

Preregistration of secondary data analysis: A template and tutorial

Olmo R. van den Akker

Department of Methodology and Statistics, Tilburg University

Lorne Campbell

Department of Psychology, University of Western Ontario

Rodica Ioana Damian

Department of Psychology, University of Houston

Andrew N. Hall

Department of Psychology, Northwestern University

Elliott Kruse

EGADE Business School, Tec de Monterrey

Stuart J. Ritchie

Social, Genetic and Developmental Psychiatry Centre, King's College

Anna E. van 't Veer

Methodology and Statistics Unit, Institute of Psychology, Leiden University

Sara J. Weston

Department of Psychology, University of Oregon

William J. Chopik

Department of Psychology, Michigan State University

Pamela E. Davis-Kean

Department of Psychology, University of Michigan

Jessica E. Kosie

Department of Psychology, Princeton University

Jerome Olsen

Department of Applied Psychology: Work, Education and Economy, University of Vienna Max Planck Institute for Research on Collective Goods

K. D. Valentine

Division of General Internal Medicine, Massachusetts General Hospital

Marjan Bakker

Department of Methodology and Statistics, Tilburg University

Preregistration has been lauded as one of the solutions to the so-called 'crisis of confidence' in the social sciences and has therefore gained popularity in recent years. However, the current guidelines for preregistration have been developed primarily for studies where new data will be collected. Yet, preregistering secondary data analyses--- where new analyses are proposed for existing data---is just as important, given that researchers' hypotheses and analyses may be biased by their prior knowledge of the data. The need for proper guidance in this area is especially desirable now that data is increasingly shared publicly. In this tutorial, we present a template specifically designed for the preregistration of secondary data analyses and provide comments and a worked example that may help with using the template effectively. Through this illustration, we show that completing such a template is feasible, helps limit researcher degrees of freedom, and may make researchers more deliberate in their data selection and analysis efforts.

Keywords: preregistration, secondary data analysis

Preregistration has been lauded as one of the key solutions to the replication crisis in the social sciences, mainly because it has the potential to prevent p-hacking by restricting researcher degrees of freedom, but also because it improves transparency and study planning, and can reduce publication bias. However, despite its growing popularity, preregistration is still in its infancy and preregistration practices are far from optimal (Claesen, Gomes, Tuerlinckx, & Vanpaemel, 2019; Veldkamp et al., 2018). Moreover, the current guidelines for preregistration are primarily relevant for studies in which new data will be collected. In this paper, we suggest that preregistration is also attainable when testing new hypotheses with pre-existing data and provide a tutorial on how to effectively preregister such secondary data analyses.

Secondary data analysis involves the analysis of existing data to investigate research questions, often in addition to the main ones for which the data were originally gathered (Grady, Cummings, & Hulley, 2013). Analyzing these datasets comes with its own challenges (Cheng & Phillips, 2014; Smith et al., 2011). For instance, common secondary datasets often include many different variables from many different respondents, sometimes measured at different points in time (e.g., the World Values Survey, Inglehart et al., 2014; the Wisconsin Longitudinal Study, Herd, Carr, & Roan, 2014). This provides ample opportunity for researchers to p-hack and increases the likelihood of obtaining spurious statistically significant results (Weston, Ritchie, Rohrer, & Przybylski, 2019).

In addition, because secondary data are often extensive and difficult to collect initially, researchers frequently analyze the same dataset multiple times to answer different research questions. Researchers are therefore not likely to come to a dataset with completely fresh eyes, and may have insight regarding associations between at least some of the variables in the dataset. Such prior knowledge may steer the researchers toward a hypothesis that they already know is in line with the data. This practice is called HARKing (Hypothesizing After Results Are Known; Kerr, 1998) and can lead to false positive results (Rubin, 2017). If HARKing goes undisclosed, it is not possible for third parties to evaluate whether the statistical tests for the hypotheses are well founded, as statistical hypothesis tests (e.g., null hypothesis significance tests, NHST) are only valid when the hypotheses are drawn up a priori (Wagenmakers, Wetzels, Borsboom, Van der Maas, & Kievit, 2012; but see Devezer et al., 2020).

Because secondary data analyses are particularly sensitive to data-driven researcher decisions, preregistering them is especially important. Other options exist to increase error control and illustrate sensitivity to flexibility in data analysis, however. For example, a multiverse analysis (Steegeen, Tuerlinckx, Gelman, & Vanpaemel, 2016) or specification-curve analysis (Simonsohn, Simmons, & Nelson, 2015) would be useful if researchers are unsure about which specific analysis is most suitable to test their hypothesis. In these approaches, all plausible analytic specifications are implemented to get an overall picture of the evidence without the need to choose a specific (and potentially biased) statistical analysis. This makes it impossible for researchers to cherry-pick variables or analyses based on their prior knowledge. However, it would still be possible to cherry-pick the range of analyses, and it is difficult to weight the results from the different analyses in an unbiased manner. It would thus be appropriate to complement these methods with a preregistration, especially when the aim is to limit the potential for p-hacking and HARKing, for both primary and secondary data analysis.

To facilitate the preregistration of secondary data analyses, a session was organized at the Society for the Improvement of Psychological Science (SIPS, see <https://improvingpsych.org>) conference in 2018 with the aim of creating an expert-generated preregistration template specifically tailored to secondary data analysis. Providing guidance on how to preregister is vital as preregistration is hard and requires practice and effort to be effective (Nosek et al., 2019). Participants in the session were experts on or had experience with secondary data analysis, preregistration, or both, thereby providing a good mix of expertise for the task at hand. The session began with analyzing the standard OSF Preregistration template (Bowman et al., 2016) and through successive rounds of discussion and testing, participants decided whether items could be edited, omitted, or added to make the template suitable for secondary data analysis. The resulting first draft of the template was further improved in the months following the conference through a digital back and forth involving the preregistration of an actual secondary data analysis. These efforts---the generation of the template and the preregistration of an example analysis---culminated in the preregistration template presented here.

Specific templates like this can greatly facilitate preregistration as it gives the author guidance about what to include in the preregistration so that

all researcher degrees of freedom are covered (Veldkamp et al., 2018). As such, the template would also be well-suited as a framework for a registered report submission that focuses on secondary data. Some of the questions in the preregistration template for secondary data analysis are similar to the questions in more 'traditional' templates; others aim to solve the challenges unique to the preregistration of secondary data analysis, such as the increased need for transparency about the process leading up to the preregistration.

The template presented here is not the only preregistration template for secondary data analysis. Mertens and Kryptos (2019) simultaneously developed a template consisting of 10 questions based on the AsPredicted template (see <https://aspredicted.org>). Our template differs from that template in two ways. First, it involves 25 questions and therefore captures a wider array of researcher degrees of freedom. For example, our template includes specific questions about defining and handling outliers, and the specification of robustness checks, both of which give leeway for data-driven decisions in secondary data analyses (Weston et al., 2019). Moreover, a more comprehensive template gives researchers the option to use as many or as few of the questions as they want, in order to tailor their preregistration to specific study needs. Second, our template comes with elaborate comments and a worked example that we hope makes the preregistration of secondary data analysis more concrete. We think both these contributions are helpful to researchers looking to preregister their secondary data analysis.

Using the template to preregister a secondary data analysis: Template questions and example answers, with guiding comments in italics

Part 1: Study information

Question 1: Provide the working title of your study.

