



The Preregistration of Secondary Data Analysis

Olmo van den Akker, 08-03-2019

Meta-Research Center

MET^A

META-RESEARCH CENTER

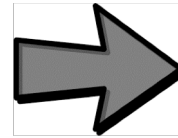
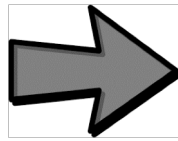
Tilburg School of Social and Behavioral Sciences



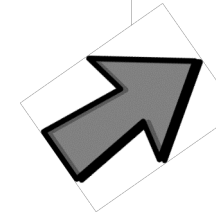
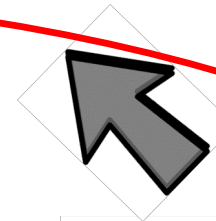
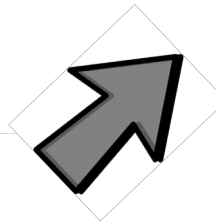
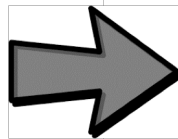
Secondary Data

Secondary Data

Primary data



Secondary data



Existing data



Examples of Secondary Data

European Social Survey

About ESS Findings Methodology Data and Documentation Learning

The European Social Survey

We're an academically driven cross-national survey using high methodological standards to provide freely available data for 36 countries. [MORE...](#)



World Bank Open Data

Free and open access to global development data

Search data e.g. GDP, population, Indonesia


Browse by Country or Indicator

MOST RECENT

- Demystifying machine learning for disaster risk management
G. Molinaro, Mar 06, 2019
- Moving towards gender equality: A new index looks at legal reforms to help women's economic inclusion
S. Iqbal, Feb 27, 2019
- New Mali survey data now available!
M. Tiberti, Feb 27, 2019
- Like manna from heaven? The sustainability of Open Source projects

WHAT YOU CAN LEARN WITH OPEN DATA

Poverty headcount ratio at \$1.90 a day (2011 PPP) (% of population)




Data from World Bank

Extreme Poverty

The proportion of the world's population living in extreme poverty has dropped significantly

THE NEW World Development Indicators



Do children who work attend school?
In Sub-Saharan Africa, HIV is more common among women than men
Globally, more than 1 in 10 people live in the open

Indicator Code Time coverage Region coverage Get data

| Indicator | Code | Time coverage | Region coverage | Get data |
|--------------------------------------|----------------|---------------|-----------------|----------|
| Population, total | SP.POP.TOTL | 1950-2019 | ALL | API |
| Population growth (annual %) | SP.POP.GROW | 1950-2019 | ALL | API |
| Birth rate, crude (per 1,000 people) | SP.DNY.CBRT.O1 | 1950-2019 | ALL | CSV |
| Death rate, crude (per 1,000 people) | SP.DNY.CBRT.O1 | 1950-2019 | ALL | Dataviz |

