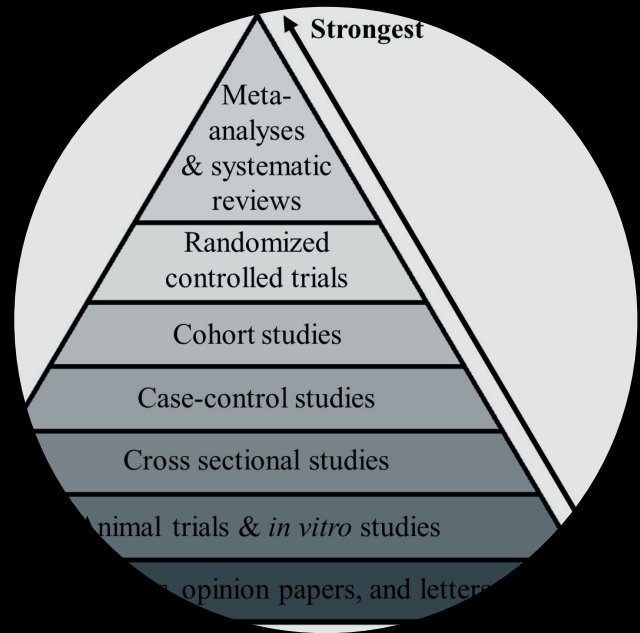


Meta-analysis data management with the {metamanager} package

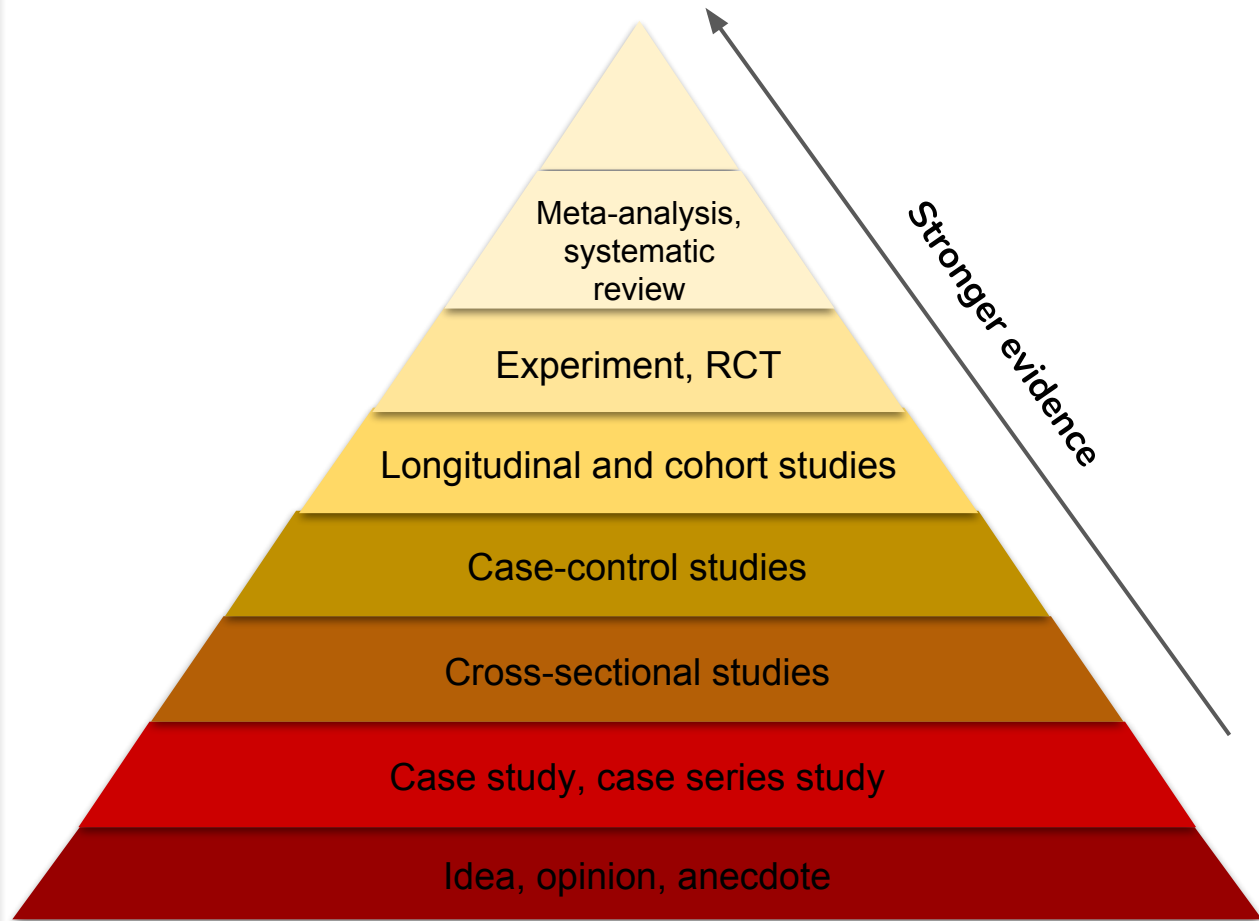
Tamas Nagy, PhD
Eotvos Lorand University
Budapest, Hungary



