In every analysis there are **multiple decisions** that have to be taken about the data.

This gives rise to **multiple** alternative – or a **multiverse** – of data sets

The Garden of Forking Paths

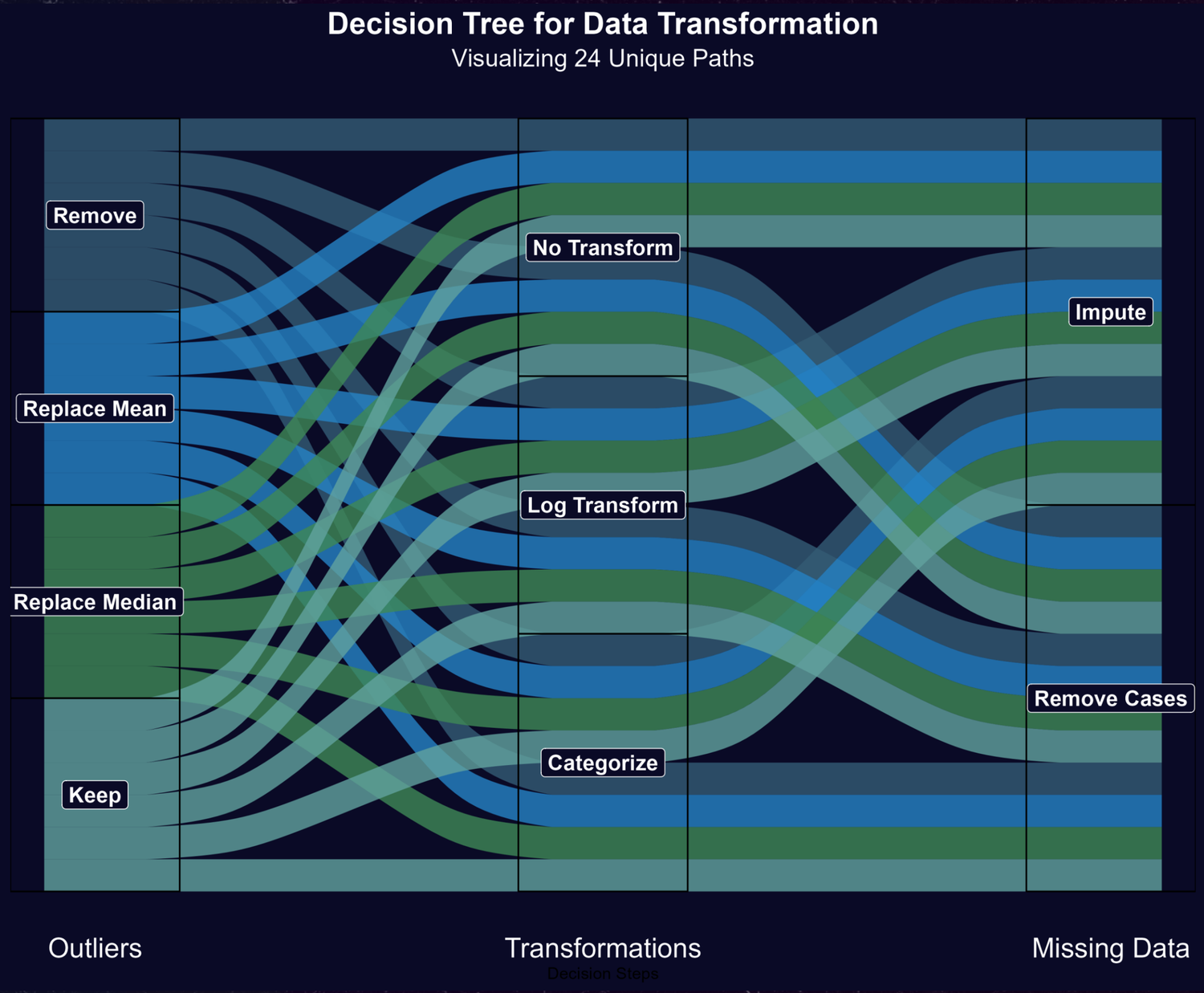
How much data should be collected?

Should specific measures be combined or transformed?

Should some observations be excluded?

Exploiting decisions increases likelihood of false positive results

Restricting yourself to an arbitrary set of choices leads a arbitrary data set, and an arbitrary statistical result



Simmons, Nelson, & Simonsohn, 2011

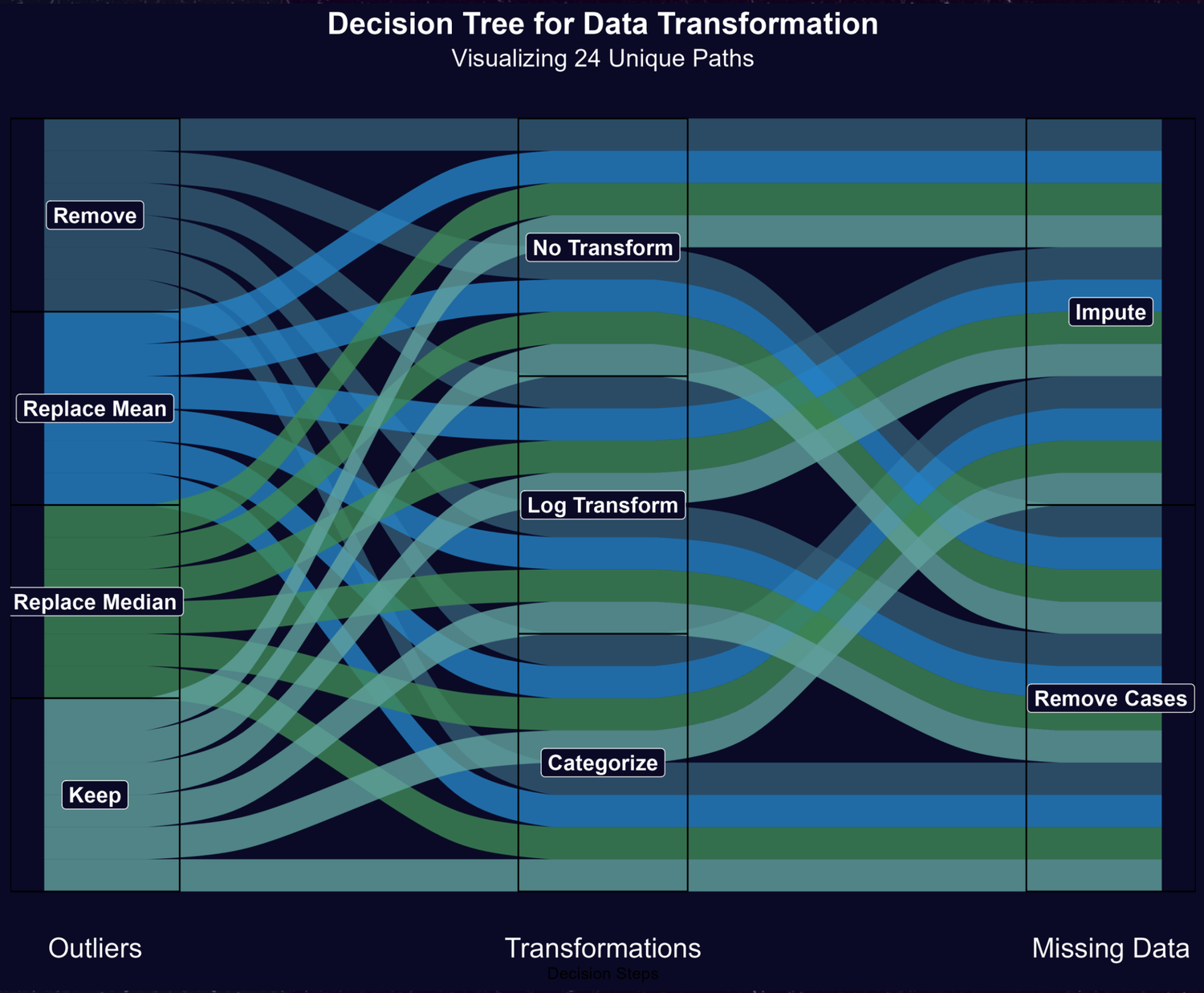
What does it look like?

- Researchers typically report one analysis
  → Replace outliers with mean value
  → Log transform skewed variable
  → Remove missing data casewise

- The decision affects the results

- **How can you be convinced it is not a false positive?**

- Instead: Report all combinations of **reasonable alternatives**



Simmons, Nelson, & Simonsohn, 2011

PROMISES & PITFALLS

TRANSPARENCY

Provide insight into decision process

ROBUSTNESS CHECK

Is the effect the same across arbitrary decisions

IMPACT OF CHOICES

Understand which choices have implications for conclusions, and improve future studies

NON-ARBITRARINESS

Not all decisions are equal. Selecting different covariate can answer a different question

ONLY ONE RAW DATA SOURCE

Only address decisions at the data cleaning and analysis stage. But decisions that affect results can happen before.

Harder, 2020; Del Giudice & Gangestad, 2021

DEGREES OF FREEDOM IN THE RESEARCH PIPELINE

STUDY DESIGN

WHAT MEASURE SHOULD I USE?

HOW OFTEN SHOULD I MEASURE?

DATA COLLECTION

SHOULD MORE DATA BE COLLECTED?

SHOULD I EXCLUDE PARTICIPANTS?

DATA PROCESSING

HOW SHOULD I TRANSFORM VARIABLES?

HOW DO I HANDLE MISSING DATA

DATA ANALYSIS

WHAT CONTROL VARIABLES TO USE?

ALLOW FOR RANDOM EFFECTS?

1 EXAMPLE
Preprocessing Choices

- **Re-analysis Durante et al., 2013:** "Effect of fertility on religiosity and political attitudes"

  Study 1: 205 participants; Study 2: 502

  → Women in a relationship are more religious when fertile than single women

Relationship status – What is defined as single?

(1) not dating/romantically involved with anyone

(2) dating or involved with only one partner

(3) engaged or living with my partner

(4) married

Single

Relationship

Steegen et al., 2016

1 EXAMPLE

Preprocessing Choices

- **Re-analysis Durante et al., 2013:** “Effect of fertility on religiosity and political attitudes”

  Study 1: 205 participants; Study 2: 502

  → Women in a relationship are more religious when fertile than single women

Relationship status – What is defined as single?

(1) not dating/romantically involved with anyone

(2) dating or involved with only one partner

(3) engaged or living with my partner

(4) married

Fertility – how do you calculate it?

cycle length → next menstrual onset → cycle day

→ **Fertile or not?**

Day 6-14 (Durante et al, 2012)

Day 7-14 (Durante et al, 2014)

Day 9-17 (Durante & Arsena, 2015) …

Steegen et al., 2016

1 EXAMPLE

Preprocessing Choices

- **Variables:**
  - **Religiosity** – 3 item scale
  - **Fertility** – high/low
    → cycle length (e.g. < 14 vs > 17; < 17 vs > 18…) = 5 alternatives
    → cycle onset = 2 alternatives
- **Relationship Status** – single / not single
  → ambiguous item “dating or involved”
  = 3 alternatives
- **Exclusions**
  → based on reported/estimated cycle length = 3 alternatives
  → based on uncertainty of cycle = 2 alternatives

5 × 2 × 3 × 3 × 2 = 180 combinations

after inconsistencies were excluded - 120

Steegen et al., 2016

1 Preprocessing Choices

Multiverse Analysis

This app shows how different choices in constructing the data leads to different analysis results.