World Development Indicators

OSFHOME

My Quick Files My Projects Search Support Donate Olmo van den Akker

Public projects

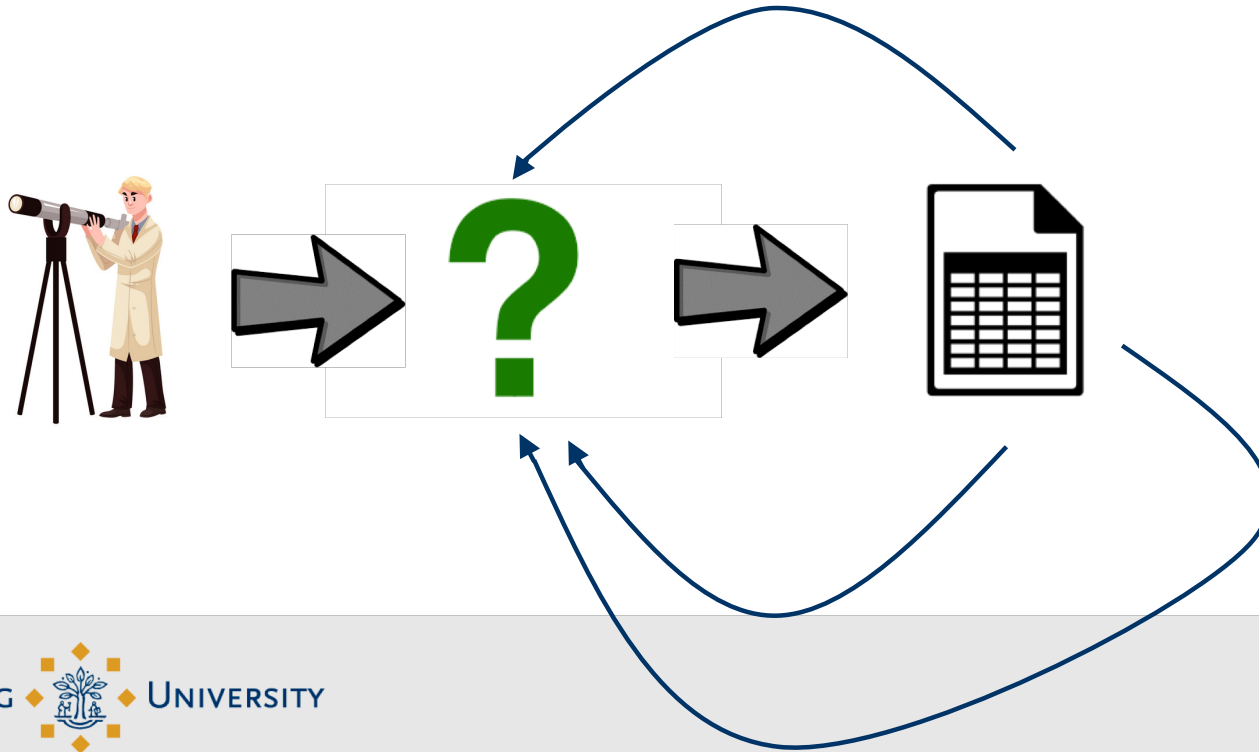
- Database of Articles with Open Science Badges**
Fidler, Singleton Thorn, Kambouris & 1 more
- What heuristics do researchers use when assessing the outcomes of multiple studies?**
Van den Akker
- Secondary Data Preregistration**
Weston, Mellor, Bakker & 15 more
- Sex differences in trust and trustworthiness**
Van den Akker

Public components

- Sex Differences in Trust and Trustworthiness - A Meta-Analysis of the Trust Game and the Gift-Exchange Game**
Sex differences in trust and trustworthiness / Sex Differences in Trust and Trustworthiness - A Meta-Analysis of the Trust Game and the Gift-Exchange Game
Van den Akker
- Hack-A-Thon: Secondary Data Template**
Preregistration hack-a-shop / Hack-A-Thon: Secondary Data Template
Weston & Bakker
- Preregistration hack-a-shop**