**We are living in a world where
facts can be regarded as
subjective therefore we need
more evidence to support
claims**



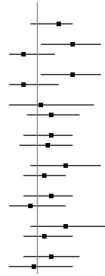
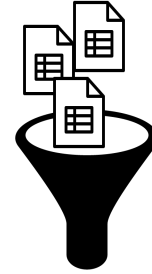
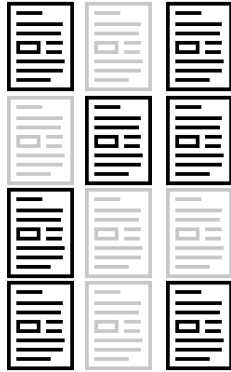
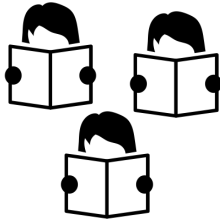
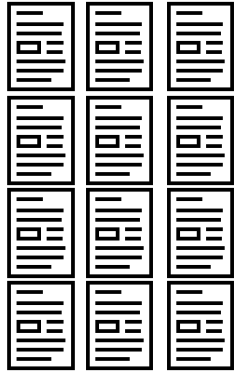
Pyramid of scientific evidence



What is the point of a meta-analysis

- Helps to draw conclusions for a whole field
- Synthesize research findings in an area across several independent studies
- May help to show the “true” effect
- It is possible to test new hypotheses if data are available from the articles

The process of doing a meta-analysis



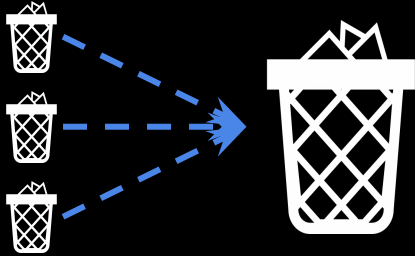
1. Search literature
2. Screen articles
for eligibility

3. Select the studies of interest
4. Extract data from published
articles

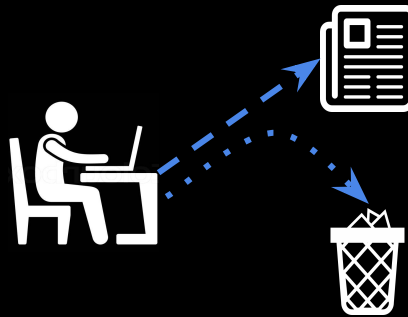
5. Calculate effect sizes
and aggregate across
studies
6. Statistical inference

Shortcomings of meta-analysis

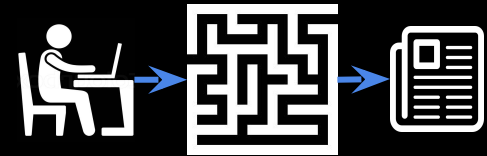
Garbage in -
garbage out



Publication bias



Questionable
research practices



There are standards that regulate how to do a meta-analysis properly

- A **protocol** defines the aims and methods
- Decisions about article screening and data extraction should be made **at least by two independent researchers**
- Assessing the quality of the included studies
- But usually the amount of work requires more collaborators