Do religious people follow the golden rule? Assessing the link between religiosity and prosocial behavior using data from the Wisconsin Longitudinal Study.

We specifically mention the data set we are using so that readers know we are preregistering a secondary data analysis. Clarifying this from the outset is helpful because readers may look at such

preregistrations differently than they look at preregistrations of primary data analyses.

Question 2: Name the authors of this preregistration.

Josiah Carberry (JC) – ORCID iD:
<https://orcid.org/0000-0002-1825-0097>

Pomona Sprout (PS) – Personal webpage:
https://en.wikipedia.org/wiki/Hogwarts_staff#Pomona_Sprout

When listing the authors, add an ORCID iD or a link to a personal webpage so that you and your co-authors can be easily identified. This is particularly important when preregistering secondary data analyses because you may have prior knowledge about the data that may influence the contents of the preregistration. If a reader has access to a personal profile that lists prior research, they can judge whether any prior knowledge of the data is plausible and whether it potentially biased the data analysis. That is, whether it introduced systematic error in the testing because researchers selected or encouraged one outcome or answer over others (Merriam-Webster, n.d.).

Question 3: List each research question included in this study.

RQ1 = Are more religious people more prosocial than less religious people?

RQ2 = Does the relationship between religiosity and prosociality differ for people with different religious affiliations?

Research questions are often used as a steppingstone for the development of specific and testable hypotheses and can therefore be phrased on a more conceptual level than hypotheses. Note that it is perfectly fine to skip the research questions and only preregister your hypotheses.

Question 4: Please provide the hypotheses of your secondary data analysis. Make sure they are specific and testable, and make it clear what your statistical framework is (e.g., Bayesian inference, NHST). In case your hypothesis is directional, do not forget to state the direction. Please also provide a rationale for each hypothesis.

“Do to others as you would have them do to you” (Luke 6:31). This golden rule is taught by all ma-

for religions, in one way or another, to promote prosociality (Parliament of the World's Religions, 1993). Religious prosociality is the idea that religions facilitate behavior that is beneficial for others at a personal cost (Norenzayan & Shariff, 2008). The encouragement of prosocial behavior by religious teachings appears to be fruitful: a considerable amount of research shows that religion is positively related to prosocial behavior (e.g., Friedrichs, 1960; Koenig, McGue, Krueger, & Boucard, 2007; Morgan, 1983). For instance, religious people have been found to give more money to, and volunteer more frequently for, charitable causes than their non-religious counterparts (e.g., Grønberg & Never, 2004; Lazerwitz, 1962; Pharoah & Tanner, 1997). Also, the more important people viewed their religion, the more likely they were to do volunteer work (Youniss, McLellan, & Yates, 1999). Based on the above we expect that religiosity is associated with prosocial behavior in our sample as well. To assess this prediction, we will test the following hypotheses using a null hypothesis significance testing framework: H0(1) = In men and women who graduated from Wisconsin high schools in 1957, there is no association between religiosity and prosociality. H1(1) = In men and women who graduated from Wisconsin high schools in 1957, there is a positive association between religiosity and prosociality.

Just like in primary data analysis, a good hypothesis is specific (i.e., it includes a specific population), quantifiable, and testable. A one-sided hypothesis is suitable if theory, previous literature, or (scientific) reasoning indicates that your effect of interest is likely to be in a certain direction (e.g., $A < B$). Note that we provided detailed information about the theory and previous literature in our answer. This is crucial for secondary data analysis because it allows the reader to assess the thought process behind the hypotheses. Readers can then judge for themselves whether they think the hypotheses logically follow from the theory and previous literature or that they may have been tainted by the authors' prior knowledge of the data. Ideally, your preregistration already contains the framework for the introduction of the final paper. Moreover, writing up the introduction now instead of post hoc forces you to think clearly about the way you arrived at the hypotheses and may uncover flaws in your reasoning that can then be corrected before data collection begins.

Part 2: Data description

Question 5: Name and describe the dataset(s), and if applicable, the subset(s) of the data you plan to use. Useful information to include here is the type of data (e.g., cross-sectional or longitudinal), the general content of the questions, and some details about the respondents. In the case of longitudinal data, information about the survey's waves is useful as well.

To answer our research questions we will use a dataset from the Wisconsin Longitudinal Study (WLS; Herd, Carr, & Roan, 2014). The WLS provides long-term data on a random sample of all the men and women who graduated from Wisconsin high schools in 1957. The WLS involves twelve waves of data. Six waves were collected from the original participants or their parents (1957, 1964, 1975, 1992, 2004, and 2011), four were collected from a selected sibling (1977, 1994, 2005, and 2011), one from the spouse of the original participant (2004), and one from the spouse of the selected sibling (2006). The questions vary across waves and are related to domains as diverse as socio-economic background, physical and mental health, and psychological makeup. We will use the subset consisting of the 1957 graduates who completed the follow-up 2003-2005 wave of the WLS dataset because it includes specific modules on religiosity and volunteering.

Like the WLS data we use in our example, many large-scale datasets are outlined in detail in an accompanying paper. It is important to cite papers like this, but also to mention the most relevant information in the preregistration so that readers do not have to search for the information themselves. Sometimes information about the dataset is not readily available. In those cases, be especially candid with the information you have about the dataset because the data you provide may be the only information about the data available to readers of the preregistration.

Question 6: Specify the extent to which the dataset is open or publicly available. Make note of any barriers to accessing the data, even if it is publicly available.

The dataset we will use is publicly available, but you need to formally agree to acknowledge the funding source for the Wisconsin Longitudinal Study, to cite the data release in any manuscripts,

working papers, or published articles using these data, and to inform WLS about any published papers for use in the WLS bibliography and for reporting purposes. To do this you need to submit some information about yourself on the website: (<https://www.ssc.wisc.edu/wlsresearch/data/downloads>). You will then receive an email with a download link.

It is important to check whether the data is open or publicly available also to other researchers. For example, it could be that you have access via the organization providing the data (explain this in your answer to Q7), but that does not necessarily mean that it is publicly available to others. An example of publicly available data that is difficult to access would be data for which you need to register a profile on a website, or for which the owners of the data need to accept your request before you can have access.

Question 7: How can the data be accessed? Provide a persistent identifier or link if the data are available online or give a description of how you obtained the dataset.

The data can be accessed by going to the following link and searching for the variables that are specified in Q12 of this preregistration: https://www.ssc.wisc.edu/wlsresearch/documentation/browse/?label=&variable=&wave_108=on&searchButton=Search

When available, report the dataset's persistent identifier (e.g., a DOI) so that the data can always be retrieved from the Internet. In our example, we could only provide a link, but we added instructions for the reader to retrieve the data. In general, try to bring the reader as close to the relevant data as possible, so instead of giving the link to the overarching website, give the link to the part of the website where the data can easily be located.

Question 8: Specify the date of download and/or access for each author.

PS: Downloaded 12 February 2019; Accessed 12 February 2019.

JC: Downloaded 3 January 2019 (estimated); Accessed 12 February 2019.

We will use the data accessed by JC on 12 February 2019 for our statistical analyses.

