INCREASING TRANSPARENCY THROUGH A MULTIVERSE ANALYSIS

Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel
Data are not just “collected” or “observed”

Data are constructed: translating observed raw data to processed data set ready for analysis
  - discretization
  - transformation
  - data exclusion
  - ...

In data construction there are often many researcher degrees of freedom (Simmons, Nelson & Simonsohn, 2011; Gelman & Loken, 2014): researcher has to choose among several options

Different choices can lead to different data sets: multiverse of data sets

Different data sets can lead to different statistical results: multiverse of results
Multiverse analysis

- Researchers typically consider a single processed data set: **single data set analysis**
- Problematic since choosing among data processing options is often arbitrary
  - arbitrary? when there are different reasonable options + no clear justification
- Arbitrariness is inherited by result
- Is the result robust?
  - which other reasonable results could have been obtained?
Researchers typically consider a single processed data set: **single data set analysis**

Problematic since choosing among data processing options is often arbitrary
- arbitrary? when there are different reasonable options + no clear justification

Arbitrariness is inherited by result

Is the result robust?
- which other reasonable results could have been obtained?

**Solution:** multiverse analysis
- if data processing choices are arbitrary, without a clear justification...
- ... consider all reasonable options
- perform the analysis of interest across multiverse of data sets
- summarize results across multiverse
The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle

Kristina M. Durante¹, Ashley Rae¹, and Vladas Griskevicius²

¹College of Business, University of Texas, San Antonio, and ²Carlson School of Management, University of Minnesota

Abstract

Each month, many women experience an ovulatory cycle that regulates fertility. Although research has found that this cycle influences women's mating preferences, we proposed that it might also change women's political and religious views. Building on theory suggesting that political and religious orientation are linked to reproductive goals, we tested how fertility influenced women's politics, religiosity, and voting in the 2012 U.S. presidential election. In two studies with large and diverse samples, ovulation had drastically different effects on single women and women in committed relationships. Ovulation led single women to become more liberal, less religious, and more likely to vote for Barack Obama. In contrast, ovulation led women in committed relationships to become more conservative, more religious, and more likely to vote for Mitt Romney. In addition, ovulation-induced changes in political orientation mediated women's voting behavior. Overall, the ovulatory cycle not only influences women's politics but also appears to do so differently for single women than for women in relationships.
Raw data

- answers to three statements on religiosity
- answers to five statements on social political attitudes
- answer to “what is your current romantic relationship status?” (choose 1 of 4 answer options)
- answers to fertility related questions
  - the start date of the last period
  - the typical cycle length
  - how sure are you about the start date of the last period
  - ...

Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps

1. **cycle day assessment**: transforming and combining variables
2. **fertility assessment**: dichotomization (high vs low fertility)
   - high in fertility when cycle day is between 7 and 14
   - low in fertility when cycle day is between 17 and 25
3. **relationship status assessment**: dichotomization (single vs committed relationship)
4. exclusion criteria based on cycle length
5. exclusion criteria based on variable sure
Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps:

→ 1 processed data set

→ Single data set analysis: effect of fertility x relationship status
Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps

→ 1 processed data set

→ single data set analysis: effect of fertility x relationship status

- ... on religiosity

![Graph showing religiosity composite score by relationship status (Not Ovulating vs Ovulating) and relationship status (Single vs In Committed Relationship).](image)

F(1,159) = 6.46, p = .012
Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps:

1. Processed data set

2. Single data set analysis: effect of fertility x relationship status

- ... on social political attitudes

![Graph showing social political attitudes by relationship status and ovulation status](image)

F(1, 299) = 12.26, p = .001
Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps:

→ 1 processed data set

→ single data set analysis: effect of fertility x relationship status

- ... on religiosity: $F(1,159) = 6.46$, $p = .012$
- ... on social political attitudes: $F(1,299) = 12.26$, $p = .001$
Translating the observed, raw data to the processed data ready for analysis involved several data processing steps:

1. one processed data set

2. single data set analysis: effect of fertility x relationship status
   
   - ... on religiosity: $F(1,159) = 6.46, p = .012$
   
   - ... on social political attitudes: $F(1,299) = 12.26, p = .001$

3. are these results robust? which results do we get with other processing choices?
Data processing: Durante et al. (2013)

Translating the observed, raw data to the processed data ready for analysis involved several data processing steps

→ 1 processed data set

→ single data set analysis: effect of fertility x relationship status

   - ... on religiosity: $F(1,159) = 6.46$, $p = .012$
   - ... on social political attitudes: $F(1,299) = 12.26$, $p = .001$

→ are these results robust? which results do we get with other processing choices?