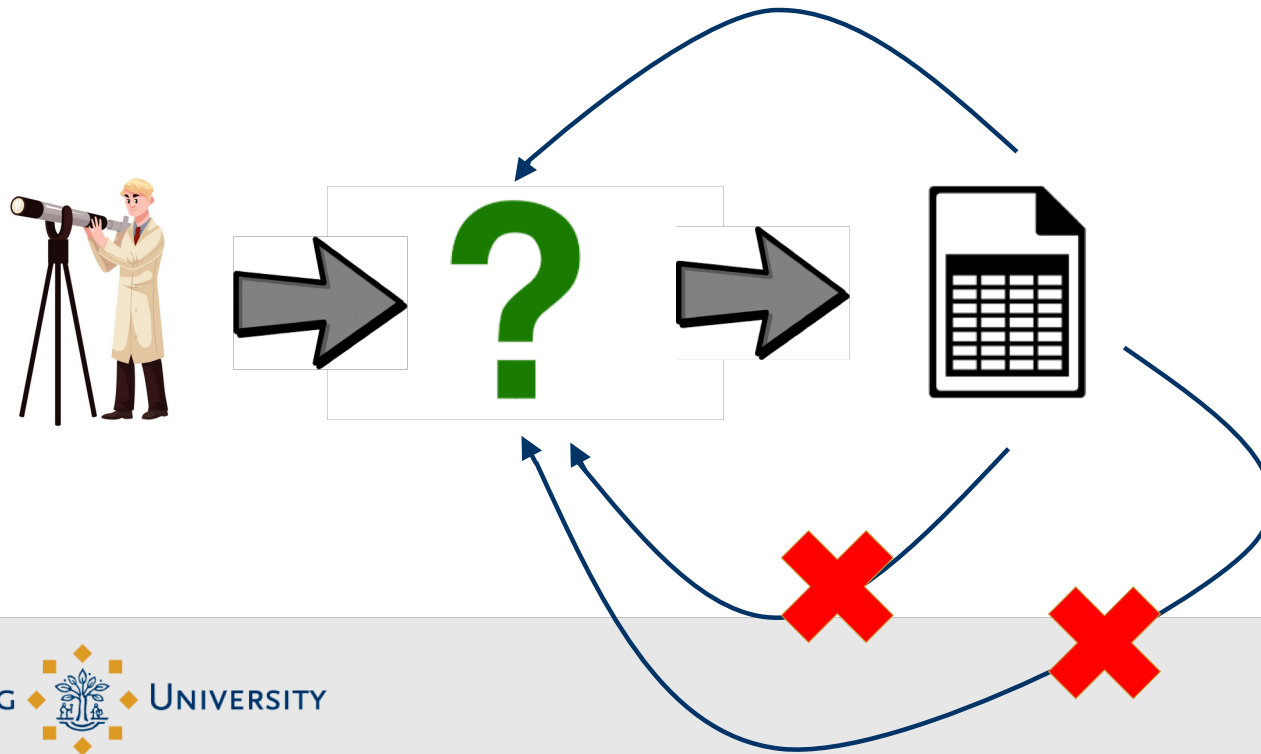
Characteristics of Secondary Data

1. Many different variables and respondents → “Data buffet”



Secondary Data

2. Unique → Many analyses → Prior knowledge of the data



A Preregistration Template for Secondary Data Analysis



SOCIETY FOR THE
IMPROVEMENT OF
PSYCHOLOGICAL SCIENCE

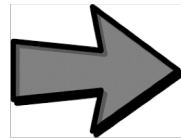
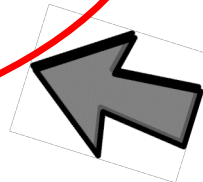
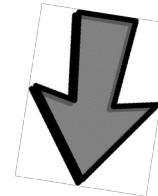


Table 1.

Nine Recommendations to Increase Transparency in Analysis of Pre-existing Data


- 1 Provide links to codebooks and data access instructions.
- 2 Provide the data access form or email.
- 3 List the specific variables used in the analyses.
- 4 Post the analytic script prior to analysis.
- 5 Describe prior by others research using the same variables in the same data set.
- 6 Describe any analyses, published or unpublished, that you have done on any of the variables used in the current study.
- 7 Register analysis plan prior to examining data.
- 8 Preregister analyses for upcoming and not-yet released waves of data.
- 9 Post analytic script prior to analysis.



SOCIETY FOR THE
IMPROVEMENT OF
PSYCHOLOGICAL SCIENCE

OSFHOME ▾ Search Support Donate Sign Up Sign In

Secondary Data Preregistration Files Wiki Analytics Registrations


 Secondary Data Preregistration Public 15

Contributors: Sara J. Weston, David Mellor, Marjan Bakker, Olmo van den Akker, Lorne Campbell, Stuart J. Ritchie, William J. Chopik, Rodica I. Damian, Jessica Kosie, Courtney K. Soderberg, Charles R. Ebersole, Brian Brown, Pamela Davis-Kean, Andrew Hall, Elliott Kruse, Jerome Olsen, K.D. Valentine, Thuy-vy Nguyen


Affiliated Institutions: Center For Open Science

Forked from osf.io/djgww on 2018-09-11 05:37 PM

Date created: 2018-09-11 05:37 PM | Last Updated: 2018-11-16 11:07 PM

Category:  Instrumentation

Description: This project contains a template and instructions for registering a research project that uses an existing dataset.

License: CC0 1.0 Universal 

A Tutorial

RQ → Are more religious people more prosocial than less religious people?



Data → Wisconsin Longitudinal Study



How to solve the “Data buffet” issue?

Specify your variables and statistical analyses in detail!

RQ → Are more religious people more prosocial than less religious people?

- Status and Descriptive Variables
- Ancillary Data
- 1957 - 1964 (graduates)
- 1975 (graduates)
- 1977 (siblings)
- 1992 - 1993 (graduates)
- 1993 - 1994 (siblings)
- 2003 - 2005 (graduates and spouses)
- 2004 - 2007 (siblings and sibling spouses)
- 2011 CAPI (graduates and siblings)
- Archived variables

| | | |
|---|-------------------------------|------------------------------|
| Phone: Health | HTML Codebook | PDF Codebook |
| Phone: Cognition Similarities | HTML Codebook | PDF Codebook |
| Phone: Employment | HTML Codebook | PDF Codebook |
| Phone: Pensions | HTML Codebook | PDF Codebook |
| Phone: Job characteristics | HTML Codebook | PDF Codebook |
| Graduate Phone: Access to Health Care and Insurance | HTML Codebook | PDF Codebook |
| Phone: Other income | HTML Codebook | PDF Codebook |
| Phone: Assets | HTML Codebook | PDF Codebook |
| Phone: Cognition - word recall | HTML Codebook | PDF Codebook |
| Phone: Volunteering | HTML Codebook | PDF Codebook |
| Phone: Alcohol | HTML Codebook | PDF Codebook |
| Phone: Religion | HTML Codebook | PDF Codebook |
| Phone: Internet | HTML Codebook | PDF Codebook |
| Phone: Psychological well-being | HTML Codebook | PDF Codebook |

How to solve the “Data buffet” issue?