State here for each author when the dataset was initially downloaded (e.g., for previous analyses or merely to obtain the data) and when either metadata or the actual data (specify which) was first accessed (e.g., to identify variables of interest or to help fill out this form). Also, specify the author whose downloaded data you will use for the statistical analyses. This information is crucial in light of the reproducibility of your study because it is possible that the data has been edited since you last downloaded or accessed it. If you cannot retrieve when you downloaded or accessed the data, estimate those dates. In case you collected the data yourself to answer another research question, please state the date you first looked at the data. Finally, because not everybody will use the same date format it is important to state the date you downloaded or accessed the data unambiguously. For example, avoid dates like 12/02/2019 and instead use 12 February 2019 or December 2nd, 2019.

Question 9: If the data collection procedure is well documented, provide a link to that information. If the data collection procedure is not well documented, describe, to the best of your ability, how data were collected.

The WLS data was and is being collected by the University of Wisconsin Survey Center for use by the research community. The origins of the WLS can be traced back to a state-sponsored questionnaire administered during the spring of 1957 at all Wisconsin high school to students in their final year. Therefore, the dataset constitutes a specific sample not necessarily representative of the United States as a whole. Most panel members were born in 1939, and the sample is broadly representative of white, non-Hispanic American men and women who completed at least a high school education. A flowchart for the data collection can be found here: <https://www.ssc.wisc.edu/wlsresearch/about/flowchart/cor459d7.pdf>

While describing the data collection procedure, pay specific attention to the representativeness of the sample, and possible biases stemming from the data collection. For example, describe the population that was sampled from, whether the aim was to acquire a representative / regional / convenience sample, whether the data collectors were aware of this aim, the data collectors' recruitment efforts, the procedure for running participants,

whether randomization was used, and whether participants were compensated for their time. All of this information can be used to judge whether the sample is representative of a wider population or whether the data is biased in some way, which crucially determines the conclusions that can be drawn from the results. In addition, thinking about the representativeness of a dataset is a crucial part of the planning stage of the research. For example, you might come to the conclusion that the dataset at hand is not suitable after all and opt for a different dataset, thereby preventing research waste. Finally, it is good practice to describe what entity originally collected the data (e.g., your own lab, another lab, a multi-lab collaboration, a (national) survey collection organization, a private organization) because different data sources may have different purposes for collecting the data, which may also result in biased data.

Question 10: Some studies offer codebooks to describe their data. If such a codebook is publicly available, link to it here or upload the document. If not, provide other available documentation. Also provide guidance on what parts of the codebook or other documentation are most relevant.

The codebook for the dataset we use can be found here: <https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k>. We will mainly use questions from the mail survey about religion and spirituality, and the phone survey on volunteering, but will also use some questions from other modules (see the answer to Q12).

Any documentation is welcome here, as readers will use this documentation to make sense of the dataset. If applicable, provide the codebook for the entire dataset but guide the reader to the relevant parts of the codebook so they do not have to search for the relevant parts extensively. Alternatively, you can create your own data dictionaries/codebooks (Arslan, 2019; Buchanan et al., 2019). If, for some reason codebook information cannot be shared publicly, provide an explanation.

Part 3: Variables

Question 11: If you are going to use any manipulated variables, identify them here. Describe the variables and the levels or treatment arms of each

variable (note that this is not applicable for observational studies and meta-analyses). If you are collapsing groups across variables this should be explicitly stated, including the relevant formula. If your further analysis is contingent on a manipulation check, describe your decisions rules here.

Not applicable.

Manipulated variables in secondary datasets usually originate from another study investigating another research question. You may, therefore, need to adapt the manipulated variable to answer your own research question. For example, it may be necessary to relabel or even omit one of the treatment arms. Please provide a careful log of all these adaptations so that readers will have a clear grasp of the variable you will be using and how it differs from the variable in the original dataset. Any resources mentioned in the answer to Q10 may be useful here as well.

Question 12: If you are going to use measured variables, identify them here. Describe both outcome measures as well as predictors and covariates and label them accordingly. If you are using a scale or an index, state the construct the scale/index represents, which items the scale/index will consist of, how these items will be aggregated, and whether this aggregation is based on a recommendation from the study codebook or validation research. When the aggregation of the items is based on exploratory factor analysis (EFA) or confirmatory factor analysis (CFA), also specify the relevant details (EFA: rotation, how the number of factors will be determined, how best fit will be selected, CFA: how loadings will be specified, how fit will be assessed, which residuals variance terms will be correlated). If you are using any categorical variables, state how you will code them in the statistical analyses.

Religiosity (IV): Religiosity is measured using a newly created scale with a subset of items from the Religion and Spirituality module of the 2004 mail survey (described here: https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gmail_religion). The scale includes general questions about how religious/spiritual the individual is and how important religion/spirituality is to them. Importantly, the questions are not specific to a particular denomination and are on the

same response scale. The specific variables are as follows:

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?
7. il007rer: How important is it for you to be with other people who are the same religion as you?
8. il008rer: How important do you think it is for people of your religion to marry other people who are the same religion?
9. il009rer: How strongly do you believe that one should stick to a particular faith?
10. il010rer: How important was religion in your home when you were growing up?
11. il011rer: When you have important decisions to make in your life, how much do you rely on your religious or spiritual beliefs?
12. il012rer: How much would your spiritual or religious beliefs influence your medical decisions if you were to become gravely ill?

The levels of all of these variables are indicated by a Likert scale with the following options: (1) Not at all; (2) Not very; (3) Somewhat; (4) Very; (5) Extremely, as well as 'System Missing' (the participant did not provide an answer) and 'Refused' (the participant refused to answer the question). Variables il006rer, il008rer, and il012rer additionally include the option 'Don't know' (the participant stated that they did not know how to answer the question). We will use the average score (after omitting non-numeric and 'Don't know' responses) on the twelve variables as a measure of religiosity. This average score is constructed by ourselves and was not already part of the dataset.

Prosociality (DV): In line with previous research (Konrath, Fuhrel-Forbis, Lou, & Brown, 2012), we will use three measures of prosociality that measure three aspects of engagement in other-oriented activities (see Brookfield, Parry, & Bolton, 2018 for the link between prosociality and volunteering). The prosociality variables come from the Volunteering module of the 2004 phone survey. The codebook

of that module can be found here: <https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gvol>. The three measures of prosociality we will use are:

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

This dichotomous variable assesses whether or not the participant has engaged in any volunteering activities in the last 12 months. The levels of this variable are yes/no. Yes will be coded as '1', no will be coded as '0'.

2. gv109re: Number of graduate's other volunteer activities in the past 12 months.

This variable is a summary index providing a quantitative measure of the participant's volunteering activities. Scores on this variable range from 1 to 5 and reflect the number of the previous five questions to which the participant answered YES. The previous five questions assess whether or not the participant volunteered at any of the following organization types: (1) religious organizations; (2) school or educational organization; (3) political group or labor union; (4) senior citizen group or related organization; (5) other national or local organizations. For each of these questions the answer 'yes' is coded as 1 and the answer 'no' is coded as 0.

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

This is a numerical variable that provides information on how many hours per month, on average, the participant volunteered.

The three variables will be treated as separate measures in the dataset and do not require manual aggregation.

Number of Siblings (Covariate): We will include the participant's number of siblings as a control variable because many religious families are large (Pew Research Center, 2015) and it can be argued that cooperation and trust arise more naturally in larger families because of the larger number of social interactions in those families. To measure participants' number of siblings we used the variable gk067ss: The total number of siblings ever born from the 2004 phone survey Siblings module (see: <https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gsib>). This is a numerical variable with the possibility for the participant to state "I don't know". At the interview participants were instructed to include "siblings born alive but no longer living, as well as those alive now

and to include step-brothers and step-sisters and children adopted by their parents.”