   → multiverse analysis
Data processing: other reasonable options

1. cycle day assessment: transforming and combining variables
2. fertility assessment: dichotomization (high vs low fertility)
   - high in fertility when cycle day is between 7 and 14
   - low in fertility when cycle day is between 17 and 25
3. relationship status assessment: dichotomization (single vs committed relationship)
4. exclusion criteria based on cycle length
5. exclusion criteria based on variable sure
Data processing: other reasonable options

1. cycle day assessment: transforming and combining variables
2. **fertility assessment: dichotomization (high vs low fertility)**
   - high in fertility when cycle day is between 7 and 14
   - low in fertility when cycle day is between 17 and 25
3. relationship status assessment: dichotomization (single vs committed relationship)
4. exclusion criteria based on cycle length
5. exclusion criteria based on variable sure
Data processing: other reasonable options

1. cycle day assessment: transforming and combining variables

2. fertility assessment: dichotomization (high vs low fertility)
   - high in fertility when cycle day is between 6 and 14
   - low in fertility when cycle day is between 17 and 27
   (Durante et al., 2011)

3. relationship status assessment: dichotomization (single vs committed relationship)

4. exclusion criteria based on cycle length

5. exclusion criteria based on variable sure
1. cycle day assessment: transforming and combining variables
2. **fertility assessment: dichotomization (high vs low fertility)**
   - high in fertility when cycle day is between 9 and 17
   - low in fertility when cycle day is between 18 and 25
   (Durante et al., 2012)
3. relationship status assessment: dichotomization (single vs committed relationship)
4. exclusion criteria based on cycle length
5. exclusion criteria based on variable sure
Data processing: other reasonable options

1. Cycle day assessment: transforming and combining variables
2. Fertility assessment: dichotomization (high vs low fertility)
   - High in fertility when cycle day is between 8 and 14
   - Low in fertility when cycle day is between 1 and 7 or 15 and 28
   (Durante et al., 2014)
3. Relationship status assessment: dichotomization (single vs committed relationship)
4. Exclusion criteria based on cycle length
5. Exclusion criteria based on variable sure
Data processing: other reasonable options

1. cycle day assessment: transforming and combining variables
2. **fertility assessment: dichotomization (high vs low fertility)**
   - high in fertility when cycle day is between 9 and 17
   - low in fertility when cycle day is between 1 and 8 or 18 and 28
   (Durante et al., 2015)
3. relationship status assessment: dichotomization (single vs committed relationship)
4. exclusion criteria based on cycle length
5. exclusion criteria based on variable sure
Data processing: other reasonable options

1. cycle day assessment (3 choice options)
2. fertility assessment (5 choice options)
3. relationship status assessment (3 choice options)
4. exclusion criteria based on cycle length (3 choice options)
5. exclusion criteria based on variable sure (2 choice options)
1. cycle day assessment (3 choice options)
2. fertility assessment (5 choice options)
3. relationship status assessment (3 choice options)
4. exclusion criteria based on cycle length (3 choice options)
5. exclusion criteria based on variable sure (2 choice options)

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets
Multiverse analysis: results

perform analysis on each data set in multiverse → multiverse of results
Multiverse analysis: results

religiosity - fertility x relationship status
Multiverse analysis: results

religiosity - fertility x relationship status
Multiverse analysis: results

religiosity - fertility x relationship status

The effect is too fragile to be taken seriously.
Multiverse analysis: results

social political attitudes - fertility x relationship status
Multiverse analysis: results

social political attitudes - fertility x relationship status

49%
Multiverse analysis: results

social political attitudes - fertility x relationship status

49%

the belief in the effect heavily depends on the belief in the different processing choices
→ closer inspection
Multiverse analysis: results

social political attitudes - fertility x relationship status

the belief in the effect heavily depends on the belief in the different processing choices → closer inspection
Multiverse analysis: results

social political attitudes - fertility x relationship status

if you trust R1, there is a significant interaction effect

if you trust R2, there is no significant interaction effect

if you trust R3 or don’t know what to trust
- you are uncertain about the interaction effect
- your main conclusion should be a methodological one, i.e. we need to deflate the multiverse
Deflating the multiverse

- **Better theory** (assessment of fertility)
Deflating the multiverse

- Better theory (assessment of fertility)
Deflating the multiverse

- **Better theory** (assessment of fertility)
- **Cleaner design** (assessment of relationship status)
Deflating the multiverse

- **Better theory** (assessment of fertility)
- **Cleaner design** (assessment of relationship status)
Multiverse analysis: discussion

