Open access • Journal Article • DOI:10.1037/ABN0000424

A step-by-step guide on preregistration and effective data sharing for psychopathology research. — Source link

Angelos-Miltiadis Krypotos, Irene Klugkist, Gaëtan Mertens, Iris M. Engelhard

Institutions: Utrecht University

Published on: 01 Aug 2019 - Journal of Abnormal Psychology (American Psychological Association)

Topics: Workflow, Data anonymization and Data sharing

A Step-By-Step Guide on Preregistration and Effective Data Sharing for Psychopathology Research

Angelos-Miltiadis Krypotos[1,2], Irene Klugkist[3], Gaëtan Mertens[1], & Iris M. Engelhard[1]

[1] Department of Clinical Psychology, Utrecht University, Utrecht, The Netherlands
[2] Department of Health Psychology, KU Leuven, Leuven, Belgium
[3] Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

Article to appear at the *Journal of Abnormal Psychology. Please note that there may be deviations between this and the final article. Please refer to the original article for citation/reference.*

Author Note

GM and IME are supported by a VICI grant awarded by the Netherlands Organization for Scientific Research (NWO; 453-15-005) to IME. The authors thank Nicolas Perez for programming, together with AMK, the software presented in this paper.

Correspondence concerning this article should be addressed to Angelos-Miltiadis Krypotos, Utrecht University, Department of Clinical Psychology, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands. E-mail: amkrypotos@gmail.com.

Abstract

Data analysis in psychopathology research typically entails multiple stages of data preprocessing (e.g., coding of physiological measures), statistical decisions (e.g., inclusion of covariates), and reporting (e.g., selecting which variables best answer the research questions). The complexity and lack of transparency of these procedures have resulted in two troubling trends: the central hypotheses and analytical approaches are often selected after observing the data, and the research data are often not properly indexed. These practices are particularly problematic for (experimental) psychopathology research because the data are often hard to gather due to the target populations (e.g., individuals with mental disorders), and because the standard methodological approaches are challenging and time consuming (e.g., longitudinal studies). Here, we present a workflow that covers study preregistration, data anonymization, and the easy sharing of data and experimental material with the rest of the research community. This workflow is tailored to both original studies and secondary statistical analyses of archival datasets. In order to facilitate the implementation of the described workflow, we have developed a free and open-source software program. We argue that this workflow will result in more transparent and easily shareable psychopathology research, eventually increasing and replicability reproducibility in our research field.

Keywords: replicability, reproducibility, experimental psychopathology, R

A Step-By-Step Guide on Preregistration and Effective Data Sharing for Psychopathology Research

The main goals of psychopathology research are to unveil the factors that contribute to the genesis and maintenance of mental disorders, and to develop relevant prevention and intervention programs (Marks & Yardley, 2004; van den Hout, Engelhard, & McNally, 2017). This research area often requires challenging data accumulation methods (Comer & Kendall, 2013), including longitudinal research in samples at risk of developing mental disorders, and demanding research protocols (e.g., randomized control trials, RCT). Given these challenges, it is crucial to make the most of the collected data.

The timely answering of research questions depends on how reliable the published literature is. Recent findings in psychology, however, suggest that many popular effects cannot be reproduced (e.g., Open Science Collaboration, 2015; Świątkowski & Dompnier, 2017). There are scientific, ethical, and practical reasons that make such low reproducibility deleterious for psychopathology research. Scientifically, a finding with low reproducibility is not informative, and it slows the progress of our field. Ethically, unreliable research findings stall the development of effective interventions for mental disorders. Practically, unreproducible psychopathology research is a waste of resources and patients' time (Baker, McFall, & Shoham, 2008). Arguably, psychopathology research can only progress by studies that are *replicable* (i.e., repetition of the results using similar procedures but a new data set) and *reproducible* (i.e., obtaining the same results as the original study by using the same procedures and data) (Brandt et al., 2014; Goodman, Fanelli, & Ioannidis, 2016).