Dependent Variable ?

Religiosity Study 1

Next menstrual onset ?

NMO1

Relationship status ?

R1

Fertility ?

F1

Exclusion of women based on cycle length ?

ECL1

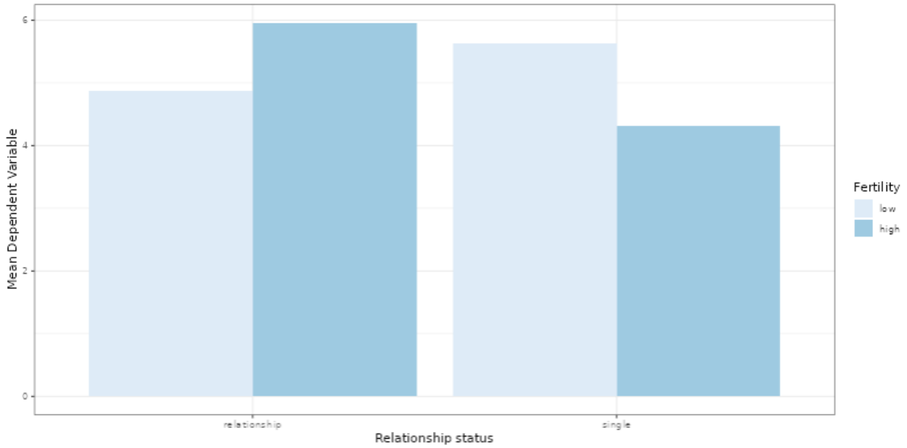
Exclusion of women based on certainty ratings ?

EC1

General info | Single data set | Single data set analysis | Multiverse

Graphical summary

The barplot represent the dependency between the chosen dependent variable and the relationship status and fertility status (n= 163)

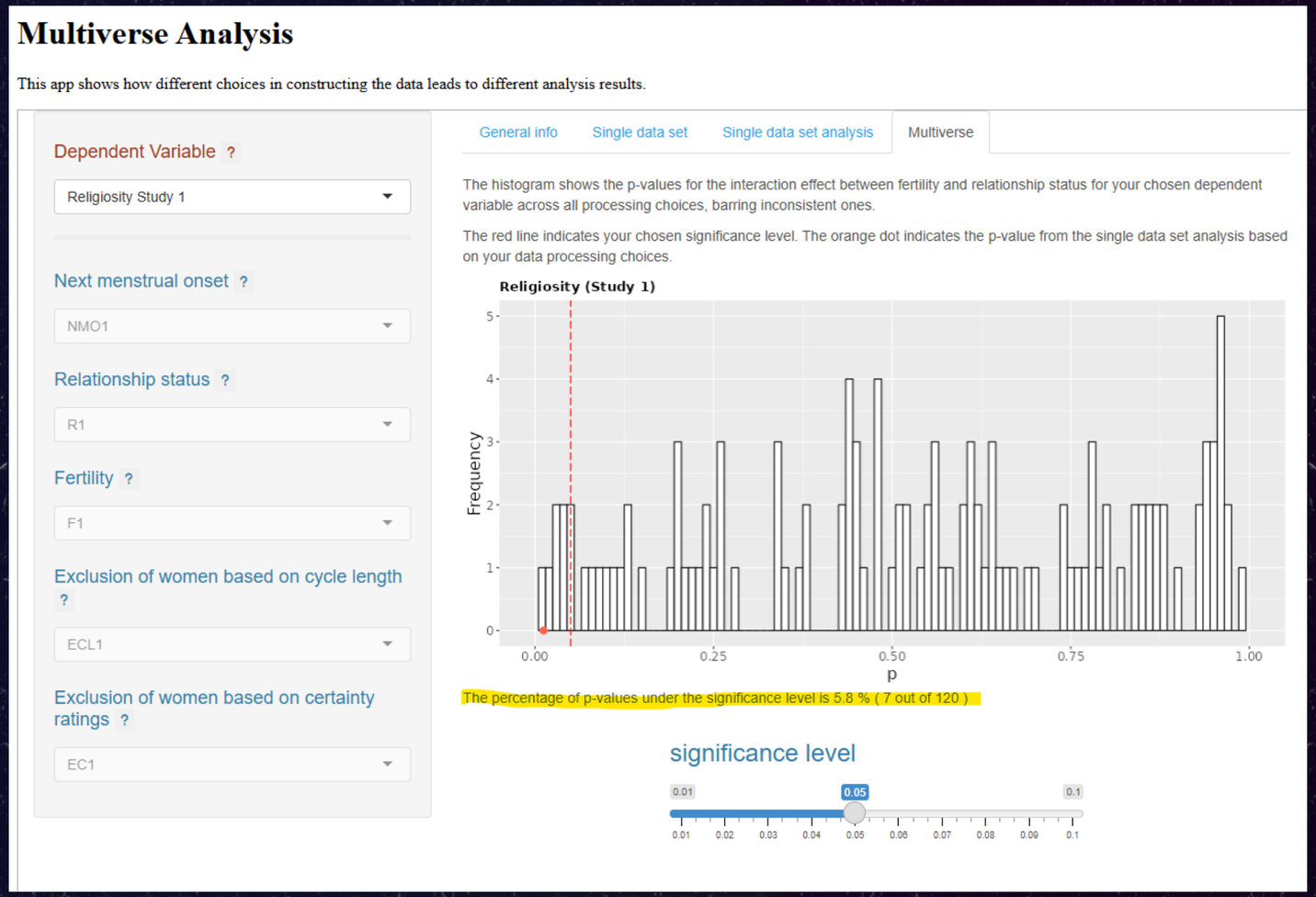


ANOVA results

Effects	F-value	DF	SS	MS	p-value
Relationship	0.746	1	6.633	6.633	0.389
Fertility	0.061	1	0.543	0.543	0.805
Interaction	6.46	1	57.408	57.408	0.012

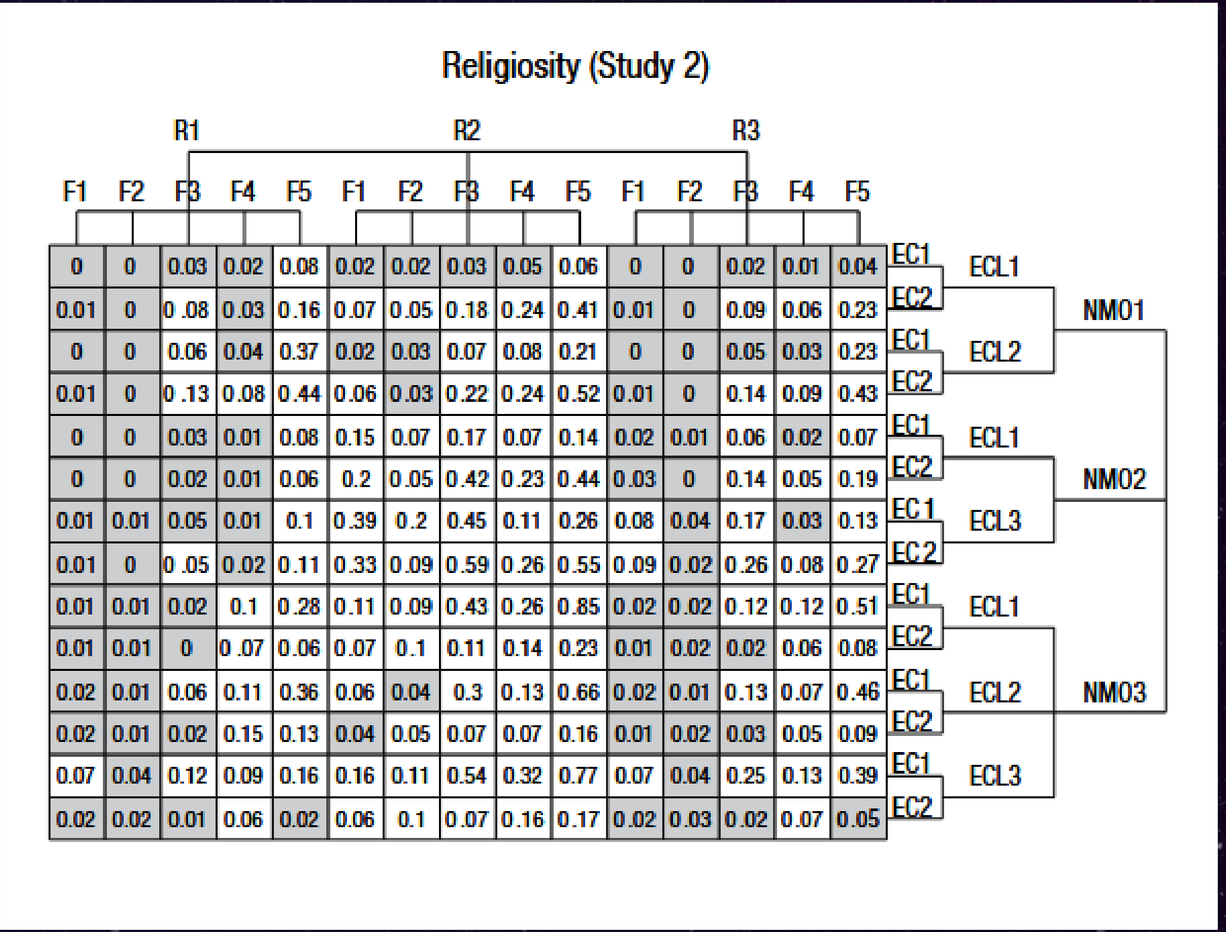
https://r.tquant.eu/KULeuven/Multiverse/

1 Preprocessing Choices



https://r.tquant.eu/KULeuven/Multiverse/

1 Preprocessing Choices



Steegen et al., 2016

2 EXAMPLE

ESM Preprocessing

- Experience sampling data of positive and negative affect
- H1: Negative Affect is elevated in individuals with psychosis
- H2: Negative Affect is related to stress – more so in individuals with psychosis
- H3: Emotional inertia is elevated in individuals with psychosis

Weermeijer et al., 2022

2 EXAMPLE

ESM Preprocessing

- Variables:
  - Exclusion of persons based on compliance
    →5%-50% in steps of 5
    = 11 alternatives
  - Exclusion of first day
    → yes no
    = 2 alternatives
- Sum score Negative affect items
  → mean, mode or max
  = 3 alternatives

11*2*3 = 66 datasets

401 – 456 participants

Weermeijer et al., 2022

2 ESM Preprocessing

H1: Negative Affect is elevated in individuals with psychosis

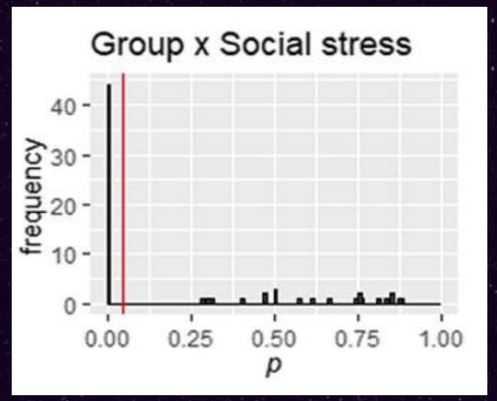
→ significant across all datasets

H2: Negative Affect is related to stress – more so in psychosis

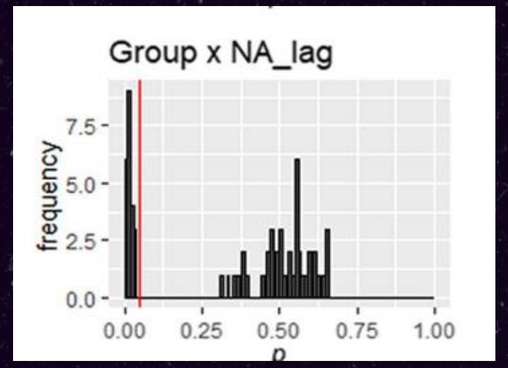
→ the interaction effect is **not significant** in datasets using **NA mean**

H3: Emotional inertia is elevated in individuals with psychosis

→ the interaction effect is **only significant** in datasets using NA mean



66% sign.



