Preregistration

# Compliance of funder requirements of open access publication preregistration

Henrik Danielsson[1], Gustav Nilsonne[2,3], Lovisa Österlund[1], Johanna Nählinder[1], David Lawrence[1]

[1] Linköping University

[2] Stockholm University

[3] Karolinska Institutet

*07. September 2018*

## Study Information

| | |
|---|---|
| **Title** | Compliance of funder requirements of open access publication preregistration |
| **Research questions** | The main research question of this project is: how does the introduction of open access publishing mandates by science funders affect the degree of open access publication? |
| | Additionally, we aim to describe trends in open science publication of Swedish research, with respect to fraction of open access publication and types of open access. |

**Hypotheses**

- H1: The introduction of open access mandates will be aqssociated with an increase in open access publishing.

## Sampling Plan

Data sources: + We will collect bibliometric information and funding information from web of science via their API. + We will collect open access status of published articles from Unpaywall via the R package 'roadoi' (Jahn 2018).

Inclusion criteria: + Swedish affiliation for the first author. This is because every funding agency has different requirements and to make the project feasible we limit ourselves to Swedish funding agencies where we know when different open access requirements were introduced. Swedish affiliation is defined by "Sweden" or "Sverige" in the affiliation information for the article.

- Data will be collected for the following publication years 2007 to 2017. This peridod of time will cover the introduction of open access policies by major funders. The larger grant agencies in Sweden who has introduced an open access requiremnet are the following:

Vetenskapsrådet - The Swedish research agency staring from grants awarded 2010.

Forte - . . . . staring from grants awarded 2012

Formas - . . . . . staring from grants awarded 2010

Riksbankens Jubileumsfond - . . . . .staring from grants awarded 2010

Matching procedure: + The link between articles and funding information is the article doi.

- The funding information is free text. We will identify relevant funders with regular expressions.

We will search for diarienummer too . . . . .

**Assumptions:**

- Not all articles are published in WoS-journals. On the contrary, there is a rather large difference between academic fields. We are aware of this difference. . . .

- Mandated OA was introduced 2010 (2012). We assume that there is a time -lag of xx years from. . .

- Some articles do not have doi-identificators. The share of articles with doi-identificators have risen over time. We assume that articles with doi-identificator follow the same pattern as articles without doi-identificator.

- Some articles do not have funding information, although they are obliged to.

- We assume that articles incorrectly without funding information follow the same pattern as articles with funding information.

- Most articles are co-authored. Our study will identify all articles with at least one Swedish author. (Gustav: Fast där uppe säger vi svensk förstaförfattare, Henrik: Kunde vi definiera att det var förstaförfattare enkelt. Var det inte bara ett sökfält för affiliations?)

- We assume that the affiliations are correct, i.e. that the word "Sweden" is correctly used in affiliations

---

**Existing data**      **Registration prior to analysis of the data**. As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

---

**Explanation of existing data**      All data are available, but the combination of data from different sources is not. To avoid harking, one of the authors searched the first 200 entries from web of science to use as pilot data. On the pilot data, two other authors could optimize the regular expression to find different spellings of the grant agencies. Pilot data was also used to make sure that the coverage of funding information was sufficient (need a number?) in Crossref.

| | |
|---|---|
| **Data collection procedures** | Move something from above? |
| **Sample size** | Sample or population? |
| | The sample size will be all data that is available in the combination of data sources above. |
| **Sample size rationale** | The data is limited by when the requirement to publish open access by the main Swedish grant agencies was introduced. We collect data from 3 years before that was introduced to establish a baseline of degree of open acceess publications. Then we take all data up to 2017 to make sure that we have data for whole years. (Henrik: Vad sa vi om täckningsgraden i olika källor? Kunde vi börja 2007 som det står tidigare?) Data is also limited to one data source for each type of data. We have chosen the data sources with the best coverage of data that we are interested in by doing pilot searches. Web of Science had the best coverage of bibliometric information that is available via an API (Google Scholar has better coverage, but no easy way to get the data). .... |
| **Stopping rule** | No. |

## Variables

- publication year
- current article open access status (vi vet inte hur det var då? Lovisa kollar upp något)
- article funding information
- funding agency startyear of open access publication requirement
- funding agency type of open access publication requirement
- author affilliation
- article doi

| | |
|---|---|
| **Manipulated variables** | Enter your response here. |
| **Measured variables** | Enter your response here. |
| **Indices** | OA status. Green, Bronze, other categories... Specify! |

## Design Plan

| | |
|---|---|
| **Study type** | **Observational Study**. Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs. |
| **Blinding** | Data will be collected by two of the authors. The affiliations and oainformation will be given codes. Then two of the other authors will analyze the data without being aware of the meaning of the codes. |
| **Study design** | The design of this observational study is a longitudinal study.... |
| **Randomization** | No. |

## Analysis Plan

We will conduct a time series analysis and descriptive analyses. The dependent variable will be open access status of the articles. The independent variables will have different levels, ..... Each article has oa-status, funding information, first author, and year. The author has an affiliation.

| | |
|---|---|
| **Statistical models** | Enter your response here. |

| | |
|---|---|
| **Transformations** | No. |
| **Follow-up analyses** | Follow-up analyses will be made to investigate if the main result is different depending on affiliation (which university in Sweden). |
| **Inference criteria** | Enter your response here. |
| **Data exclusion** | Enter your response here. |
| **Missing data** | The degree of missing data from the different data sources will be documented. Will we do an analysis of these? |
| **Assumptions (optional)** | Enter your response here. |
| **Exploratory analyses (optional)** | Enter your response here. |
| **Analysis scripts (optional)** | Enter your response here. |

## Other

| | |
|---|---|
| **Other (Optional)** | Enter your response here. |

## References

@Manual{, title = {roadoi: Find Free Versions of Scholarly Publications via Unpaywall}, author = {Najko Jahn}, year = {2018}, note = {R package version 0.5.1}, url = {https://CRAN.R-project.org/package=roadoi}, }