Replicability in psychology is often hampered by the formation of a study's hypotheses after the results are already known (Kerr, 1998; Nosek, Spies, & Motyl, 2012). Hypothesizing based on the results could inflate the rate of false-positives and lead to non-informative conclusions. A proposed way to demonstrate that a research hypothesis has been formed *prior* to the beginning of a study, as well as to avoid the temptation of post-hoc decisions, is *preregistration* (Chambers, 2017; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). In the case of an original study, preregistration refers to the a-priori documentation of the research questions, hypotheses, methods, and statistical analyses (although changes in the document can still be made later on; see below). In the case of *secondary analyses* (i.e., follow-up statistical analyses of an archival data set), preregistration refers mainly to the documentation of the research hypotheses and statistical analyses. Preregistration is routinely used in RCTs (e.g., *clinicaltrials.gov* in the United States and *eudract.ema.europa.eu* in Europe) but not in other types of psychopathology research. Because psychopathology research provides the foundation for the follow-up development of clinical interventions (van den Hout et al., 2017), preregistration may further increase the reproducibility and replicability of these studies.

The reproducibility of a study can be demonstrated by making the full data set available, together with the relevant analyses' scripts. Importantly, the availability of the data, analyses' scrips, and the accompanying material also enable the easier and more accurate replication studies by independent research labs. As such, the availability of the data and material, when held to current ethical standards, can help the further advancement of our field.

Preregistration of scientific studies and the sharing of data/material are not new ideas (Chambers, 2017; Klein et al., 2018). Still, they have yet to be widely implemented in psychopathology research. We see at least three factors that hamper the implementation of open science practices in psychopathology research. First, although there is a growing awareness about the need to preregister a study (e.g., van't Veer & Giner-Sorolla, 2016), there is no consensus about the type of information that should be included in a preregistration document. Similar concerns apply to the open sharing of research data. Second, there have been major developments in how research is done, with a plethora of software options and websites now being incorporated in researchers' workflows. However, to effectively employ these new tools often requires extensive time and effort investments from students and researchers. Third, there are concerns regarding the open sharing of sensitive information often obtained from participants in psychopathology research.

To address these problems, we have formulated *six* steps for the effective preregistration of studies and data sharing in psychopathology research (see Appendix). This workflow is tailored towards original studies as well as secondary statistical analyses of archival datasets.

In order to facilitate the easy follow-up of the proposed workflow, we have developed *The Preregistration and Sharing Software* (*pss*; Figure 1), which can be downloaded for free at https://github.com/AngelosPsy/pssr. Our software provides a suite of functions for a project's preregistration and data sharing, and the logging of changes made in any of the files (Figure 2). Specifically, *pss* uses the popular version control system, '*git',* to keep track of all changes made in the files. For example, when a new line of code is added to the scripts, the software points to which files were changed and which changes were applied, without creating new copies of the files. Version control systems (Bryan, 2018; Vuorre & Curley, 2018) have multiple advantages. They allow researchers to work on the same files throughout the project, rather than having to create new files when new versions of the manuscript, analyses, or data are created. The smaller number of files can be easily organized into a single digital folder that can be shared with the scientific community. Finally, researchers can easily track down which changes and decisions were made by whom at each time point. This last point is especially useful for longitudinal projects and for justifying the contribution of each collaborator in a project. A detailed tutorial of *pss* can be found at https://github.com/AngelosPsy/pssr_tutorial.

Steps for preregistering a study and sharing the research data

STEP 1: Determination of research questions and predictions

Traditionally, the research questions and hypotheses of a study are communicated with the rest of the community through the presentation of the results in a research article or a conference. Today, best practices mandate that both the research questions and hypotheses are known before beginning data collection/analyses in the form of a preregistration document (Chambers, 2017; Kerr, 1998). Without preregistration, researchers could selectively report outcomes that support their hypotheses, leading to *researcher degrees of freedom* (or *the garden of forking paths*). Unspecified predictions and researcher degrees of freedom have resulted in considerable skepticism on a range of findings (Gelman & Loken, 2013). A preregistration document provides many advantages to the researchers, including that they can now take full credit for their predictions and that they are protected from criticism about whether the study was performed as planned (Wagenmakers & Dutilh, 2016).