Agreeableness (Covariate): We will include the summary score for agreeableness (ih009rec, see https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gmail_values) in the analysis as a control variable because a previous study (on the same dataset, see the answer to Q18) we were involved in showed a positive association between agreeableness and prosociality. Because previous research also indicates a positive association between agreeableness and religiosity (Saroglou, 2002) we need to include agreeableness as a control variable to disentangle the influence of religiosity on prosociality and the influence of agreeableness on prosociality. The variable ih009rec is a sum score of the variables ih003rer-ih008rer (To what extent do you agree that you see yourself as someone who is talkative / is reserved [reverse coded] / is full of energy / tends to be quiet [reverse coded] / who is sometimes shy or inhibited [reverse coded] / who generates a lot of enthusiasm). All of these were scored from 1 to 6 (1 = “agree strongly”, 2 = “agree moderately”, 3 = “agree slightly”, 4 = “disagree slightly”, 5 = “disagree moderately”, 6 = “disagree strongly”), while participants could also refuse to answer the question. If a participant refused to answer one of the questions, that participant’s score was not included in the sum score variable ih009rec.

If you are using measured variables, describe them in such a way that readers know exactly what variables will be used in the statistical analyses. Because secondary datasets often involve many measured variables, there is ample room to select variables after doing an analysis. It is therefore essential to be exhaustive here. Variables you do not mention here should not pop up in your analysis later unless you have a good reason for it. As you can see, we clearly label the function of each variable, the specific items related to that variable, and the item’s response options. It could be that you choose to combine items into an index or scale that have not been combined like that in previous studies. Carefully detail this process and indicate that you constructed the index or scale yourself to avoid confusion. Finally, note that we include covariates to be able to make statements about the causal effect of religion on prosociality. This is common practice in the social sciences, but causal inference is complex

and there may be better solutions in other situations, and even in this situation. Please see Rohrer (2018) for more information about causation in observational data.

Question 13: Which units of analysis (respondents, cases, etc.) will be included or excluded in your study? Taking these inclusion/exclusion criteria into account, indicate the (expected) sample size of the data you’ll be using for your statistical analyses to the best of your knowledge. In the next few questions, you will be asked to refine this sample size estimation based on your judgments about missing data and outliers.

Initially, the WLS consisted of 10,317 participants. As we are not interested in a specific group of Wisconsin people, we will not exclude any participants from our analyses. However, only 7,265 participants filled out the questions on prosociality and the number of siblings in the phone survey and only 6,845 filled out the religiosity items in the mail survey (Herd et al., 2014). This corresponds to a response rate of 73% and 69% respectively. Because we do not know whether the participants that the mail survey also did the phone survey, our minimum expected sample size is $10,317 * 0.73 * 0.69 = 5,297$.

Provide information on the total sample size of the dataset, the sample size(s) of the wave(s) you are going to use (if applicable), as well as the number of participants that provided data on each of the questions and/or scales to be used in the data analyses. In our sample we do not exclude any participants, but if you have a research question about a certain group you may need to exclude participants based on one or more characteristics. Be very specific when describing these characteristics so that readers with no knowledge of the data are able to redo your moves easily. For our WLS dataset, it is impossible to know the exact sample size without inspecting the data. If that is the case, provide an estimate of the sample size. If you provide an estimate, try to be conservative and pick the lowest sample size of the possible options. If it is impossible to provide an estimate, it is also possible to mask the data. For example, it is possible to add random noise to all values of the dependent variable. In that case, it is impossible to pick up any real effects and you are essentially blind to the data. Similarly, it is possible to blind yourself to real effects in the data by having someone relabel the treatment levels so you cannot link them to the treatment levels anymore. These

and other methods of data blinding are clearly described by Dutilh, Sarafoglou, and Wagenmakers (2019).

Question 14: What do you know about missing data in the dataset (i.e., overall missingness rate, information about differential dropout)? How will you deal with incomplete or missing data? Based on this information, provide a new expected sample size.

The WLS provides a documented set of missing codes. In Table 1 below you can find missingness information for every variable we will include in the statistical analyses. ‘System missing’ refers to the number of participants that did not or could not complete the questionnaire. ‘Partial interview’ refers to the number of participants that did not get that particular question because they were only partially interviewed. The rest of the codes are self-explanatory.

Importantly, some respondents refused to answer the religiosity questions. These respondents apparently felt strongly about these questions, which could indicate that they are either very religious or very anti-religious. If that is the case, the respondent’s propensity to respond is directly associated with their level of religiosity and that the data is missing not at random (MNAR). Because it is not possible to test the stringent assumptions of the modern techniques for handling MNAR data we will resort to simple listwise deletion. It must be noted that this may bias our data as we may lose respondents who are very religious or anti-religious. However, we believe this bias to be relatively harmless given that our sample still includes many respondents that provided extreme responses to the items about the importance of the different facets of religion (see: https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gmail_religion). Moreover, because our initial sample size is very large, statistical power is not substantially compromised by omitting these respondents. That being said, we will extensively discuss any potential biases resulting from missing data in the limitations section of our paper.

Employing listwise deletion leads to an expected minimum number of $10,317 * 0.30 * 0.70 * 0.64 = 1,387$ participants for the binary logistic regression, and an expected minimum number of $10,317 * 0.24 * 0.70 * 0.64 = 1,109$ (gv109re) and $10,317 * 0.23 * 0.70 * 0.64 = 1,063$ (gv111re) for the linear regressions.

Provide descriptive information, if available, on the amount of missing data for each variable you will use in the statistical analyses and discuss potential issues with the pattern of missing data for your planned analyses. Also provide a plan for how the analyses will take into account the presence of missing data. Where appropriate, provide specific details how this plan will be implemented. This can be done by specifying a step-by-step protocol for how you will impute any missing data. You could first explain how you will assess whether the data are missing at random (MAR) missing completely at random (MCAR) or missing not at random (MNAR), and then state that you will use technique X in case of MAR data, technique Y in case of MCAR data, and technique Z in case of MNAR data. For an overview of the types of missing data, and the different techniques to handle missing data, see Lang & Little (2018). Note that the missing data technique we used in our example, listwise deletion, is usually not the best way to handle missing data. We decided to use it in this example because it gave us the opportunity to illustrate how researchers can describe potential biases arising from their analysis methods in a preregistration.

If you cannot specify the exact number of missing data because the dataset does not provide that information, provide an estimate. If you provide an estimate, try to be conservative and pick the lowest sample size of the possible options. If it is impossible to provide an estimate, you could also mask the data (see Dutilh, Sarafoglou, & Wagenmakers, 2019). It is good practice to state all missingness information with relation to the total sample size of the dataset.

Table 1

An overview of the missing values for all variables we will use in our analyses.