33% sign.

3 EXAMPLE

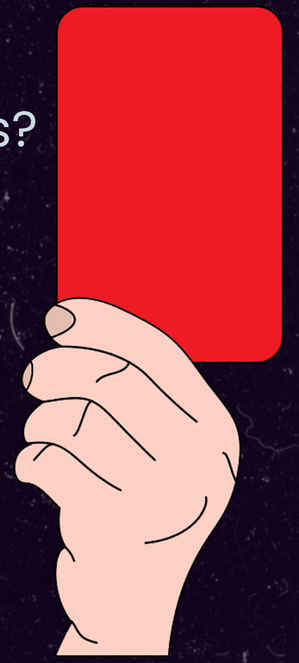
Data Analysis Choices

Same data many analysts Silberzahn et al., 2018 (outsourced multiverse)

29 Research Teams got data and a research question

Are referees more likely to give red cards to dark skinned than light skinned players?

- Would you treat each observation independent?
- How would you cluster observations?
- Would you control for player position?



Silberzahn et al., 2018

3 EXAMPLE

Data Analysis Choices

- Variables:

  → players skin tone (1 = very light; 5 = very dark)

  → number of red cards

  → Covariates (e.g.) player position, number of games played

- Analyses differences

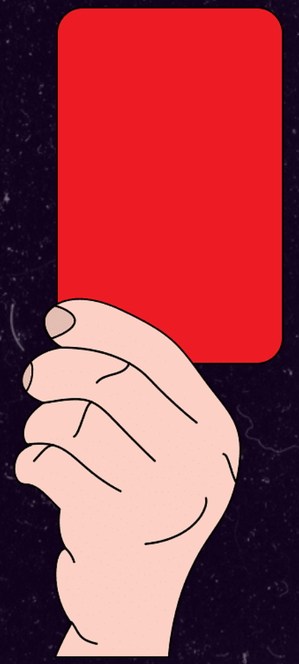
  → Treatment of nonindependence

  → Number of covariates

  → Statistical Model

29 teams = 29 analyses

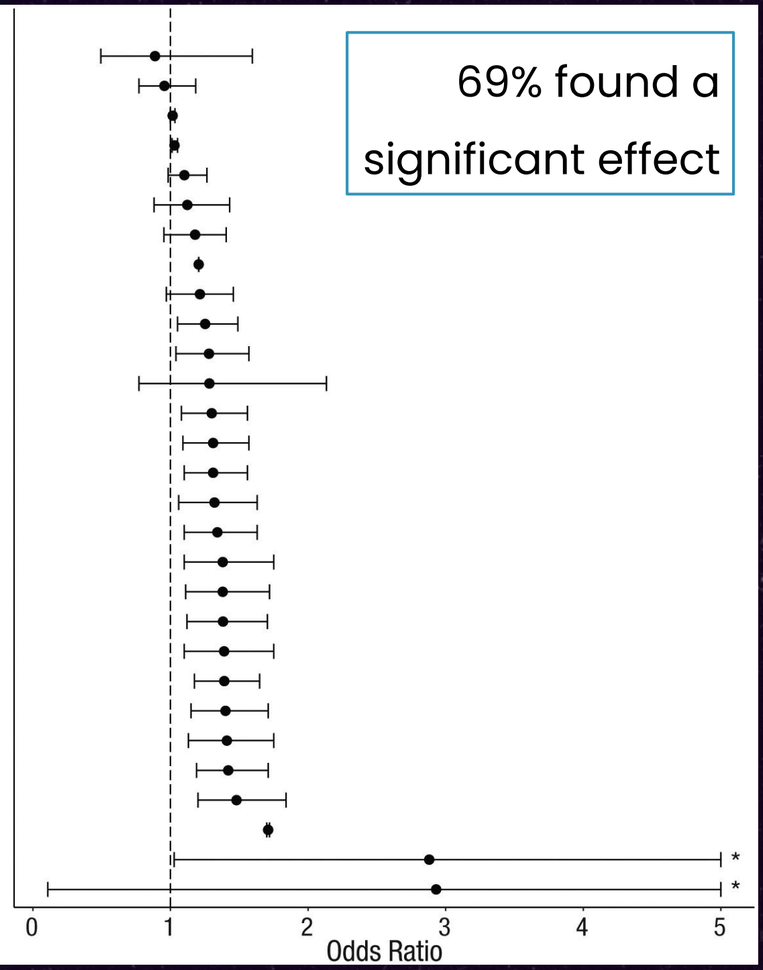
21 unique covariate combinations



Silberzahn et al., 2018

3 Data Analysis Choices

Team	Analytic Approach	Odds Ratio
12	Zero-Inflated Poisson Regression	0.89
17	Bayesian Logistic Regression	0.96
15	Hierarchical Log-Linear Modeling	1.02
10	Multilevel Regression and Logistic Regression	1.03
18	Hierarchical Bayes Model	1.10
31	Logistic Regression	1.12
1	OLS Regression With Robust Standard Errors, Logistic Regression	1.18
4	Spearman Correlation	1.21
14	WLS Regression With Clustered Standard Errors	1.21
11	Multiple Linear Regression	1.25
30	Clustered Robust Binomial Logistic Regression	1.28
6	Linear Probability Model	1.28
26	Hierarchical Generalized Linear Modeling With Poisson Sampling	1.30
3	Multilevel Logistic Regression Using Bayesian Inference	1.31
23	Mixed-Model Logistic Regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear Probability Model, Logistic Regression	1.34
5	Generalized Linear Mixed Models	1.38
24	Multilevel Logistic Regression	1.38
28	Mixed-Effects Logistic Regression	1.38
32	Generalized Linear Models for Binary Data	1.39
8	Negative Binomial Regression With a Log Link	1.39
20	Cross-Classified Multilevel Negative Binomial Model	1.40
13	Poisson Multilevel Modeling	1.41
25	Multilevel Logistic Binomial Regression	1.42
9	Generalized Linear Mixed-Effects Models With a Logit Link	1.48
7	Dirichlet-Process Bayesian Clustering	1.71
21	Tobit Regression	2.88
27	Poisson Regression	2.93



doi/10.1177/2515245917747646

APPLIED EXAMPLE

Data: Mental Health & Social Contact During COVID-19 Fried et al. 2019

ID	Day	time	beepvar	Stress_1	Stress_2	Anxiety_1	Anxiety_2	Depressi...	Depressi...	Fatigue	Hunger
1	2020-03-16	2020-03-16 12:00:00	1	1	1	2	1	1	1	2	3
1	2020-03-16	2020-03-16 15:00:00	2	2	1	2	1	1	1	1	2
1	2020-03-16	2020-03-16 18:00:00	3	1	1	3	1	1	1	2	3
1	2020-03-16	2020-03-16 21:00:00	4	1	1	3	2	1	1	2	2
1	2020-03-17	2020-03-17 12:00:00	1	1	1	2	1	1	1	2	3
1	2020-03-17	2020-03-17 15:00:00	2	3	2	2	1	1	1	1	2
2	2020-03-29	2020-03-29 12:00:00	1	2	1	2	2	1	1	3	1
2	2020-03-29	2020-03-29 15:00:00	2	2	1	2	2	1	1	3	1
2	2020-03-29	2020-03-29 18:00:00	3	Invalid Number	Invalid Number	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...
2	2020-03-29	2020-03-29 21:00:00	4	2	2	1	1	1	1	3	3
3	2020-03-16	2020-03-16 12:00:00	1	Invalid Number	Invalid Number	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...	Invalid Num...
3	2020-03-16	2020-03-16 15:00:00	2	2	1	1	1	1	1	2	1
3	2020-03-16	2020-03-16 15:01:49	2	1	2	1	1	3	2	2	3
3	2020-03-16	2020-03-16 18:00:00	3	1	1	1	1	1	1	2	4
3	2020-03-16	2020-03-16 21:00:00	4	1	1	1	1	1	1	1	1

https://github.com/LJRCloos/Multiverse-Workshop

APPLIED EXAMPLE

Data: Mental Health & Social Contact During COVID-19 Fried et al. 2019

What is the relation between negative affect and social media use?

1. Negative Affect Operationalization = 4 Alternatives
   - Stress + Anxiety + Depression + Fatigue
   - Depression
   - Stress
   - Anxiety
2. Response Time = 3 Alternatives
   - All
   - Remove > 30 minutes
   - Remove > 60 minutes
3. Compliance = 2 Alternatives
   - All
   - Remove < 50%

4*3*2 = 24 Datasets

APPLIED EXAMPLE

Download and preprocess data

Define the criteria

```
Negative Affect (4 levels)
{Composite score: (Stress + Anxiety + Depression + Fatigue)/4
Only Stress
Only Depression
Only Anxiety}

Response Time (3 levels)
{Keep all responses
Remove responses where Duration > 2700 seconds
Remove responses where Duration > 1800 seconds}

{Compliance (2 levels)
Keep all responses
Remove cases where Compliance < 75}
```

Multiverse-Workshop > Multiverse analysis.r > ...

```r
# Install necessary packages
install.packages(c("tidyverse", "dplyr", "stringr"))
library(tidyverse)
library(dplyr)
library(stringr)

# Download the data from OSF
data <- read.csv2("https://osf.io/t7g4f/download", sep = ",")

# Rename columns
data <- data |> rename_with(~c("Stress_1", "Stress_2", "Anxiety_1", "Anxiety_2", "Depression_1", "Depression_2",
                               "Fatigue", "Hunger",  "Loneliness", "Anger", "Social_Contact",
                               "Social_Media", "Music", "Procrastination", "Ourdoors", "OccupiedCovid",
                               "HealthCovid", "IsolationCovid"),
                               .cols = starts_with("Q"))

# Transform Duration column to numeric, Number ID, order columns
data <- data |> mutate(Duration = case_when(Duration == "Expired" ~ NA_character_,
                                            TRUE ~ Duration)) |>
                mutate(Duration = as.numeric(Duration)) |>
                nest(.by = ID) |>
                mutate(ID = row_number()) |>
                unnest(cols = c(data)) |>
                relocate(c("ID", "Day", "time", "beepvar"))

# Compute index for Stress, Anxiety, and Depression
data <- data |> mutate(Stress = (Stress_1 + Stress_2) / 2,
                       Anxiety = (Anxiety_1 + Anxiety_2) / 2,
                       Depression = (Depression_1 + Depression_2) / 2)

# Compute compliance per participant (e.g. number of duration is not NA)
data <- data |> group_by(ID) |>
                mutate(Compliance = sum(!is.na(Duration))/56*100) |>
                ungroup()
```

