02\_descriptive\_stats

Timo Roettger

2024-08-21

## General characteristics

## Loading required package: pacman

## Rows: 943 Columns: 29  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (25): ApproximateTitle, major\_issue, study\_type, preregistered, synthesi...  
## dbl (2): article\_id, year  
## lgl (2): synth\_location, synth\_access  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## [1] "no issues" "other" "not a full article"  
## [4] "yes" "no access" "not in English"   
## [7] NA

799 of articles were coded as in principle assessible ( 84.7%) ). Only those articles were further analysed. 58 of articles could not be accessed by the coder (6.2%); 24 of articles were coded as not being a full article, i.e. conference abstract or an erratum ( 2.5%); 2 of articles were coded as not being written in English ( 0.2%); and 60 of articles were coded as having other issues ( 6.4%) were coded as other issues, including papers that did not use MoBa.

## [1] 2022 2023 2021 2020 2019 2018 2017 2015 2016 2014 2013 2012 2011 2010 2009  
## [16] 2008 2007 NA 2024

## [1] "Secondary analysis of MoBa data" "Data synthesis - Meta-analysis."  
## [3] "no empirical data" "Unclear"

Of those articles that were in principle assessible, 701 of articles were coded as secondary analysis of MoBa data (87.7%) and 32 of articles were coded as being part of a meta analysis (including other epidiomological data sets, 4%). Only those two categories were further evaluated. 29 of articles were coded as not presenting data (i.e. editorials or reviews, 3.6%) and 37 of articles could not be categorized (“unclear) but were considered not further assessible (i.e. editorials or reviews, 4.6%). This left us with 733 articles that were further analysed.

## [1] "No" "Yes"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, preregistered)`

## [1] "Open Science Framework (osf.io)"

## [1] "Yes" "No"

## [1] "Hypotheses;Research question;Variables;Data analysis"   
## [2] "Research question;Variables;Data processing;Model structure"

In the entire corpus, there were 3 studies that reported to have preregistered aspects of their analysis. With the available information we were able to recover the original preregistration (albeit it required some digging). Upon further scrutiny, all three preregistrations preregistered some aspects of the analysis including research question, variables and data analysis, but LOOK DEEPER INTO THOSE THREE

## [1] "No"

None of the studies reported to use synthesized data, which in principle would allow to reproduce their analysis without violating data protection laws.

## [1] "no" "yes" NA

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, other\_data)`

## [1] "Upon request from the authors"   
## [2] "Supplementary materials hosted by the journal"   
## [3] "Available from a third party"   
## [4] "Unclear"   
## [5] "An online, third-party repository (e.g., OSF, GitHub, FigShare etc.)"

## [1] "No" "Yes"

## [1] "unclear" "metadata" "summary statistics"  
## [4] "other data" NA

## # A tibble: 5 × 3  
## other\_data\_type n prop  
## <chr> <int> <dbl>  
## 1 metadata 8 47.1   
## 2 other data 3 17.6   
## 3 summary statistics 2 11.8   
## 4 unclear 3 17.6   
## 5 <NA> 1 5.88

17 of assessed articles were coded as reporting to share resources that could be considered some form of non-primary data (2.3%), mostly additional meta data or summary statistics. This information was reported to be found in supplementary materials hosted by the journal (n = 8), available “upon request” from the authors (n = 4), or available from a third party (n = 3). Only one article pointed to an open online third-party repository like the OSF or Github and one article remained too vague to recover where to access the information. All in all, out of the 17 only 11 could be actually accessed.

## [1] "no" "freely available"   
## [3] "upon request" "authors point to third party"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, analysis)`

## [1] "An online, third-party repository (e.g. OSF, GitHub, FigShare, etc.)"  
## [2] "Upon request from the authors"   
## [3] "Supplementary materials hosted by the journal"   
## [4] "Personal or institutional website"

## [1] "Yes" "No" NA

31 of assessed articles were coded as reporting to share some form of analysis protocols either in form of step-by-step instructions or code (5%). Out of these, 16 articles pointed to an open online third-party repository like the OSF or Github, 7 articles pointed to supplementary materials hosted by the journal, 1 articles pointed to personal or institutional websites, and 7 were described as available “upon request” from the authors. All in all, out of the 7, 19, 5 only 24 could actually be accessed.

## [1] "no" "sensitivity" "sensitivity & power"  
## [4] "power"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, robust\_check)`  
## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.

472 of assessed articles reported neither a power analysis nor a sensitivity analysis (64%). 26 of assessed articles reported a power analysis (3.5%), 221 of assessed articles reported a sensitivity analysis (30%), and 18 of assessed articles reported both a power and sensitivity analysis (2.4%).

## [1] "no" "freely available"   
## [3] "authors point to third party" NA

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, materials)`

## [1] "Supplementary materials hosted by the journal"   
## [2] "Upon request from the authors"   
## [3] "Available from a third party"   
## [4] "Personal or institutional website"   
## [5] "An online, third-party repository (e.g. OSF, GitHub, FigShare, etc.)"

## [1] "Yes" "No"

NA of assessed articles were coded as reporting to share some form of additional materials (3.4%). Out of these, 2 articles pointed to an open online third-party repository like the OSF or Github, 15 articles pointed to supplementary materials hosted by the journal, 4 articles pointed to third party, and 2 were described as available “upon request” from the authors. All in all, out of the NA only 17 could actually be accessed.

## [1] "no" "yes" "unclear"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, replication)`

11 of assessed articles were coded as reporting to replicate a previous study (1.5%). On further inspection… CHECK THEM INDIVIDUALLY

## [1] "statement says yes" "statement says no" "no statement"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, coi)`

142 of assessed articles were coded as not reporting whether or not a conflict of interest applied to their study (19.3%); 488 of assessed articles were coded as reporting that there was no conflict of interest applied to their study (66.2%); and 107 of assessed articles were coded as reporting whether that there in fact was a conflict of interest applied to their study (14.5%)

## [1] "Yes" "No"

## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.  
## Joining with `by = join\_by(year, isOA)`

661 of assessed articles were coded as being open access according to our definition (89.7%);

## Overall plot over time