The preregistration document typically begins with the research questions, followed by the hypotheses. Hypotheses can be *confirmatory* or *exploratory* (De Groot, 2014). This distinction is helpful, for instance, in clarifying which hypotheses are designed to confirm a specific prediction *before* seeing the data and which are formulated to explore potential data patterns *after* seeing (some of) the data. Specifically, confirmatory hypotheses are used for studies that are designed to rigorously test a theoretical prediction in a highly constrained context with strict limits on researcher degrees of freedom. These hypotheses should be formulated prior to data analysis and should describe the predicted data pattern as detailed as possible. To illustrate, the hypothesis "anxiety scores in individuals with anxiety disorders will be lower after cognitive behavior therapy than after the control intervention" is vague because "anxiety scores" can be defined in various ways. Including the definition of "anxiety scores" (e.g., trait anxiety as measured by the trait subscale of the State-Trait Anxiety Inventory; Spielberger, Gorsuch, & Lushene, 1970) will result in a stronger hypothesis. The size of the effect (e.g., a δ of .50) could also be specified, although having such specific hypotheses is uncommon in psychology (Berger & Delampady, 1987; Meehl, 1954).

Exploratory hypotheses, on the other hand, can be formed at any time during a study and they may not include specific predictions about the data pattern. To return to the previous example, a possible exploratory hypothesis could be that the between-group differences are moderated by baseline severity of the anxiety disorder (e.g., Wolitzky-Taylor, Arch, Rosenfield, & Craske, 2012). Secondary analyses of archival data are typical examples of exploratory research (Nosek, Ebersole, DeHaven, & Mellor, 2018). Follow-up confirmatory studies can be done to test whether results of exploratory research are trustworthy (De Groot, 2014).

Due to publication pressure (Simmons, Nelson, & Simonsohn, 2011), it could be tempting to present an exploratory study as confirmatory (Nosek et al., 2012). Although this approach may help in the publication of a study, it is deleterious to the field as it may give rise to false-positive results. Preregistration of hypotheses can help in distinguishing between exploratory and confirmatory hypotheses. Nonetheless, while confirmatory hypotheses can result in strong conclusions, exploratory research remains important because it may inspire future confirmatory studies, and it is often performed with data that are difficult to collect (e.g., functional magnetic resonance imaging scans of individuals with a low-prevalence mental disorder).

To assist with this step, we have created a preregistration template that is available within *pss* (see Figure 3 and the Appendix). Our software also supports two other commonly used templates for preregistration: the OSF (from the osf.io website) and the *aspredicted* (from the aspredicted.org website).

STEP 2: Methods and statistical plan

The amount of methodological details included in the preregistration document depends on whether it refers to an original study or to secondary analyses. In the former case, it is advisable that a clear description of the used material (including stimuli and questionnaires), procedures, equipment, and protocol is included in the pre-registration document. This document should also include sufficient information on the used experimental paradigm, so that readers can follow each step of the method and independent labs can replicate the study. Notably, experimenters who aim to replicate a study are expected to have sufficient training and experience with the methods they employ; a detailed protocol can help in this direction. Researchers should also preregister the design of the study (between-subject or within-subject) together with information about whether the study is experimental, longitudinal, or crosssectional. It is also strongly advisable that the preregistration document includes information about the potential blinding of the experimenters, the method of data acquisition (e.g., online questionnaires), and the sampling method. Lastly, to preregister a meta-analysis or systematic review, researchers are advised to use the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009).

In the case of secondary analyses, it is sufficient to refer to the initial study (e.g., by providing the *digital object identifier* of the original published record), and the way the data were acquired (e.g., by providing a weblink). The preregistration document should also include specifics on the planned statistical analyses.