File Edit Selection View Go Run Terminal Help

Multiverse analysis

Multiverse analysis.r 1, M

Multiverse-Workshop > Multiverse analysis.r > ...

```r
# Install necessary packages
install.packages(c("tidyverse", "dplyr", "stringr"))
library(tidyverse)
library(dplyr)
library(stringr)

# Download the data from OSF
data <- read.csv2("https://osf.io/t7g4f/download", sep = ",")

# Rename columns
data <- data |> rename_with(~c("Stress_1", "Stress_2", "Anxiety_1", "Anxiety_2", "Depression_1", "Depression_2",
                               "Fatigue", "Hunger",  "Loneliness", "Anger", "Social_Contact",
                               "Social_Media", "Music", "Procrastination", "Ourdoors", "OccupiedCovid",
                               "HealthCovid", "IsolationCovid"),
                               .cols = starts_with("Q"))

# Transform Duration column to numeric, Number ID, order columns
data <- data |> mutate(Duration = case_when(Duration == "Expired" ~ NA_character_,
                                            TRUE ~ Duration)) |>
                mutate(Duration = as.numeric(Duration)) |>
                nest(.by = ID) |>
                mutate(ID = row_number()) |>
                unnest(cols = c(data)) |>
                relocate(c("ID", "Day", "time", "beepvar"))

# Compute index for Stress, Anxiety, and Depression
data <- data |> mutate(Stress = (Stress_1 + Stress_2) / 2,
                       Anxiety = (Anxiety_1 + Anxiety_2) / 2,
                       Depression = (Depression_1 + Depression_2) / 2)

# Compute compliance per participant (e.g. number of duration is not NA)
data <- data |> group_by(ID) |>
                mutate(Compliance = sum(!is.na(Duration))/56*100) |>
                ungroup()

## Make 24 Datasets

```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS COMMENTS

```
+                               .cols = starts_with("Q"))
> # Transform Duration column to numeric, Number ID, order columns
> data <- data |> mutate(Duration = case_when(Duration == "Expired" ~ NA_chara$
+                                             TRUE ~ Duration)) |>
+                 mutate(Duration = as.numeric(Duration)) |>
+                 nest(.by = ID) |>
+                 mutate(ID = row_number()) |>
+                 unnest(cols = c(data)) |>
+                 relocate(c("ID", "Day", "time", "beepvar"))
Warning message:
There was 1 warning in `mutate()`.
i In argument: `Duration = as.numeric(Duration)`.
Caused by warning:
! NAs introduced by coercion
> # Compute index for Stress, Anxiety, and Depression
> data <- data |> mutate(Stress = (Stress_1 + Stress_2) / 2,
+                        Anxiety = (Anxiety_1 + Anxiety_2) / 2,
+                        Depression = (Depression_1 + Depression_2) / 2)
> # Compute compliance per participant (e.g. number of duration is not NA)
> data <- data |> group_by(ID) |>
+                 mutate(Compliance = sum(!is.na(Duration))/56*100) |>
+                 ungroup()
> ## Make 24 Datasets
>
```

powershell

R Interactive

Ask Copilot

Copilot is powered by AI, so mistakes are possible. Review output carefully before use.

or type # to attach context

@ to chat with extensions

Type / to use commands

/help What can you do?

Ask Copilot

Multiverse analysis.r Current file

00:01:07

main* 1 0

R 4.4.3: 9872 Ln 38, Col 1 Spaces: 4 UTF-8 CRLF R

14:42
6/03/2025

APPLIED EXAMPLE

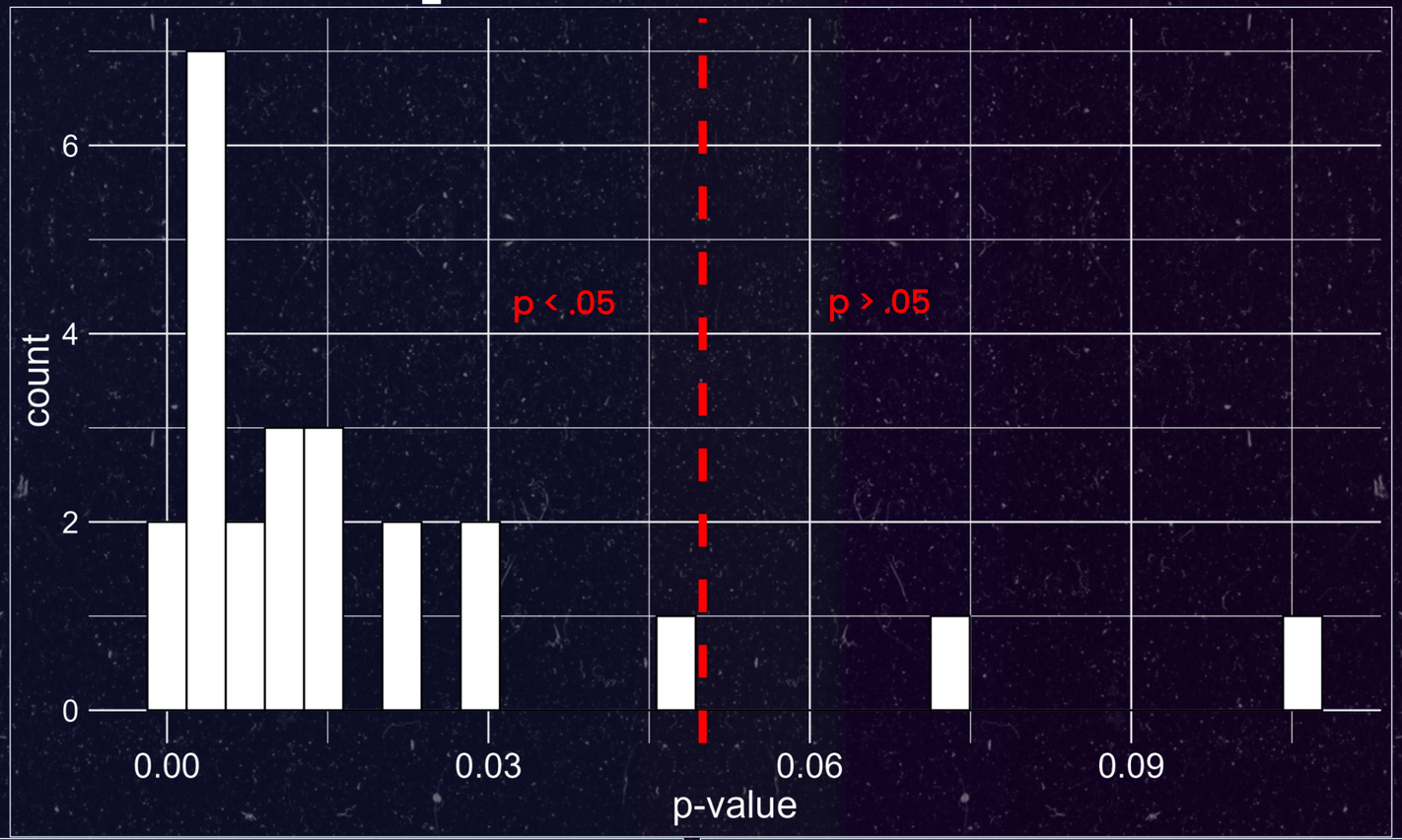
New prompt

```
82
83  # On each dataset run the analyses: relationship between negative affect and
    for (dataset_name in names(datasets)) {
      dataset <- datasets[[dataset_name]]

      # Compute correlation between negative affect and social media use
      correlation <- cor.test(dataset$Negative_Affect, dataset$Social_Media, met

      # Print the results
      cat("Dataset:", dataset_name, "\n")
      cat("Correlation:", correlation$estimate, "\n")
      cat("p-value:", correlation$p.value, "\n\n")
    }
```

Not entirely correct

```r
# On each dataset run the analyses: relationship between negative affect and social media use
library(lme4)
library(broom.mixed)
library(lmerTest)

# Save the results
results <- expand_grid(dataset_name = names(datasets), coefficient = c("Intercept", "Social Media")) |>
  mutate(estimate = NA_real_, std.error = NA_real_, p.value = NA_real_, conf.low = NA_real_, conf.high = NA_real_)

# Run the analysis on each dataset
for (dataset_name in names(datasets)) {
  dataset <- datasets[[dataset_name]]

  # Compute correlation between negative affect and social media use
  model <- lmer(Negative_Affect ~ Social_Media + (1 + Social_Media|ID),
                REML=FALSE, lmerControl(optCtrl=list(maxit=100L)),
                data = dataset)

  # if model doesnt converge print error message
  if (is.null(model@optinfo$conv$lme4$warnings)) {
      print(paste("Model converged for", dataset_name))
      # Extract the fixed effects results
      fixed_effects <- tidy(model, effects = "fixed", conf.int = TRUE)

      # Update the results dataframe
      results <- results |>
        mutate(estimate = ifelse(dataset_name == !!dataset_name & coefficient == "Intercept", fixed_effects$estimate[1], estimate),
           std.error = ifelse(dataset_name == !!dataset_name & coefficient == "Intercept", fixed_effects$std.error[1], std.error),
           p.value = ifelse(dataset_name == !!dataset_name & coefficient == "Intercept", fixed_effects$p.value[1], p.value),
           conf.low = ifelse(dataset_name == !!dataset_name & coefficient == "Intercept", fixed_effects$conf.low[1], conf.low),
           conf.high = ifelse(dataset_name == !!dataset_name & coefficient == "Intercept", fixed_effects$conf.high[1], conf.high),
           estimate = ifelse(dataset_name == !!dataset_name & coefficient == "Social Media", fixed_effects$estimate[2], estimate),
           std.error = ifelse(dataset_name == !!dataset_name & coefficient == "Social Media", fixed_effects$std.error[2], std.error),
           p.value = ifelse(dataset_name == !!dataset_name & coefficient == "Social Media", fixed_effects$p.value[2], p.value),
           conf.low = ifelse(dataset_name == !!dataset_name & coefficient == "Social Media", fixed_effects$conf.low[2], conf.low),
           conf.high = ifelse(dataset_name == !!dataset_name & coefficient == "Social Media", fixed_effects$conf.high[2], conf.high))
  } else {
      print(paste("Model did not converge for", dataset_name))
  }
}
```