Variable	System missing	Don't know	Inappropriate	Refused	Not ascertained	Partial interview	Could not code	Remaining	Remaining (%)
il001rer	3,471	0	0	190	0	0	0	6,656	64
il002rer	3,471	0	0	212	0	0	0	6,634	64
il003rer	3,471	0	0	191	0	0	0	6,655	65
il004rer	3,471	0	0	241	0	0	0	6,605	64
il005rer	3,471	0	0	201	0	0	0	6,645	64
il006rer	3,471	1	0	201	0	0	0	6,644	64
il007rer	3,471	0	0	192	0	0	0	6,654	65
il008rer	3,471	1	0	199	0	0	0	6,646	64
il009rer	3,471	0	0	219	0	0	0	6,627	64
il010rer	3,471	0	0	190	0	0	0	6,656	65
il011rer	3,471	0	0	190	0	0	0	6,656	65
il012rer	3,471	1	0	198	0	0	0	6,647	64
gv103re	3,052	0	3,955	1	0	182	0	3,127	30
gv109re	3,052	0	4,590	0	0	182	0	2,493	24
gv111re	3,052	50	4,716	0	0	182	0	2,317	23
gk067ss	3,052	21	0	0	0	0	0	7,244	70

Question 15: If you plan to remove outliers, how will you define what a statistical outlier is in your data? Please also provide a new expected sample size. Note that this will be the definitive expected sample size for your study and you will use this number to do any power analyses.

The dataset probably does not involve any invalid data since the dataset has been previously ‘cleaned’ by the WLS data controllers and any clearly unreasonably low or high values have been removed from the dataset. However, to be sure we will create a box and whisker plot for all continuous variables (the dependent variables gv109re and gv111re, the covariate gk067ss, and the scale for religiosity) and remove any data point that appears to be more than 1.5 times the IQR away from the 25th and 75th percentile. Based on normally distributed data, we expect that 2.1% of the data points will be removed this way, leaving 1,358 out of 1,387 participants for the binary regression with gv103re as the outcome variable and 1,086 out of 1,109 participants, and 1,041 out of 1,063 participants for the linear regressions with gv109re and gv111re as the outcome variables, respectively.

Estimate the number of outliers you expect for each variable and calculate the expected sample size of your analysis. The expected sample size is required to do a power analysis for the planned statistical tests (Q21) but also prevents you from discarding a significant portion of the data during or after the statistical analysis. If it is impossible to provide such an estimate, you can mask the data and make a more informed estimation based on these masked data (see Dutilh, Sarafoglou, & Wagenmakers, 2019). If you expect to remove many outliers or if you are unsure about your outlier handling strategy, it is good practice to preregister analyses including and excluding outliers. To see how decisions about outliers can influence the results of a study, see Bakker and Wicherts (2014) and Lonsdorf et al. (2019). For more information about outliers in the context of preregistration, see Leys, Delacre, Mora, Lakens, and Ley (2019).

Question 16: Are there sampling weights available with this dataset? If so, are you using them or are you using your own sampling weights?

The WLS dataset does not include sampling weights and we will not use our own sampling

weights as we do not seek to make any claims that are generalizable to the national population.

Because secondary data samples may not be entirely representative of the population you are interested in, it can be useful to incorporate sampling weights into your analysis. You should state here whether (and why) you will use sampling weights, and provide specifics on exactly how you will use them. To implement sampling weights into your analyses, we recommend using the “survey” package in R (Lumley, 2004).

Part 4: Knowledge of data

Question 17: List the publications, working papers (in preparation, unpublished, preprints), and conference presentations (talks, posters) you have worked on that are based on the dataset you will use. For each work, list the variables you analyzed, but limit yourself to variables that are relevant to the proposed analysis. If the dataset is longitudinal, also state which wave of the dataset you analyzed.

Importantly, some of your team members may have used this dataset, and others may not have. It is therefore important to specify the previous works for every co-author separately. Also mention relevant work on this dataset by researchers you are affiliated with as their knowledge of the data may have been spilled over to you. When the provider of the data also has an overview of all the work that has been done using the dataset, link to that overview.

Both authors (PS and JC) have previously used the Graduates 2003–2005 wave to assess the link between Big Five personality traits and prosociality. The variables we used to measure the Big Five personality traits were ih001rei (extraversion), ih009rei (agreeableness), ih017rei (conscientiousness), ih025rei (neuroticism), and ih032rei (openness). The variables we used to measure prosociality were ih013rer (“To what extent do you agree that you see yourself as someone who is generally trusting?”), ih015rer (“To what extent do you agree that you see yourself as someone who is considerate to almost everyone?”), and ih016rer (“To what extent do you agree that you see yourself as someone who likes to cooperate with others?”). We presented the results at the ARP conference in St. Louis in 2013 and we are currently finalizing a manuscript based on these results.

Additionally, a senior graduate student in JC's lab used the Graduates 2011 wave for exploratory analyses on depression. She linked depression to alcohol use and general health indicators. She did not look at variables related to religiosity or prosociality. Her results have not yet been submitted anywhere.

An overview of all publications based on the WLS data can be found here: <https://www.ssc.wisc.edu/wlsresearch/publications/pubs.php?topic=ALL>.

It is important to specify the different ways you have previously used the data because this information helps you to establish any knowledge of the data you may already have. This prior knowledge will need to be provided in Q18. If available, include persistent identifiers (e.g., a DOI) to any relevant papers and presentations.

Understandably, there is a subjectivity involved in determining what constitutes "relevant" work or "relevant" variables for the proposed analysis. We advise researchers to use their professional judgment and when in doubt always mention the work or variable so readers can assess their relevance themselves. In the worked example, the exploratory analysis by the student in JC's lab is probably not relevant, but because of the close affiliation of the student to JC, it is good to include it anyway.

Listing previous works based on the data also helps to prevent a common practice identified by the American Psychological Association (2019) as unethical: the so-called "least publishable unit" practice (also known as "salami-slicing"), in which researchers publish multiple papers on closely related variables from the same dataset. Given that secondary datasets often involve many closely related variables, this is a particularly pernicious issue here.

Question 18: What prior knowledge do you have about the dataset that may be relevant for the proposed analysis? Your prior knowledge could stem from working with the data first-hand, from reading previously published research, or from codebooks. Also provide any relevant knowledge of subsets of the data you will not be using. Provide prior knowledge for every author separately.

In a previous study (mentioned in Q17) we used three prosociality variables (ih013rer, ih015rer, and ih016rer) that may be related to the prosociality variables we use in this study. We found that

ih013rer, ih015rer, and ih016rer are positively associated with agreeableness (ih009rec). Because previous research (on other datasets) shows a positive association between agreeableness and religiosity (Saroglou, 2002) agreeableness may act as a confounding variable. To account for this we will include agreeableness in our analysis as a control variable. We did not find any associations between prosociality and the other Big Five variables.

It is important to denote your prior knowledge diligently because it provides information about possible biases in your statistical analysis decisions. For example, you may have learned at an academic conference or in a footnote of another paper that the correlation between two variables is high in this dataset. If you do a test of this hypothesis, you already know the test result, making the interpretation of the test invalid (Wagenmakers, et al., 2012). In cases like this, where you have direct knowledge about a hypothesized association, you should disregard doing a confirmatory analysis altogether or do one based on a different dataset.

Any indirect knowledge about the hypothesized association does not preclude a confirmatory analysis but should be transparently reported in this section. In our example, we mentioned that we know about the positive association between agreeableness and prosociality, which may say something about the direction of our hypothesized association given the association between agreeableness and religiosity. Moreover, this prior knowledge urged us to add agreeableness as a control variable. Thus, aside from improving your preregistration, evaluating your prior knowledge of the data can also improve the analyses themselves.