(Dis-)confirmation of research predictions. The results of a study could lead to the (dis-)confirmation of the predictions or the conclusion that there is insufficient evidence for arguing for or against the research predictions. Because different statistical approaches can yield different results (Shafer, 1982; Silberzahn et al., 2018), the preregistration form should specify which statistical tests will be performed. Below, we extend how this can be done in case of Null-Hypothesis Significance Testing (NHST), Bayesian Hypothesis Testing (BHT), model selection, and correlational analyses.

NHST is useful for testing the *existence* of differences between groups/conditions. Importantly, the null hypothesis is either rejected or not rejected. However, within a frequentist context, there are procedures for finding support for the null hypothesis (Lakens, 2017). One of these is *equivalence testing* (Lakens, 2017; Wellek, 2010) in which the null hypothesis is defined as the existence of an interesting effect, with that effect falling within the *equivalence bound.* For example, someone could define a Cohen's *d* between -0.2 and 0.2 as the absence of the effect. The null hypothesis then contains two expressions: the effect is smaller than -0.2 or the effect is larger than 0.2.

The alternative hypothesis is that the effect is less extreme as the defined equivalence bound. After defining the two hypotheses, a statistical approach is used (e.g., two one-sided t-tests or a 95% confidence interval) in order to reject the hypothesis that the observed effect is large enough to be judged as worthwhile (defined by the equivalence bound; Lakens, 2017). In our example, if the upper and lower bound of the confidence interval are .12 and .17, respectively, then the null hypothesis is rejected (the interval is not entirely larger than 0.2), and we conclude that the observed effect is not relevantly different from zero (as expressed in the alternative hypothesis).

Within the NHST framework, there is a wide debate as to whether an α level of 0.05 or lower (e.g., 0.005) should be used, or whether researchers should be allowed to determine their α based on their research question (Benjamin et al., 2018; Lakens et al., 2018). Given the different opinions, we advise that the general α level and the α level after correcting for multiple comparisons are specified in the preregistration of a study.

Another approach that allows both the confirmation and dis-confirmation of a research hypothesis is to follow a Bayesian procedure, such as Bayesian Hypothesis Testing (BHT; e.g., Dienes, 2014; Kruschke, 2011; Krypotos, Klugkist, & Engelhard, 2017). In BHT, relative evidence for the alternative and null hypotheses is accumulated from the collected data, which makes it possible to compare the alternative to the null hypothesis and vice versa (Kruschke, 2011). One point of caution when using Bayesian statistics is the selection of meaningful *prior probability distributions* (see Krypotos, Blanken, Arnaudova, Matzke, & Beckers, 2017; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010 for tutorials). The prior probability distributions represent the beliefs that a person has about the parameters in a study (e.g., β_0 and β_1 parameters of a regression model) before observing the data. The careful selection of prior distributions is particularly important because the results of BHT will change when different priors are used. It is against Bayesian inference to choose priors based on the direction of the observed results (Dienes, 2016). As such, it is recommended that the definition of prior distributions is included in the preregistration document, together with the level of evidence for (dis-)confirming each hypothesis (Jeffreys, 1961; Wetzels et al., 2011).

Another method to draw statistical conclusions from a study is to define statistical models and compare them using diverse model selection criteria (e.g., Akaike information criterion, Bayesian information criterion). In these cases, researchers are advised to describe their choices regarding the parameters of the models, the criteria for model selection, and the threshold for deciding which model is preferred. In the case of Bayesian modeling, it is strongly advised that the prior distribution of each parameter is mentioned in the preregistration document. Any ambiguity in terms of model definition/selection could increase the degrees of freedom of the researchers and create skepticism as to whether the presented results were influenced by potential biases during data analyses.