Specify your variables and statistical analyses in detail!

RQ → Are more religious people more prosocial than less religious people?

1. il001rer: How religious are you?
2. il002rer: How spiritual are you?
3. il003rer: How important is religion in your life?
4. il004rer: How important is spirituality in your life?
5. il005rer: How important was it, or would it have been if you had children, to send your children for religious or spiritual instruction?
6. il006rer: How closely do you identify with being a member of a religious group?

How to solve the “Data buffet” issue?

Specify your variables and statistical analyses in detail!

RQ → Are more religious people more prosocial than less religious people?

1. il001rer: How religious are you?
2. il002rer: How spiritual are you?
3. il003rer: How important is religion in your life?
4. il004rer: How important is spirituality in your life?
5. il005rer: How important was it, or would it have been if you had children, to send your children for religious or spiritual instruction?
6. il006rer: How closely do you identify with being a member of a religious group?

1. gv103re: Did the graduate do volunteer work in the last 12 months?

2. gv111re: How many hours did the graduate volunteer during a typical month in the last 12 months?

How to solve the “Data buffet” issue?

Specify your variables and statistical analyses in detail!

RQ → Are more religious people more prosocial than less religious people?

Analysis 1:

$$VolunteerBinary = \beta_0 + ReligiosityIndex * \beta_1 + Controls + Error\ term$$

Analysis 2:

$$VolunteerCont = \beta_0 + ReligiosityIndex * \beta_1 + Controls + Error\ term$$

Don't forget outliers and missing data!

How to solve the “Prior knowledge” issue?

Specify your prior knowledge of the data in detail!

→ What amounts of prior knowledge do you already have for the specific dataset?

“In our previous study conducted on this data set, we found that agreeableness predicted (continuous) prosociality, which was normally distributed with mean 4.4 and standard deviation 1.9. We have not analyzed the religiosity variable at all. In addition, previous research (by other researchers) on this data set show that socio-economic status is related to religiosity and that prosociality is linked to mental health. Therefore, one could say that we are not completely blind to potential inter-correlations involving the variables in this study.”

How to solve the “Prior knowledge” issue?

Specify your prior knowledge of the data in detail!

“All three authors (OA, MB, SW) have worked with this data set before. Together, we previously used the Graduates 2003-2005 wave to assess the link between Big Five personality traits and prosociality. We presented these results at the ARP conference in St. Louis in 2013 and we are currently working on a manuscript. Additionally, a senior graduate student in MB’s lab used the data set for exploratory analyses on depression; these results have not yet been submitted anywhere.”

Take Away

Preregistration of secondary data is not that hard!

Specify your prior knowledge of the data set from your own previous use the data and from other researcher's previous use of the data, preferably for each author separately

→ Link to our preregistration template: <https://osf.io/x4gzt> (Tutorial out soon)

An Alternative

Preregistration of secondary analyses

Ad

Preprint (PDF Available) · February 2019 · with 35 Reads

DOI: 10.31234/osf.io/ph4q7

[Cite this publication](#)

<https://psyarxiv.com/ph4q7>



Gaëtan Mertens

11 21.83 · Utrecht University



Angelos-Miltiadis Krypotos

11 25.99 · KU Leuven & Utrecht University

Abstract

The preregistration of a study's hypotheses, methods, and data-analyses steps is becoming a popular psychological research practice. To date, most of the discussion on study preregistration has focused on the preregistration of studies that include the collection of original data. However, much of the research in psychology relies on the (re-)analysis of existing data (i.e., secondary analyses). Importantly, this type of studies is different from original studies as researchers cannot change major aspects of the study (e.g., experimental manipulations, sample size). Here, we provide arguments as to why it is useful to preregister secondary analyses, discuss practical considerations, and propose a preregistration template, tailored for studies using secondary analyses. We argue that the preregistration of hypotheses and data-analyses for secondary analyses is an important step towards more transparent psychological research.

The End



Olmo van den Akker

ovdakker@gmail.com

[@DenOlmo](#)