- shows the robustness/fragility of a finding
- helps identifying the sources of the fragility
- is not a formal test to assess strength of evidence
- is primarily meant to be used on own data
- is subjective, but less so than a single data set analysis
- is not restricted to p-values
  - e.g., Bayes factors are equally sensitive to choices in data construction
- in case of pre-registration: pre-register a multiverse analysis
- shows a small part only of a larger multiverse of statistical results
  - there is not only a data multiverse but also a model multiverse (e.g., Gelman & Loken, 2014; Patel, Burford, & Ioannidis, 2015)
Thank you for listening

Interested to do a PhD on responsible research practices? Contact wolf.vanpaemel@kuleuven.be

Data processing: other reasonable options

Exclusion criteria

answers to fertility related questions
  - the start date of the last period
  - the start date of the period before the last period
  - the typical cycle length
  - the expected start date of the next period
  - how sure are you about the start of the last period
  - how sure are you about the start date of the period before the last period

exclude women who are not sure about their start dates
(Durante et al., 2014)
Data processing: other reasonable options

Exclusion criteria

answers to fertility related questions
- the start date of the last period
- the start date of the period before the last period
- the typical cycle length
- the expected start date of the next period
- how sure are you about the start of the last period
- how sure are you about the start date of the period before the last period

exclude women who have irregular cycle lengths (calculated cycle length)
Data processing: other reasonable options

Fertility status
answers to fertility related questions
- the start date of the last period
- the start date of the period before the last period
- the typical cycle length
- the expected start date of the next period
- how sure are you about the start of the last period
- how sure are you about the start date of the period before the last period

→ dichotomization: high vs low fertility

high in fertility when cycle day is between 9 and 17
low in fertility when cycle day is between 1 and 8 or 18 and 28
(Durante et al., 2015)
Data processing: other reasonable options

Fertility status

answers to fertility related questions
- the start date of the last period
- the start date of the period before the last period
- the typical cycle length
- the expected start date of the next period
- how sure are you about the start of the last period
- how sure are you about the start date of the period before the last period

→ dichotomization: high vs low fertility

high in fertility when cycle day is between 9 and 17
low in fertility when cycle day is between 1 and 8 or 18 and 28

(Durante et al., 2015)
Deflating the multiverse

assessment of relationship status
- answer to “what is your current romantic relationship status?”
  (1) not dating/romantically involved with anyone
  (2) dating or involved with only one partner
  (3) engaged or living with my partner
  (4) married
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 1
2. fertility assessment (5 choice options) → option 1
3. relationship status assessment (3 choice options) → option 1
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

data set 1

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 2
2. fertility assessment (5 choice options) → option 1
3. relationship status assessment (3 choice options) → option 1
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets.
## Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 3
2. fertility assessment (5 choice options) → option 1
3. relationship status assessment (3 choice options) → option 1
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

Data set 3
Data set 2
Data set 1

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets
Data processing: other reasonable options

1. cycle day assessment (3 choice options)  → option 1
2. fertility assessment (5 choice options)  → option 2
3. relationship status assessment (3 choice options)  → option 1
4. exclusion criteria based on cycle length (3 choice options)  → option 1
5. exclusion criteria based on variable sure (2 choice options)  → option 1

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets. 

data set 4
data set 3
data set 2
data set 1
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 1
2. fertility assessment (5 choice options) → option 3
3. relationship status assessment (3 choice options) → option 1
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

data set 5
data set 4
data set 3
data set 2
data set 1

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets.
### Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 1
2. fertility assessment (5 choice options) → option 1
3. relationship status assessment (3 choice options) → option 2
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

Rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets:

- data set 6
- data set 5
- data set 4
- data set 3
- data set 2
- data set 1
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 1
2. fertility assessment (5 choice options) → option 1
3. relationship status assessment (3 choice options) → option 3
4. exclusion criteria based on cycle length (3 choice options) → option 1
5. exclusion criteria based on variable sure (2 choice options) → option 1

data set 7
data set 6
data set 5
data set 4
data set 3
data set 2

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 3
2. fertility assessment (5 choice options) → option 3
3. relationship status assessment (3 choice options) → option 5
4. exclusion criteria based on cycle length (3 choice options) → option 3
5. exclusion criteria based on variable sure (2 choice options) → option 2

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets.

data set 210
...
data set 7
data set 6
data set 5
data set 4
data set 3
data set 2
data set 1
Data processing: other reasonable options

1. cycle day assessment (3 choice options) → option 3
2. fertility assessment (5 choice options) → option 3
3. relationship status assessment (3 choice options) → option 5
4. exclusion criteria based on cycle length (3 choice options) → option 3
5. exclusion criteria based on variable sure (2 choice options) → option 2

rather than arbitrarily choosing 1 option and creating a single processed data set, we consider all combinations of options, which give rise to a multiverse of data sets:

- data set 210
- data set 7
- data set 6
- data set 5
- data set 4
- data set 3
- data set 2
- data set 1

data multiverse