Many studies include the computation of correlation coefficients between variables. For example, someone could correlate personality characteristics or investigate the relation between personality characteristics and a performance variable (e.g., fear learning; Gazendam, Kamphuis, & Kindt, 2013). In such cases, it is advised that apart from the *p*-values, α level, and/or Bayes factors, the predicted size of the correlation coefficient is included in the preregistration document. When more complicated models are used (e.g., mediation or moderation models), then the guidelines for reporting each model could be followed (see previous paragraph). Lastly, in the case of confirmatory factor analyses, the researchers should define each predicted factor, the items that load to each factor, as well as the chosen rotations, in the pre-registration document (Thompson, 2004). Adding these specifications will ensure that the predetermined statistical plan is clear and robust.

Sample size determination. Within NHST, a power calculation is performed *prior to data collection*. Statistical power is the probability that a test correctly rejects the null hypothesis. With values ranging between 0 to 1, the recommended power of a test is typically 0.80 (Cohen, 1988, 1992). Apart from the specified power and α levels, power calculations also depend on the expected effect size. Given that an estimated effect size is subject to variability, researchers are urged to also consider the accuracy (i.e., the width of the confidence intervals) of the predicted effect size when planning a study (Maxwell, Kelley, & Rausch, 2008). The main way (but not the only way; see McClelland, 2000) to achieve high precision around an effect size is by using large samples (Maxwell et al., 2008). However, in psychopathology research, participant recruitment is often challenging, making it difficult to recruit the sample suggested by a power analysis. To illustrate, in one study we screened 480 participants in order to find 68 participants who fitted our selection criteria (Toffolo, van den Hout, Hooge, Engelhard, & Cath, 2013). When recruiting a large enough sample size is not possible, this could be acknowledged in the preregistration document and the final report of the study.

An alternative approach for determining the sample size is to stop the data collection when the evidence crosses a threshold. This assumes that the results are checked multiple times during data collection and not just at the end. However, within NHST checking the results during data collection increases the chance of false-positives. Specifically, *p*-values are bound to cross a predefined alpha level with enough participants, *even when* the tested effect comes from the null hypothesis (Wagenmakers, 2007). As such, during data collection, a researcher could check the results and continue collecting data until a *p*-value becomes small enough. To safeguard against this, we outline two principled methods to check the results multiple times during data collection.

The first one is to use interim analyses. Interim analysis allows a researcher to compute *p*-values, as typically done in NHST, at multiple points during data collection, while controlling for false-positives by, for example, using lower α levels for every time the statistical analyses are conducted (Armitage, McPherson, & Rowe, 1969; Dodge & Romig, 1929; see Lakens, 2014 for an example). An alternative way to evaluate the results before the end of data collection is by using BHT (Bernardo & Rueda, 2002; Wagenmakers et al., 2010). With BHT, the data results can be inspected after each participant has been tested (Rouder, 2014; Schönbrodt & Wagenmakers, 2018). Researchers could consider collecting data until a threshold of evidence is met, rather than after testing a predetermined number of participants. Notably, this approach obviates the argument for listing a fixed sample size in the preregistration; this is particularly useful whenever the research involves difficult-to-recruit samples. Regardless of whether interim analysis or BHT is used, researchers are encouraged to mention the stopping rules/thresholds in their preregistration document.

When secondary analyses are performed, it is useful to determine the size of the predicted effect that can be achieved with the recruited sample size. This could help in the interpretation of the results as maybe, and whenever using NHST, no significant results arose due to the sample being insufficient for detecting the predicted effect (e.g., an effect size of Cohen's *f* of .25 is predicted, with a power of 80%, when the sample size is large enough for detecting a Cohen's *f* of .40). When the parameters of a statistical model are estimated, it can also be useful to argue why the recruited sample and the available trial size per individual can lead to reliable parameter estimation (i.e., parameter values with reasonably small confidence intervals). Such estimation can be achieved by data simulation (see *Generating analysis scripts* section).