Doing a meta-analysis is not easy



Domain specific
knowledge



Methodological
knowledge

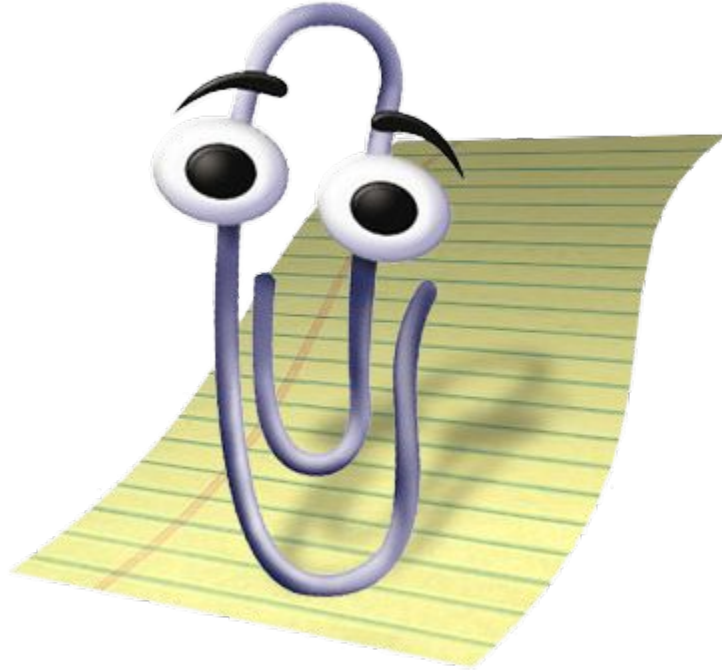


Teamwork



Data management
and organization

Data management usually organized in Excel



There should be a package for this!

- Reduce project management overhead
- File management should be automated and made reproducible
- Not straightforward to coordinate human and machine generated outputs
- Integrate with a GUI that is easy to use
- Currently, there is only one other package [{metagear}](#) for meta-analysis workflow management, but it is not using the tidyverse principles, and it has a different philosophy and focus

The {metamanager} is here to help!



- Automate file creation for manual coding (via Google drive)
- Aggregate, evaluate, validate human coded data
- Clear and reproducible outputs
- Embracing the tidyverse principles (piping, readability, simplicity of functions, seamless integration with tidy packages)

Setting up a meta-analysis project

`init_gdrive("research/meta-analysis")`: Creates a folder structure on google drive. Can be self specified or a default structure.

```
literature_search/  
screening/  
fulltext/  
screening_consensus/  
extraction/  
extraction_consensus/
```

	A	B	C	D
1	name	role	screening_effort	extraction_effort
2	Ali	trainee	0.25	0
3	Andreas	trainee	0.25	0
4	Lydia	trainee	0.25	0
5	Tamas	trainer	0.25	0.1
6	Kristoffer	trainee	0	0.4
7	Marcell	trainee	0	0.3
8	Zoli	trainer	0	0.2

`init_team()`: Creates a google sheet with team member data. Accepts a tibble or opens a google sheet with header

`init_meta("research/meta-analysis")`: Does both, using defaults

Identify duplicates

```
merge_sources(workaholism_psychinfo,  
              workaholism_pubmed,  
              workaholism_scopus,  
              .renames = c("journal" = "publication"),  
              .tidy_names = TRUE,  
              .all_char = TRUE) %>%  
  flag_duplicate_title(keys = c("psyid", "pmid", "doi", "eid", "sid")) %>%  
  make_id(c("psyid", "pmid", "doi", "eid", "sid")) %>%  
  flag_duplicate_id(keys = c("psyid", "pmid", "doi", "eid", "sid")) %>%  
  filter(duplicate_by_title != 1 & duplicate_by_id != 1)
```

Create article screening files

```
merged_records %>%
  assign_articles(team_df = team_df, effort_col = "screening_effort", seed = 42)
%>%
  add_columns(c("decision", "reason")) %>%
  save_locally(local_path = "temp/screening/",
               nesting = "reviewer",
               postfix = "screening",
               overwrite = TRUE)

# A tibble: 4,390 x 15
  decision reason identifier id      source title journal authors year abstract type level position
  <chr>      <chr> <chr>      <chr> <chr> <chr> <chr> <chr> <dbl> <chr>      <chr> <chr> <chr>
1 NA        NA      pmid        24357~ pubmed 'Res~ Journa~ Silk, ~ 2013 "Whilst n~ NA     NA     review~
2 NA        NA      pq_id       33710~ proqu~ 'U G~ NA     Young,~ 2007 Backgrou~ THES  NA     review~
3 NA        NA      doi         10.10~ Scopu~ "\"B~ Accide~ Mamm~ 2013 Two studi~ Arti~ NA     review~
4 NA        NA      doi         10.11~ Scopu~ "\"B~ Entrep~ Wolfe ~ 2015 In this s~ Arti~ NA     review~
5 NA        NA      doi         10.10~ Scopu~ "\"C~ Femini~ De E.N. 2012 "This art~ Arti~ NA     review~
6 NA        NA      doi         10.10~ Scopu~ "\"G~ Comput~ Charsk~ 2011 Does usin~ Arti~ NA     review~
# ... with 4,380 more rows, and 2 more variables: reviewer <chr>, name <chr>
```