All information like this that may influence the hypothesized association is relevant here. For example, restriction of range (Meade, 2010), measurement reliability (Silver, 2008), and the number of response options (Gradstein, 1986) have been shown to influence the association between two variables. You may have provided univariate information regarding these aspects in previous questions. In this section, you can write about how they may affect your hypothesized association.

Do note that it is unlikely that you are able to account for all the effects of prior knowledge on your analytical decisions. For example, you may have prior knowledge that you are not consciously aware of. The best way to capture this unconscious prior knowledge is to revisit previous work, think deeply about any information that might be relevant for the

current project, and present it here to the best of your ability. This exercise helps you reflect on potential biases you may have and makes it possible for readers of the preregistration to assess whether the prior knowledge you mentioned is plausible given the list of prior work you provided in Q17.

Of course, it is still possible that researchers purposefully neglect to mention prior knowledge or provide false information in a preregistration. Even though we believe that deliberate deceit like this is rare, at the end of our template we require researchers to formally “promise” to have truthfully filled out the template and that no other preregistration exists on the same hypotheses and data. A violation of this formal statement can be seen as misconduct, and we believe researchers are unlikely to cross that line.

Part 5: Analyses

Question 19: For each hypothesis, describe the statistical model you will use to test the hypothesis. Include the type of model (e.g., ANOVA, multiple regression, SEM) and the specification of the model. Specify any interactions and post-hoc analyses and remember that any test not included here must be labeled as an exploratory test in the final paper.

Our first hypothesis will be tested using three analyses since we use three variables to measure prosociality. For each, we will run a directional null hypothesis significance test to see whether a positive effect exists of religiosity on prosociality. For the first outcome (gv103re: Did the graduate do volunteer work in the last 12 months?) we will run a logistic regression with religiosity, the number of siblings, and agreeableness as predictors. For the second and third outcomes (gv109re: Number of graduate’s other volunteer activities in the past 12 months; gv111re: How many hours did the graduate volunteer during a typical month in the last 12 months?) we will run two separate linear regressions with religiosity, the number of siblings, and agreeableness as predictors. The code we will use for all these analyses can be found at <https://osf.io/e3htr>.

Think carefully about the variety of statistical methods that are available for testing each of your hypotheses. One of the classic “Questionable

Research Practices” is trying multiple methods and only publishing the ones that “work” (i.e., that support your hypothesis). Almost every method has several options that may be more or less suited to the question you are asking. Therefore, it is crucial to specify a priori which one you are going to use and how.

If you can, include the code you will use to run your statistical analyses, as this forces you to think about your analyses in detail and makes it easy for readers to see exactly what you plan to do. Ideally, when you have loaded the data in a software program you only have to press one button to run your analyses. If including the code is impossible, describe the analyses such that you could give a positive answer to the question: “Would a colleague who is not involved in this project be able to recreate this statistical analysis?”

Question 20: If applicable, specify a predicted effect size or a minimum effect size of interest for all the effects tested in your statistical analyses.

For the logistic regression with ‘Did the graduate do volunteer work in the last 12 months?’ as the outcome variable, our minimum effect size of interest is an odds of 1.05. This means that a one-unit increase on the religiosity scale would be associated with a 1.05 factor change in odds of having done volunteering work in the last 12 months versus not having done so.

For the linear regressions with ‘The number of graduate’s volunteer activities in the last 12 months’, and ‘How many hours did the graduate volunteer during a typical month in the last 12 months?’ as the outcome variables, the minimum regression coefficients of interest of the religiosity variables are 0.05 and 0.5, respectively. This means that a one-unit increase in the religiosity scale would be associated with 0.05 extra volunteering activities in the last 12 months and with 0.5 more hours of volunteering work in the last 12 months. All of these smallest effect sizes of interest are based on our own intuition.

To make comparisons possible between the effects in our study and similar effects in other studies the unstandardized linear regression coefficients will be transformed into standardized regression coefficients using the following formula: $\beta_i = \mathbf{B}_i(\mathbf{s}_i / \mathbf{s}_y)$, where \mathbf{B}_i the unstandardized regression coefficient of independent variable i , and \mathbf{s}_i and \mathbf{s}_y are the

standard deviations of the independent and dependent variable respectively.

Comment(s): A predicted effect size is ideally based on a representative preliminary study or meta-analytical result. If those are not available, it is also possible to use your own intuition. For advice on setting a minimum effect size of interest, see Lakens, Scheel, & Isager (2018) and Funder and Ozer (2019).

Question 21: Present the statistical power available to detect the predicted effect size(s) or the smallest effect size(s) of interest OR present the accuracy that will be obtained for estimation. Use the sample size after updating for missing data and outliers, and justify the assumptions and parameters used (e.g., give an explanation of why anything smaller than the smallest effect size of interest would be theoretically or practically unimportant).

The sample size after updating for missing data and outliers is 1,358 for the logistic regression with gv103re as the outcome variable, and 1,086 and 1,041 for the linear regressions with gv109re and gv111re as the outcome variables, respectively. For all three analyses this corresponds to a statistical power of approximately 1.00 when assuming our minimum effect sizes of interest. For the linear regressions we additionally assumed the variance explained by the predictor to be 0.2 and the residual variance to be 1.0 (see figure below for the full power analysis of the regression with the lowest sample size). For the logistic regression we assumed an intercept of -1.56 corresponding to a situation where half of the participants have done volunteer work in the last year (see the R-code for the full power analysis at <https://osf.io/f96rn>).

*Advice on conducting a power analysis using G*Power can be found in Faul, Erdfelder, Buchner, and Lang (2009). Advice on conducting a power analysis using R can be found here: cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html.*

Note that power analyses for secondary data analyses are unlike power analyses for primary data analyses because we already have a good idea about what our sample size is based on our answers to Q13, Q14, and Q15. Therefore, we are primarily interested in finding out what effect sizes we are able to find for a given power level or what our power is given our minimum effect size of interest. In our example,

*we chose the second option. When presenting your power analysis be sure to state the version of G*Power, R, or any other tool you calculated power with, including any packages or add-ons, and also report or copy all the input and results of the power analysis.*

Question 22: What criteria will you use to make inferences? Describe the information you will use (e.g., specify the p-values, effect sizes, confidence intervals, Bayes factors, specific model fit indices), as well as cut-off criteria, where appropriate. Will you be using one- or two-tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this, and if so, how?

We will make inferences about the association between religiosity and prosociality based on the p-values and the size of the regression coefficients of the religiosity variable in the three main regressions. We will conclude that a regression analysis supports our hypothesis if both the p-value is smaller than .01 and the regression coefficient is larger than our minimum effect size of interest. We chose an alpha of .01 to account for the fact that we do a test for each of the three regressions (0.05/3, rounded down). If the conditions above hold for all three regressions, we will conclude that our hypothesis is fully supported, if they hold for one or two of the regressions, we will conclude that our hypothesis is partially supported, and if they hold for none of the regressions we will conclude that our hypothesis is not supported.