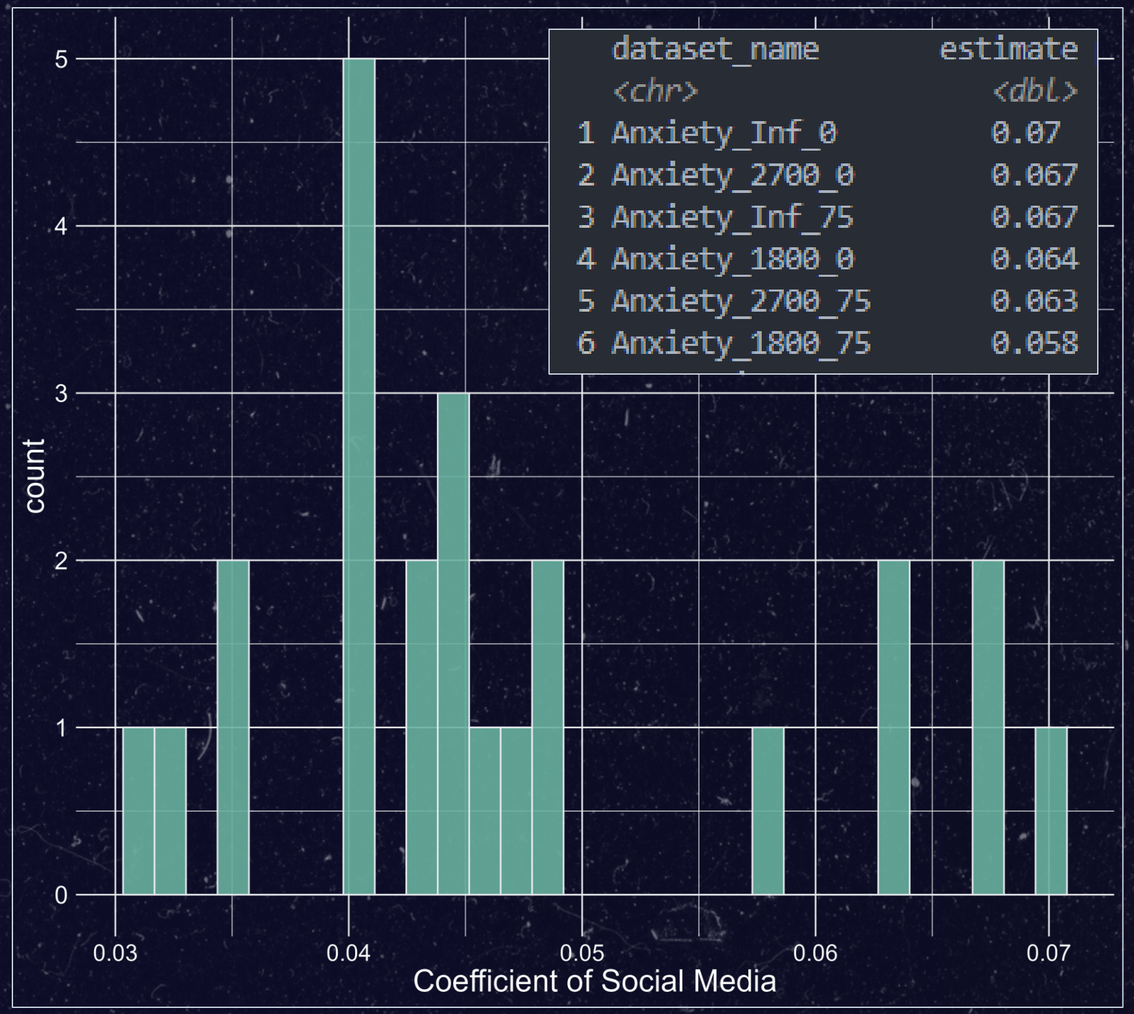
SUMMARY – p-values



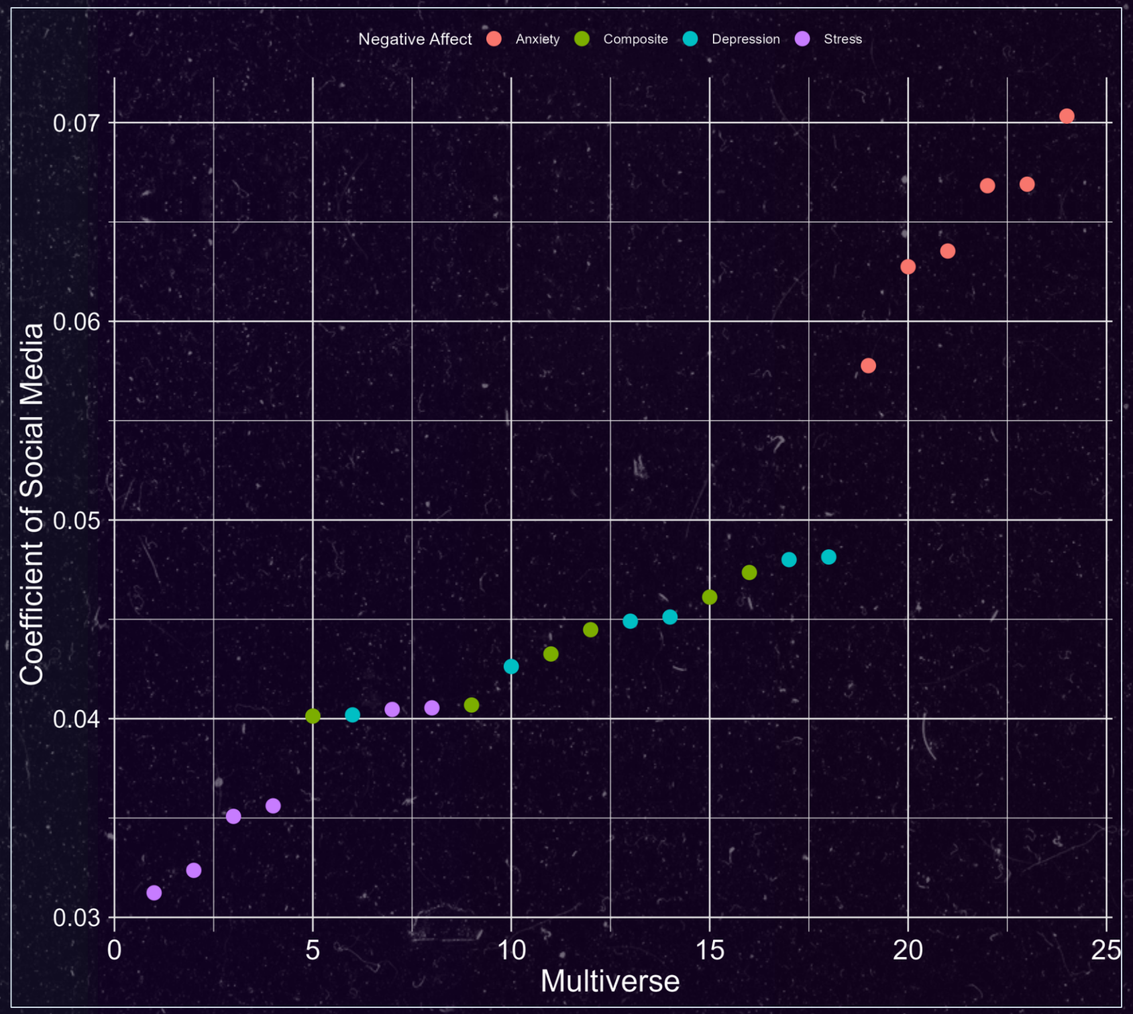
	dataset_name	estimate	p.value	conf.low	conf.high
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Anxiety_Inf_0	0.07	0.001	0.031	0.11
2	Anxiety_2700_0	0.067	0.001	0.027	0.107
3	Composite_Inf_0	0.046	0.002	0.018	0.074
4	Anxiety_Inf_75	0.067	0.002	0.025	0.108

	dataset_name	estimate	p.value	conf.low	conf.high
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Stress_1800_75	0.031	0.107	-0.007	0.069
2	Stress_1800_0	0.032	0.072	-0.003	0.068
3	Stress_2700_75	0.035	0.048	0	0.07
4	Stress_2700_0	0.036	0.031	0.003	0.068

SUMMARY - Coefficient



	dataset_name	estimate
	<chr>	<dbl>
1	Anxiety_Inf_0	0.07
2	Anxiety_2700_0	0.067
3	Anxiety_Inf_75	0.067
4	Anxiety_1800_0	0.064
5	Anxiety_2700_75	0.063
6	Anxiety_1800_75	0.058



SUMMARY

1

Operationalizing Negative Affect as Anxiety gives higher estimates

2

The overall trend is robust

3

Are the datasets equally justifiable?

DEGREES OF FREEDOM IN THE RESEARCH PIPELINE

STUDY DESIGN

WHAT MEASURE SHOULD I USE?

HOW OFTEN SHOULD I MEASURE?

DATA COLLECTION

SHOULD MORE DATA BE COLLECTED?

SHOULD I EXCLUDE PARTICIPANTS?

DATA PROCESSING

HOW SHOULD I TRANSFORM VARIABLES?

HOW DO I HANDLE MISSING DATA

DATA ANALYSIS

WHAT CONTROL VARIABLES TO USE?

ALLOW FOR RANDOM EFFECTS?

4 EXAMPLE

Measurement Decisions

Measuring Positive and Negative Affect

- Multiple items → which items to choose?
- Single items → what dimension?
- Other latent factors → intensity, arousal?

4 EXAMPLE

Measurement Decisions

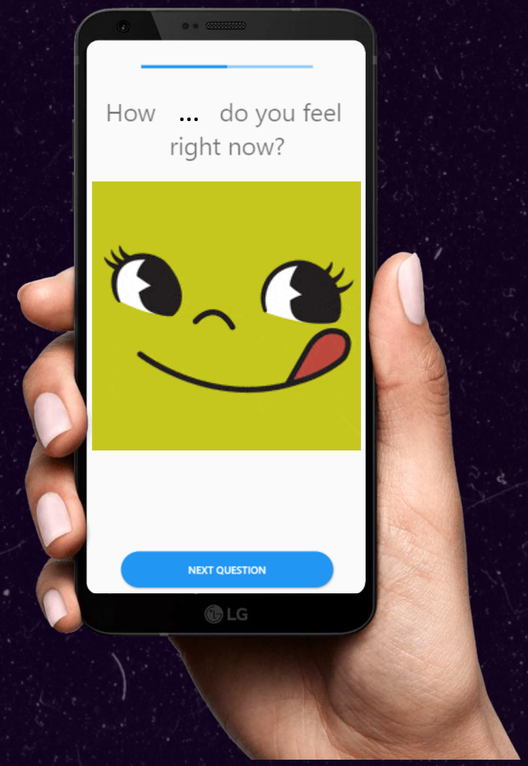
Using a Multiverse to determine the validity of multi-item measures

22 DISCRETE EMOTION ITEMS

- 8 Positive Items
  - 4 low intensity
  - 4 high intensity
  - PA Mean
  - PA Max
- 14 Negative Items
  - 7 low intensity
  - 7 high intensity
  - NA Mean
  - NA Max