Uploading and downloading gdrive folders

```
#This also converts files to a google sheets that can be edited online
upload_to_gdrive(local_path = "temp/screening/",
                 gdrive_path = "research/meta-analysis/screening",
                 overwrite = TRUE)
```



```
#Downloads all files from a google folder as a nested dataframe
screening <- get_from_gdrive(gdrive_path = "research/meta-analysis/screening",
                           all_char = FALSE)
```

Why Google drive?



Google Drive

- There are built-in functions in the package to convert data files to Google sheets
- Google sheets provide **collaborative editing** with **version control**
- **User friendly** GUI and API
- **Not obligatory:** All files are also saved locally, other cloud solutions can be used too, like github or dropbox
- It would be possible to create a shiny app for front end (e.g. <http://dev.open-meta.org/>)

Evaluate screening performance

```
screening %>%
  unnest() %>%
  calculate_agreement()
```

A tibble: 3 x 8

	name_pair <chr>	include_both <int>	exclude_both <int>	no_agreement <int>	invalid_decision <int>	all_items <int>	kr_level <chr>	kr_alpha <dbl>
1	Tom_Jerry	0	34	20	0	54	nominal	-0.216
2	Tom_Rose	1	2	7	0	10	nominal	-0.343
3	Jerry_Rose	1	5	4	0	10	nominal	0.0952

Human errors during coding of the studies

Valid reasons for exclusion

publication type
study design
sample characteristics
manipulation
outcome measures
time frame
duplicated
unavailable
unspecified reason

What we see in the file

publicatin type
pub type
publication
study design
study d
sample characteristics
sample
samle
samp
...

Correcting and highlighting human errors

```
# Correct typos in the reasons for excluded studies
correct_terms <- read_lines("valid_exclude_reasons.txt")
screening %>%
  unnest() %>%
  mutate(exclude_reason_corr = correct_categories(exclude_reason, correct_terms))
```

Humans tend to ignore variable types

```
# Safely convert col_types with human readable errors

articles %>%
  conversion_errors(c("year", "measurement_n"), as.integer)

# A tibble: 2 x 3
  file          fun          conversion_errors observations
  <chr>        <chr>        <dbl>            <list>
1 Ali_extract.csv as.integer    25               list(43, 88, 89, 123)...
2 Tamas_extract.csv as.integer     3               list(43, 288, 589)
```

Create data extraction files

```
variables_to_add <- c("subsample_id", "subsample_age", "subsample_female%",  
"measurement_point", "measurement_n", "measurement_mean", "measurement_sd",  
"study_design", "study_quality")  
  
get_from_gdrive(screening_path) %>%  
  unnest() %>%  
  filter(is.na(exclude_reason)) %>%  
  assign_articles(team_df = team_df,  
                 effort_col = "extraction_effort", seed = 42) %>%  
  add_columns(variables_to_add, before = FALSE) %>%  
  save_locally(local_path = "temp/extraction/", nesting = "reviewer", postfix =  
"screening", overwrite = TRUE)  
  
upload_to_gdrive(local_path = "temp/extraction/",  
                 gdrive_path = "research/meta-analysis/extraction",  
                 overwrite = TRUE)
```

Future plans

- Machine learning augmented screening process
- Automatic extraction of citations from included articles (pdf or wos)
- More data visualization (flow charts, keywords in abstracts, ...)
- Handle article pdfs (download, categorize, distribute, etc.)
- Tidyeval
- Fancy package stuff: more vignettes, unit tests, pkgdown site, hex sticker

How to get / contribute?

Please star, fork, open issues, PR on github!
<https://github.com/nthun/metamanager>

lifecycle experimental

Installation:

```
devtools::install_github("nthun/metamanager")
```

Thank you!