It is crucial to specify your inference criteria before running a statistical analysis because researchers have a tendency to move the goalposts when making inferences. For example, almost 40% of p-values between 0.05 and 0.10 are reported as “marginally significant”, even though these values are not significant when compared to the traditional alpha level of 0.05, and the evidential value of these p-values is low (Olsson-Collentine, Van Assen, & Hartgerink, 2019). Similarly, several studies have found that the majority of studies reporting p-values do not use any correction for multiple comparisons (Cristea & Ioannidis, 2018; Wason, Stecher, & Mander, 2014), perhaps because this lowers the chance of finding a statistically significant result. For an overview of multiple-comparison correction methods relevant to secondary data analysis, see Thompson, Wright, Bissett, and Poldrack (2019).

Question 23: What will you do should your data violate assumptions, your model not converge, or some other analytic problem arises?

When the distribution of the number of volunteering hours (gv111re) is significantly non-normal according to the Kolmogorov-Smirnov test (Massey, 1951), and/or (b) the linearity assumption is violated (i.e., the points are asymmetrically distributed around the diagonal line when plotting observed versus the predicted values), we will log-transform the variable.

It is, of course, impossible to predict every single way that things might go awry during the analysis. One of the variables may have a strange and unexpected distribution, one of the models may not converge because of a quirk of the correlational structure, and you may even encounter error messages that you have never seen before. You can use your prior knowledge of the dataset to set up a decision tree specifying possible problems that might arise and how you will address them in the analyses. Thinking through such a decision tree will make you less overwhelmed when something does end up going differently than expected.

However, note that decision trees come with their own problems and can quickly become very complex. Alternatively, you might choose to select analysis methods that make assumptions that are as conservative as possible; preregister robustness analyses which test the robustness of your findings to analysis strategies that make different assumptions; and/or pre-specify a single primary analysis strategy but note that you will also report an exploratory investigation of the validity of distributional assumptions (Williams & Albers, 2019). Of course, there are pros and cons to all methods of dealing with violations, and you should choose a technique that is most appropriate for your study.

Question 24: Provide a series of decisions about evaluating the strength, reliability, or robustness of your focal hypothesis test. This may include within-study replication attempts, additional covariates, cross-validation efforts (out-of-sample replication, split/hold-out sample), applying weights, selectively applying constraints in an SEM context (e.g., comparing model fit statistics), overfitting adjustment techniques used (e.g., regularization approaches such as ridge regression), or some other

simulation/sampling/bootstrapping method.

To assess the sensitivity of our results to our selection criterion for outliers, we will run an additional analysis without removing any outliers.

There are many methods you can use to test the limits of your hypothesis. The options mentioned in the question are not supposed to be exhaustive or prescriptive. We included these examples to encourage researchers to think about these methods, all of which serve the same purpose as preregistration: improving the robustness and replicability of the results.

Question 25: If you plan to explore your dataset to look for unexpected differences or relationships, describe those tests here, or add them to the final paper under a heading that clearly differentiates this exploratory part of your study from the confirmatory part.

As an exploratory analysis, we will test the relationship between scores on the religiosity scale and prosociality after adjusting for a variety of social, educational, and cognitive covariates that are available in the dataset. We have no specific hypotheses about which covariates will attenuate the religiosity-prosociality relation most substantially, but we will use this exploratory analysis to generate hypotheses to test in other, independent datasets.

Whereas it is not presently the norm to preregister exploratory analyses, it is often good to be clear about which variables will be explored (if any), for example, to differentiate these from the variables for which you have specific predictions or to plan ahead about how to compute these variables.

Part 6: Statement of integrity

The authors of this preregistration state that they filled out this preregistration to the best of their knowledge and that no other preregistration exists pertaining to the same hypotheses and dataset.

Summary

In this tutorial we presented a preregistration template for the analysis of secondary data and have provided guidance for its effective use. We are aware that the number of questions (25) in the template may be overwhelming, but it is important to note that not every question is relevant for every preregistration. Our aim was to be inclusive and cover all bases in light of the diversity of secondary data analyses. Even though none of the questions are mandatory, we do believe that an elaborate preregistration is preferable over a concise preregistration simply because it restricts more researcher degrees of freedom. We therefore recommend that authors answer as many questions in as much detail as possible. And, if questions are not applicable, it would be good practice to also specify why this is the case so that readers can assess your reasoning.

Effectively preregistering a study is challenging and can take a lot of time but, like Nosek et al. (2019) and many others, we believe it can improve the interpretability, verifiability and rigor of your studies and is therefore more than worth it if you want both yourself and others to have more confidence in your research findings.

The current template is merely one building block toward a more effective preregistration infrastructure and, given the ongoing developments in this area, will be a work in progress for the foreseeable future. Any feedback is therefore greatly appreciated. Please send any feedback to the corresponding author, Olmo van den Akker (ovdakker@gmail.com).

Author Contact

Correspondence concerning this article should be addressed to Olmo van den Akker, e-mail: ovdakker@gmail.com. <https://orcid.org/0000-0002-0712-3746>

Acknowledgments

The authors would like to thank the participants of the Society for the Improvement of Psychological Science (SIPS) conference in 2018 that helped to create the first draft of the preregistration template but were unable to help out with the subsequent extensions (Brian Brown, Oliver Clark, Charles Ebersole, and Courtney Soderberg).

Conflict of Interest and Funding

This research is based on data from the Wisconsin Longitudinal Study, funded by the National Institute on Aging (R01 AG009775; R01 AG033285), and was supported by a Consolidator Grant (IMPROVE) from the European Research Council (ERC; grant no. 726361).

Author Contributions

Conceptualization: OA, SW, MB

Writing (original draft): OA, LC, WC, RD, PDK, AH, JK, EK, JO, SR, KDV, AV, MB

Writing (reviewing and editing): OA, SW, LC, WC, RD, PDK, AH, JK, EK, JO, SR, KDV, AV, MB

Project administration: OA

Open Science Practices

This article is a tutorial that did not include any data or conducted analyses. It was not pre-registered. Therefore, it didn't receive any badges. The entire editorial process, including the open reviews, is published in the online supplement.