6 SINGLE ITEMS

- PA Neutral
- NA Neutral
- Valence
- PA Absent
- NA Absent
- Smiley

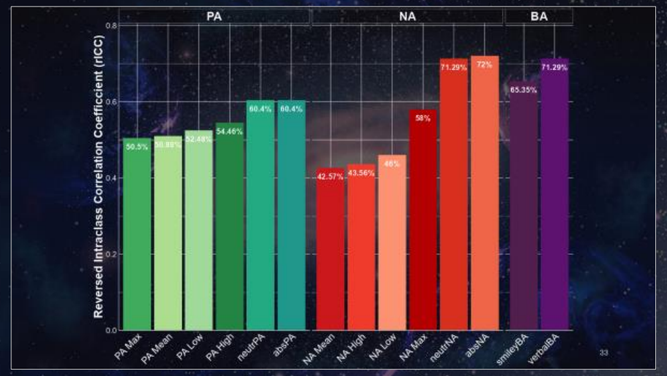


14 Measurement Operationalizations

VALIDATION

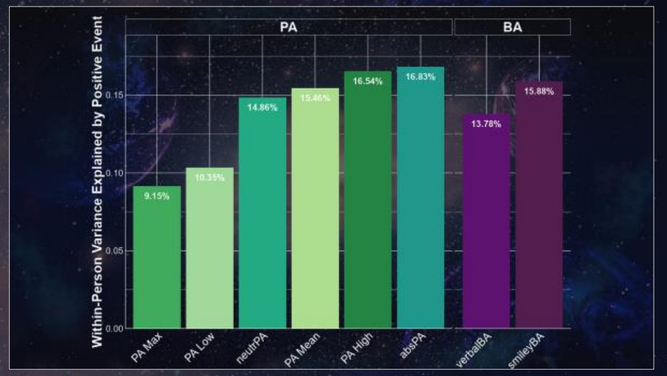
VARIANCE

Single items have higher variance



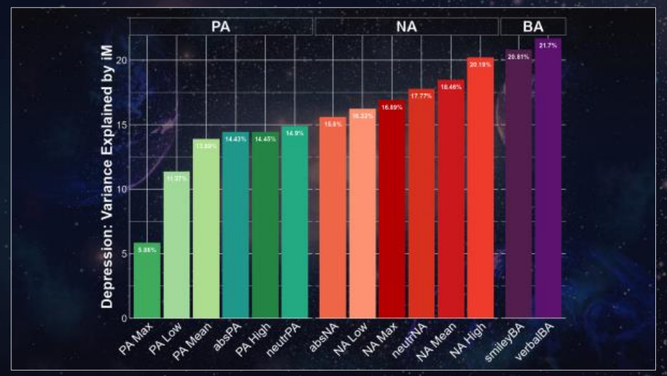
RELATIONSHIP TO EVENTS

High Intensity or Single Item



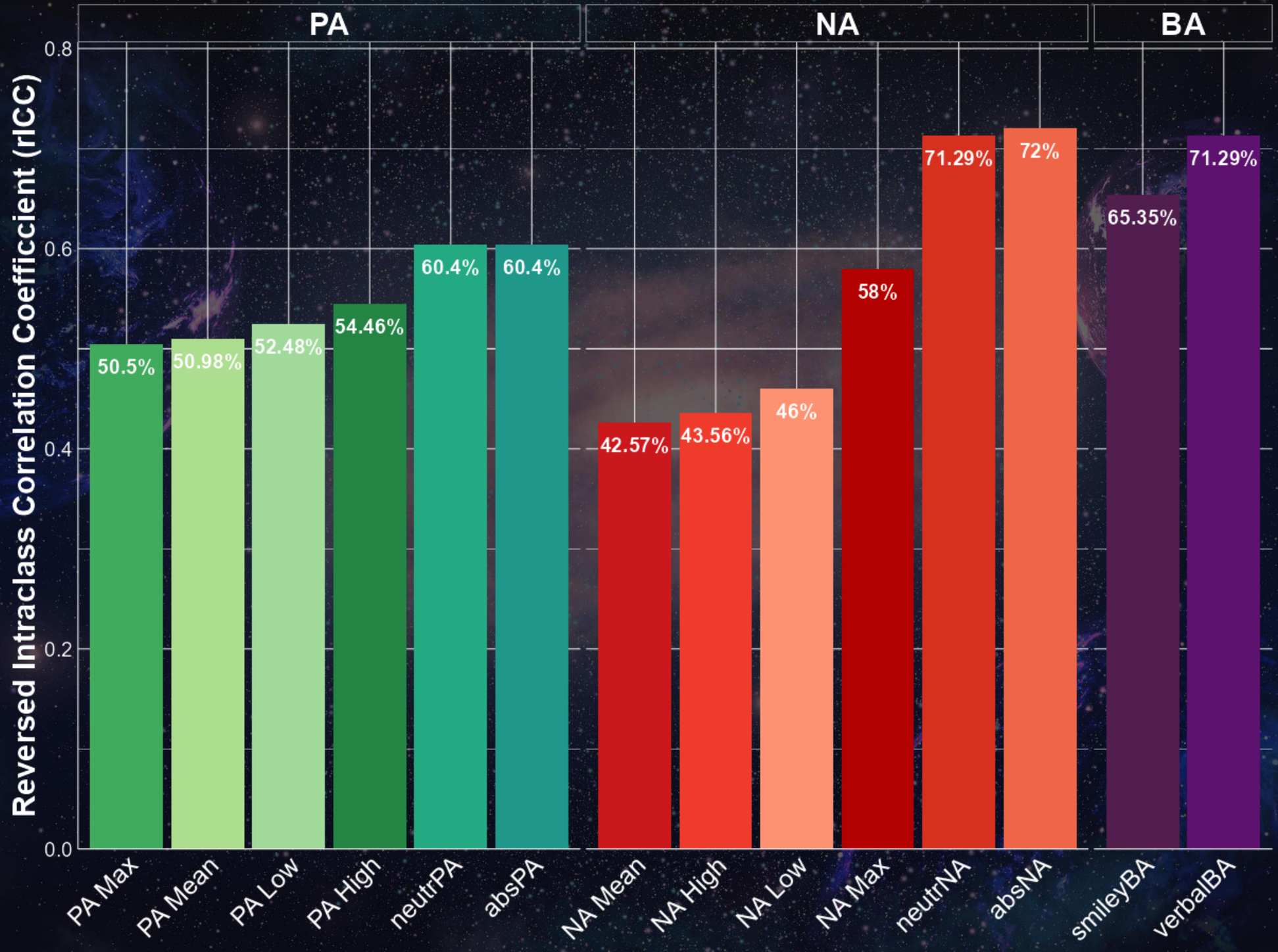
RELATIONSHIP WITH DEPRESSION

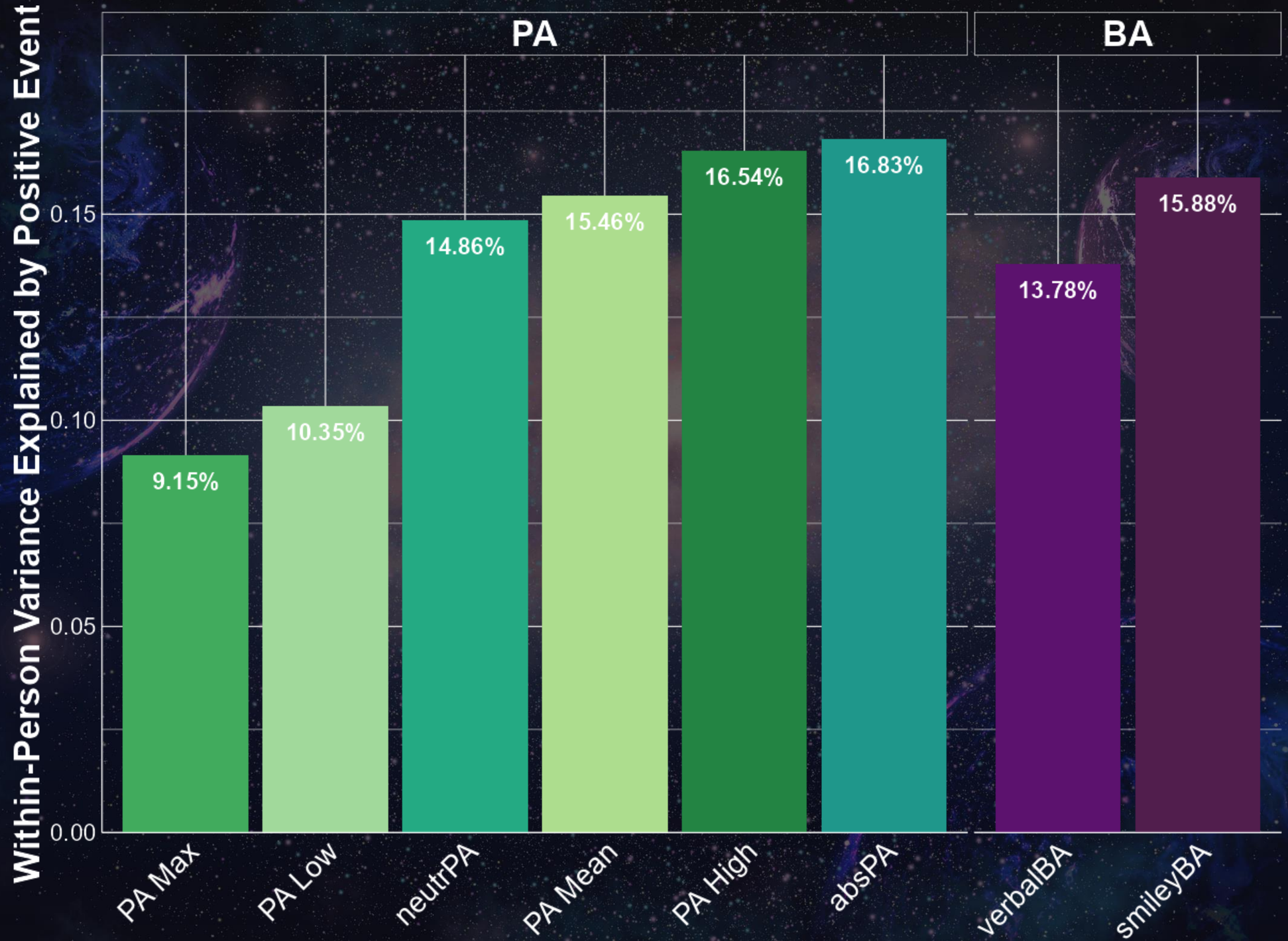
Bipolar items are best

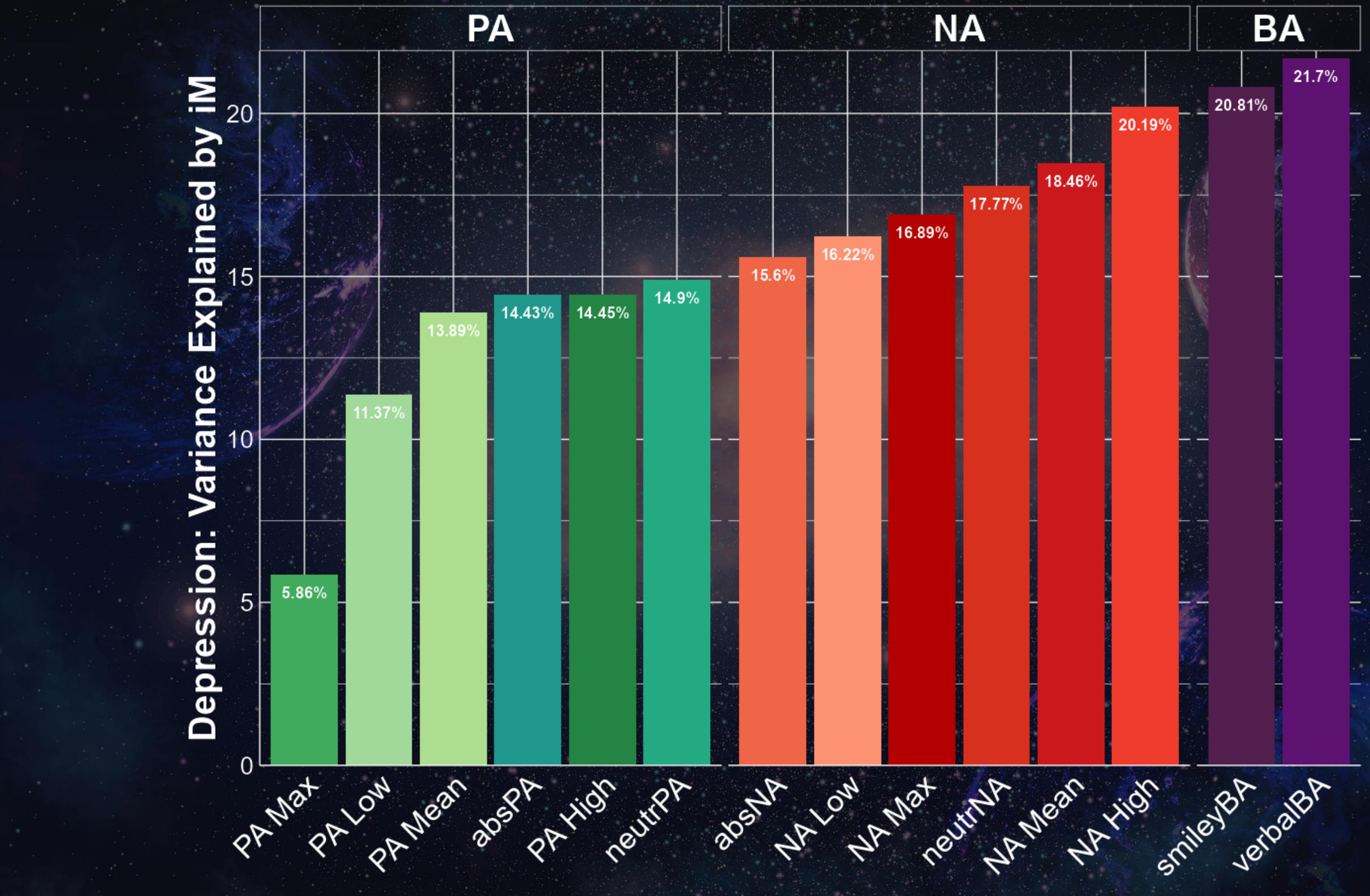


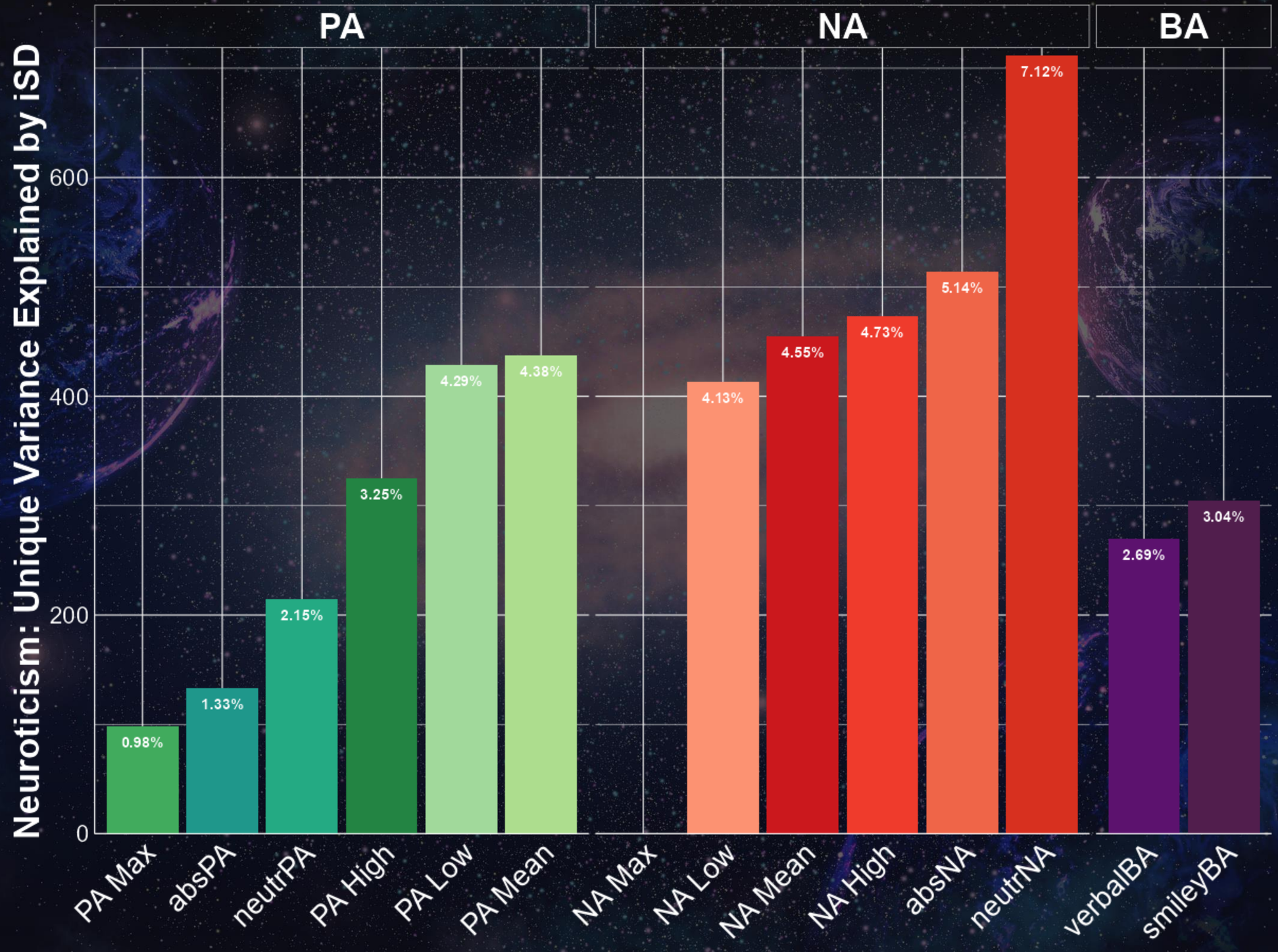
RELATIONSHIP WITH NEUROTICISM

Single items for NA are best


PA
NA
BA
Reversed Intraclass Correlation Coefficient (rICC)
0.0
0.2
0.4
0.6
0.8
50.5%
50.98%
52.48%
54.46%
60.4%
60.4%
42.57%
43.56%
46%
58%
71.29%
72%
65.35%
71.29%
PA Max
PA Mean
PA Low
PA High
neutrPA
absPA
NA Mean
NA High
NA Low
NA Max
neutrNA
absNA
smileyBA
verbalBA


PA
BA
Within-Person Variance Explained by Positive Event
0.00
0.05
0.10
0.15
9.15%
10.35%
14.86%
15.46%
16.54%
16.83%
13.78%
15.88%
PA Max
PA Low
neutrPA
PA Mean
PA High
absPA
verbalBA
smileyBA


PA
NA
BA
Depression: Variance Explained by iM
0
5
10
15
20
5.86%
11.37%
13.89%
14.43%
14.45%
14.9%
15.6%
16.22%
16.89%
17.77%
18.46%
20.19%
20.81%
21.7%
PA Max
PA Low
PA Mean
absPA
PA High
neutrPA
absNA
NA Low
NA Max
neutrNA
NA Mean
NA High
smileyBA
verbalBA


PA
NA
BA
Neuroticism: Unique Variance Explained by iSD
600
400
200
0
0.98%
1.33%
2.15%
3.25%
4.29%
4.38%
4.13%
4.55%
4.73%
5.14%
7.12%
2.69%
3.04%
PA Max
absPA
neutrPA
PA High
PA Low
PA Mean
NA Max
NA Low
NA Mean
NA High
absNA
neutrNA
verbalBA
smileyBA

5 EXAMPLE

Power Analysis

- Effect of preprocessing and measurement operationalizations on sample size
- Calculations on previous effect size estimates

  Relation between momentary affect and depression

  → Different operationalization = different effect size = different sample size estimate?

Lafit et al., 2003

5 EXAMPLE

Power Analysis

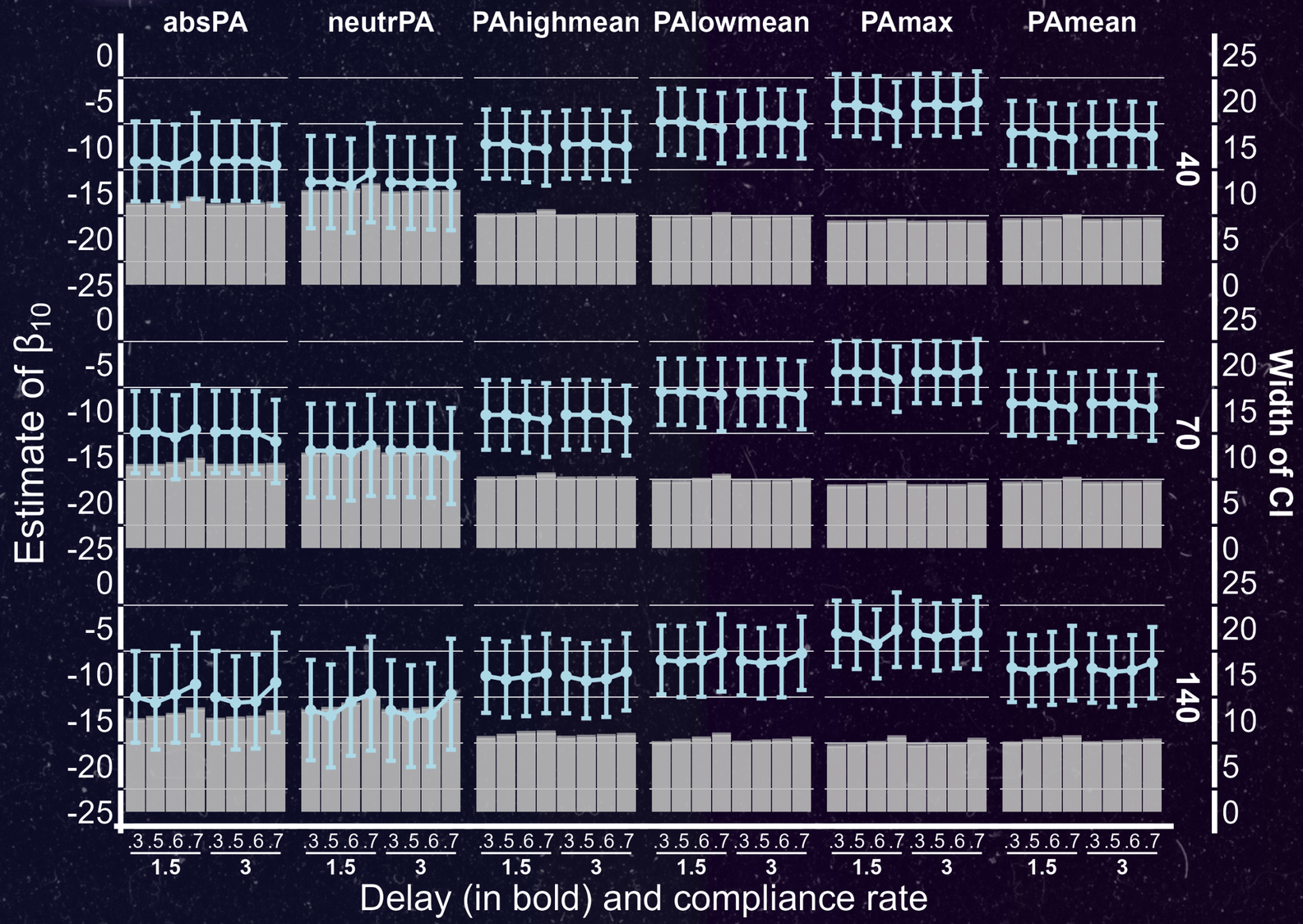
Multiverse

- Measurement operationalization
  → 4 single items (PA & NA)
  → 8 composites (PA & NA)
- Duration
  → 4 days, 1 week, 2 weeks
- Exclusion based on response delay
  → > 1.5 SD, 3 SD
- Exclusions based on compliance
  → 30%, 50%, 60% 70%