References

- American Psychological Association. (2019). Publication Practices & Responsible Authorship. Retrieved from <https://www.apa.org/research/responsible/publication>
- Arslan, R. C. (2019). How to Automatically Document Data With the codebook Package to Facilitate Data Reuse. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245919838783>
- Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the type I error rate in independent samples t tests: The power of alternatives and recommendations. *Psychological Methods*, 19(3), 409. <https://doi.org/10.1037/met0000014>
- Bowman, S. D., DeHaven, A. C., Errington, T. M., Hardwicke, T. E., Mellor, D. T., Nosek, B. A., & Soderberg, C. K. (2016, January 1). OSF Prereg Template. <https://doi.org/10.31222/osf.io/epgjd>
- Brookfield, K., Parry, J., & Bolton, V. (2018). Getting the measure of prosocial behaviors: A comparison of participation and volunteering data in the national child development study and the linked social participation and identity study. *Nonprofit and Voluntary Sector Quarterly*, 47(5), 1081-1101. <https://doi.org/10.1177/0899764018786470>
- Buchanan, E. M., Crain, S. E., Cunningham, A. L., Johnson, H. R., Stash, H. E., Papadatou-Pastou, M., ... , & Aczel, B. (2019, May 20). Getting Started Creating Data Dictionaries: How to Create a Shareable Dataset. <https://doi.org/10.31219/osf.io/vd4y3>
- Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry*, 26(6), 371-375. <https://doi.org/10.11919/j.issn.1002-0829.214171>
- Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2019, May 9). Preregistration: Comparing Dream to Reality. <https://doi.org/10.31234/osf.io/d8wex>
- Cristea, I. A., & Ioannidis, J. P. (2018). P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PLoS one*, 13(5), e0197440. <https://doi.org/10.1371/journal.pone.0197440>
- Devezer, B., Navarro, D. J., Vandekerckhove, J., & Buzbas, E. O. (2020). The case for formal methodology in scientific reform. *BioRxiv*, 2020.04.26.048306. <https://doi.org/10.1101/2020.04.26.048306>
- Dutilh, G., Sarafoglou, A., & Wagenmakers, E. J. (2019). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese*, 1-28. <https://doi.org/10.1007/s11229-019-02456-7>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Friedrichs, R. W. (1960). Alter versus ego: An exploratory assessment of altruism. *American Sociological Review*, 496-508. <http://doi.org/10.2307/2092934>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156-168. <https://doi.org/10.1177/2515245919847202>
- Gradstein, M. (1986). Maximal correlation between normal and dichotomous variables. *Journal of Educational Statistics*, 11(4), 259-261.
- Grady, D. G., Cummings, S. R., & Hulley, S. B. (2013). Research using existing data. *Designing clinical research*, 192-204. Retrieved from <https://pdfs.semanticscholar.org/343e/04f26f768c9530f58e1847aff6a4e072d0be.pdf>
- Grønbjerg, K. A., & Never, B. (2004). The role of religious networks and other factors in types of volunteer work. *Nonprofit Management and Leadership*, 14(3), 263-289. <https://doi.org/10.1002/nml.34>
- Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin longitudinal study (WLS). *International Journal of Epidemiology*, 43(1), 34-41. <https://doi.org/10.1093/ije/dys194>
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, & B. Puranen et al. (eds.). (2014). *World Values Survey: Round Six - Country-Pooled Datafile Version*: www.worldvaluessurvey.org/WVSDocumentationWV6.jsp. JD Systems Institute.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217.

- https://doi.org/10.1207/s15327957pspr0203_4
- Koenig, L. B., McGue, M., Krueger, R. F., & Bouchar, T. J. Jr. (2007). Religiousness, antisocial behavior, and altruism: Genetic and environmental mediation. *Journal of Personality*, 75(2), 265-290. <https://doi.org/10.1111/j.1467-6494.2007.00439.x>
- Konrath, S., Fuhrel-Forbis, A., Lou, A., & Brown, S. (2012). Motives for volunteering are associated with mortality risk in older adults. *Health Psychology*, 31(1), 87-96. <http://doi.org/10.1037/a0025226>
- Lang K. M. & Little T. D. (2018). Principled Missing Data Treatments. *Prevention Science*, 19(3), 284-294. <https://doi.org/10.1007/s11121-016-0644-5>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269. <https://doi.org/10.1177/2515245918770963>
- Lazerwitz, B. (1962). Membership in voluntary associations and frequency of church attendance. *Journal for the Scientific Study of Religion*, 2(1), 74-84. <http://doi.org/10.2307/1384095>
- Leyes, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, 32(1). <http://doi.org/10.5334/irsp.289>
- Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., ..., & Merz, C. J. (2019). How to not get lost in the garden of forking paths: Lessons learned from human fear conditioning research regarding exclusion criteria. <https://doi.org/10.31234/osf.io/6m72g>
- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1), 1-19. <http://doi.org/10.18637/jss.v009.i08>
- Massey F. J. Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253), 68-78.
- Meade, A. W. (2010). Restriction of Range. In N. J. Sand (Ed.), *Encyclopedia of Research Design*. SAGE Publishing. Retrieved from <https://sk.sagepub.com/reference/research-design/n388.xml>
- Merriam-Webster (n.d.). Bias. In Merriam-Webster.com dictionary. Retrieved January 26, 2021, from <https://www.merriam-webster.com/dictionary/bias>.
- Mertens, G., & Kryptos, A.M. (2019). Preregistration of Analyses of Preexisting Data. *Psychologica Belgica*, 59(1), 338-352. <http://doi.org/10.5334/pb.493>
- Morgan, S. P. (1983). A research note on religion and morality: Are religious people nice people? *Social Forces*, 61(3), 683-692. <http://doi.org/10.2307/2578129>
- Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, 322(5898), 58-62. <http://doi.org/10.1126/science.1158757>
- Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van 't Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, 23(10), 815-818. <https://doi.org/10.1016/j.tics.2019.07.009>
- Olsson-Collentine, A., Van Assen, M. A., & Hartgerink, C. H. (2019). The prevalence of marginally significant results in psychology over time. *Psychological Science*, 30(4), 576-586. <https://doi.org/10.1177/0956797619830326>
- Parliament of the World's Religions. (1993). Toward a global ethic: An initial declaration. Retrieved from https://www.weltethos.org/1-pdf/10-stiftung/declaration/declaration_english.pdf
- Pew Research Center. (2015). America's changing religious landscape. Pew Research Center. Retrieved from <https://www.pewforum.org/2015/05/12/americas-changing-religious-landscape>
- Pharoah, C., & Tanner, S. (1997). Trends in charitable giving. *Fiscal Studies*, 18(4), 427-443. <https://doi.org/10.1111/j.1475-5890.1997.tb00272.x>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27-42. <https://doi.org/10.1177/2515245917745629>
- Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, 21(4), 308-320. <https://doi.org/10.1037/gpr0000128>
- Saroglou, V. (2002). Religion and the five factors of personality: A meta-analytic review. *Personality and Individual Differences*, 32(1), 15-25. [https://doi.org/10.1016/S0191-8869\(00\)00233-6](https://doi.org/10.1016/S0191-8869(00)00233-6)

- Silver, N. C. (2008). Attenuation. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*. SAGE Publishing. Retrieved from <http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n24.xml>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, 144(6), 1146-1152. <https://doi.org/10.1037/xge0000104>
- Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., & Steinman, M. A. (2011). Conducting high-value secondary dataset analysis: An introductory guide and resources. *Journal of General Internal Medicine*, 26(8), 920-929. <https://doi.org/10.1007/s11606-010-1621-5>
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712. <https://doi.org/10.1177/1745691616658637>
- Thompson, W. H., Wright, J., Bissett, P. G., & Poldrack, R. A. (2019). Dataset Decay: the problem of sequential analyses on open datasets. *bioRxiv*, 801696. <https://doi.org/10.1101/801696>
- Veldkamp, C. L. S., Bakker, M., van Assen, M. A. L. M., Cromptoets, E. A. V., Ong, H. H., Nosek, B. A., ..., & Wicherts, J. M. (2018). Ensuring the quality and specificity of preregistrations. <https://doi.org/10.31234/osf.io/cdgyh>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6), 632-638. <https://doi.org/10.1177/1745691612463078>
- Wason, J. M., Stecher, L., & Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: Is it necessary and is it done? *Trials*, 15(1), 364. <https://doi.org/10.1186/1745-6215-15-364>
- Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of pre-Existing Datasets. *Advanced Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245919848684>
- Williams, M. N., & Albers, C. (2019). Dealing with distributional assumptions in preregistered research. *Meta-Psychology*, 3. <https://doi.org/10.15626/MP.2018.1592>
- Youniss, J., McLellan, J. A., & Yates, M. (1999). Religion, community service, and identity in American youth. *Journal of Adolescence*, 22(2), 243-253. <https://doi.org/10.1006/jado.1999.0214>