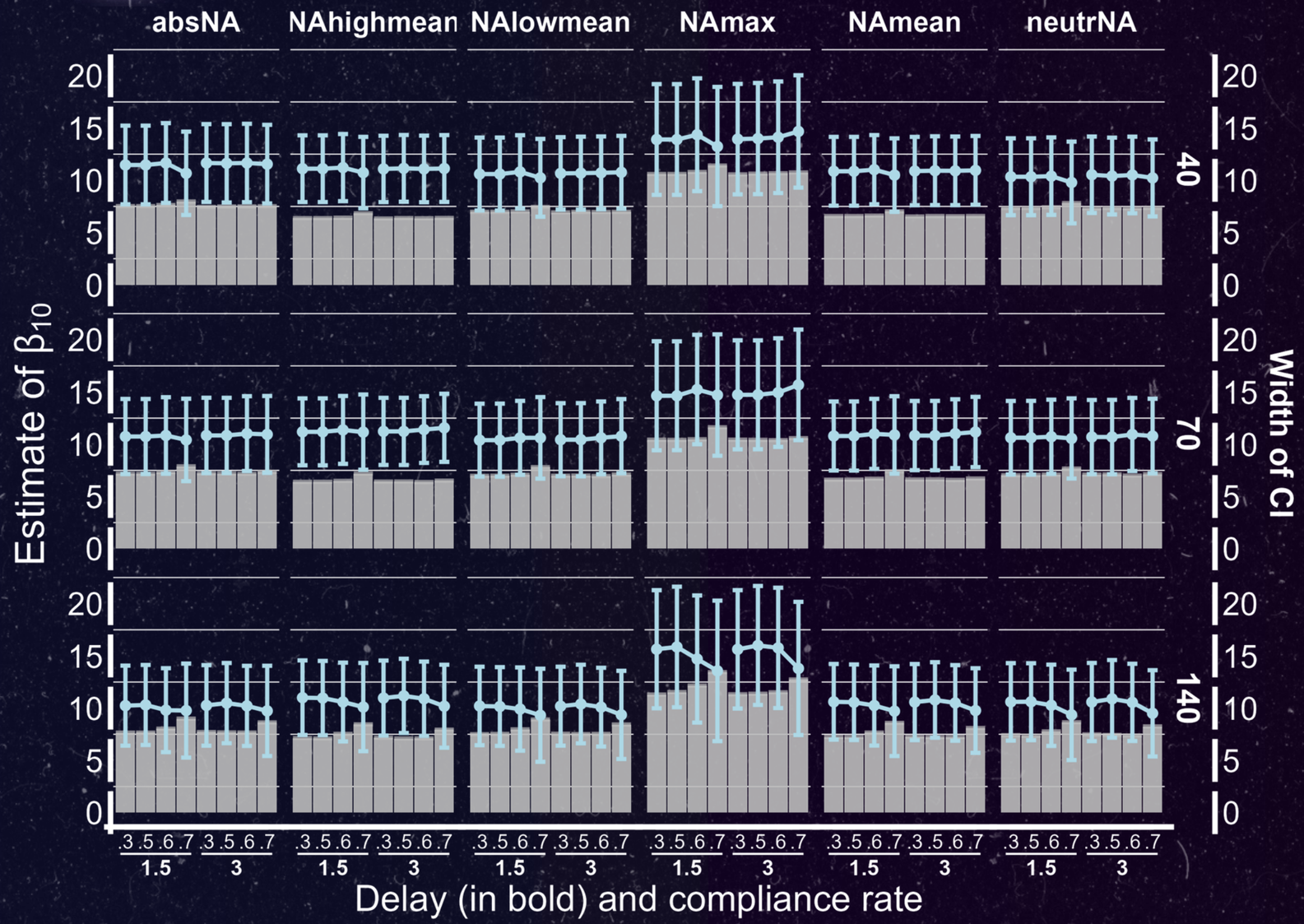
6*2*3*4 = 144 (for PA & NA each)

Lafit et al., 2003

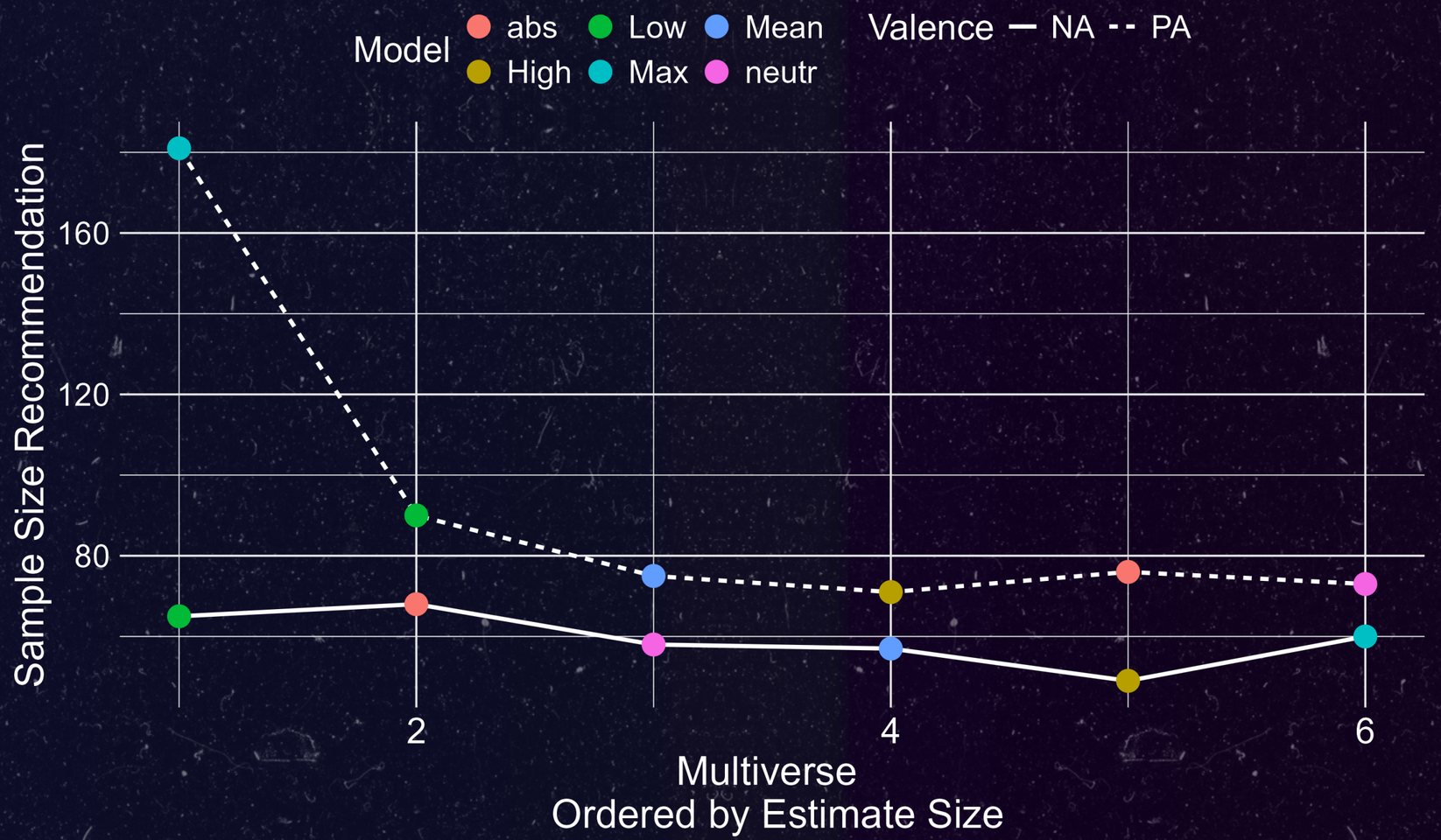
4 Power Analysis



4 Power Analysis



4 Power x Measurement



4 Power x Preprocessing

Conclusions

Provide important insights into where there is a lack of theory and standardized practices (e.g. ESM)

Multiverse analyses increase transparency and robustness of findings

Multiverse analysis with different research teams may be valuable to evaluate the outcome of different methods on the data

MORE EXAMPLES

Credé & Phillips (2017)

Power Pose Effect: 58 different datasets

Outliers, covariates, definition of DV and IV

P value only sign for original specification

Patel, Burford, & Ioannidis (2015)

Vitamin D and mortality: 8192 models

For 31% variables they observed Janus effect (1^{st} and 99^{th} percentile of effect are in opposite direction)

Simonsohn, Simmons, & Nelson (2020)

Hurricanes with feminine names cause more deaths: 1728 datasets

Only 37 are significant!

Niemeijer, Mestdagh, Kuppens (2022)

Sensor data predicting subjective sleep quality and mood: 12,600 datasets

R^2 between 7.5% – 78%

MORE EXAMPLES

control variables. Our findings indicate that the inferences regarding the presence and size of an effect on testosterone and cortisol are highly sensitive to data analytic specifications. We encourage researchers to routinely explore the influence of data analytic choices on statistical inferences and also encourage editors and reviewers to require explicit examinations of the influence of alternative data analytic specifications on the inferences that are drawn from data.

Credé & Phillips (2017)

The Multiverse of Methods: Extending the Multiverse Analysis to Address Data-Collection Decisions

Jenna A. Harder
Department of Psychology, Michigan State Ur

A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions

Marco Del Giudice and Steven W. Gangestad
Department of Psychology, University of New Mexico

Multiverse analyses in the classroom

Tom Heyman
Methodology and Statistics Unit, Institute of Psychology, Leiden University

Wolf Vanpaemel
f Psychology and Educational Sciences, KU Leuven

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect. *Psychological Assessment.* doi.org/10.1037/pas0001200

Credé, M., & Phillips, L. A. (2017). Revisiting the Power Pose Effect. *SPSP.* doi.org/10.1177/1948550617714584

Del Giudice, M., & Gangestad, S. W. (2021). A Traveler's Guide to the Multiverse. *AMPPS.* doi.org/10.1177/2515245920954925

Harder, J. (2020). The Multiverse of Methods: Extending the Multiverse Analysis to Address Data-Collection Decisions. *Perspectives on Psychological Science.* doi.org/10.1177/1745691620917678

Heyman, T., & Vanpaemel, W. (2022). Multiverse analyses in the classroom. *Meta-Psychology.* doi.org/10.15626/MP.2020.2718

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to Regulate. *Psychological Science.* doi.org/10.1177/0956797619838763

Lafit, G., Revol, J., Cloos, L., Kuppens, P., & Ceulemans, E. (2024). The effect of different operationalizations of affect and preprocessing choices on power-based sample size. *Assessment.* doi.org/10.1177/10731911241286868

Niemeijer, K., Mestdagh, M., & Kuppens, P. (2022). Tracking Subjective Sleep Quality and Mood With Mobile Sensing. *Journal of Medical Internet Research.* doi.org/10.2196/25643

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects. *Journal of Clinical Epidemiology.* doi.org/10.1016/j.jclinepi.2015.05.029

Silberzahn, R., et al (2018). Many Analysts, One Data Set. *AMPPS.* doi.org/10.1177/2515245917747646

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour.* doi.org/10.1038/s41562-020-0912-z

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science.* doi.org/10.1177/1745691616658637

Weermeijer, J., et. al. (2022). Applying multiverse analysis to experience sampling data. *BRM.* doi.org/10.3758/s13428-021-01777-1

THANK YOU

Leonie.cloos@kuleuven.be

https://github.com/LJRCloos/Multiverse-Workshop