(In)dependent variables and data manipulation. A preregistration document should define all dependent and independent variables. As mentioned above, flexibility in the variables that are included in the analyses could result in different results. As such, and especially in the case of confirmatory research, the dependent and independent variables should be explicitly stated, together with the statistical analyses that will include these variables. For example, it would be insufficient to mention that “we will use different Analyses of Variance (ANOVA) for all the main variables of a study”, as it is neither clear what type of ANOVA will be used (e.g., one-way ANOVA, repeated measures ANOVA) nor what the independent and dependent variables are. When variables are used for exploration, specification of the independent and dependent variables could be included in an exploratory hypotheses section.

Collected data are often manipulated before being analyzed. These data manipulations include the exclusion of outlying values, data transformations (e.g., log-values), and computation of summary statistics (e.g., means). Notably, different data manipulation procedures (e.g., outlier corrections) can lead to different outcomes. By mentioning the exact data cleaning procedure, the preregistration document will alleviate confusion and post-hoc decisions regarding which data cleaning approach was followed and why. As stated earlier, if data manipulations other than those mentioned in the preregistration document are deemed more appropriate after seeing the data, they can still be used as long as this is explicitly acknowledged (e.g., defined as exploratory).

Generating analysis scripts. A useful exercise after determining the statistical analyses is to simulate data according to the study’s predictions. This helps in specifying the predicted data pattern (e.g., interactions between variables). While generating this script, the researcher(s), could also explore extreme values in the data and decide how such cases will be handled when the real data are available.

A result of this exercise is the generation of an *analysis script* that can be used for reading and analyzing the experimental data at the end of the study. The analysis script provides a record of how to conduct the statistical analyses, which is useful for detecting potential errors, and saves time whenever similar scripts are used between studies. When authors choose to simulate data, they can also see how each data correction decision (e.g., removal of outliers) influences the final outcome. Lastly, the analyses of the confirmatory hypotheses are hard-coded, which prevents the analyses being determined based on the results of the study.

The generation of an analysis script requires knowledge of a scripting language (e.g., R, Python). Relying on a scripting language, rather than using mouse-click programs, is extremely useful for reproducible research (e.g., Gandrud, 2016), and we encourage researchers to take advantage of such scripting languages.

STEP 3: Run a pilot study/analysis

Before preregistering a study, it is advised to run a pilot study. There are good reasons for this: a pilot allows for testing many aspects of the main study including the recruitment rates, randomization, procedures, and the general feasibility of the project (Leon, Davis, & Kraemer, 2011; Thabane et al., 2010). Pilot studies are often used to determine the effect size, which is used for calculating the required sample size of the main study. This approach, however, has been heavily criticized given that small sample studies, which are commonly used in pilot studies, often do not give an accurate estimate of the effect size (Leon et al., 2011).

Before the main analyses of a secondary data set are performed, *pilot analyses* can be performed using parts of the data set (e.g., 10% of the data). In case no data are available, then test analyses could be performed on simulated data as described in Step 2, ‘Methods and statistical plan’. Potential revisions of the preregistration document could follow the pilot results. However, although pilot studies are common in experimental psychopathology, this is not the case in other types of studies (e.g., in RCTs). This is why our preregistration template does not require that a pilot study is run.

STEP 4: Material gathering

Together with the preregistration of the study, it is advisable that all study material (e.g., the files used for running a program collecting reaction time data) are gathered and shared online (see also STEP 1). In order to assure that future users of these material acknowledge your work, it is useful to obtain a *copyright license*. Copyright licenses describe the conditions that should be met so that the licensor grants permission for the use of the data and material by a third party. Three conditions are usually covered: 1) the attribution requirement (anyone who uses the data/material should give credit to the licensor), 2) the copyleft requirement (new work derived from using the licensed data/material should be released using the original license), and 3) non-commerciality (commercial use of the licensed data/material is not permitted). These conditions are described in prepared licenses (e.g., the Creative Commons Attribution 4.0 International Public License; see also choosealicense.com/licenses/ and http://www.dcc.ac.uk/resources/how-guides). Researchers may want to create their own bespoke license, but this requires a good understanding of the relevant laws. Researchers can easily license their data and material online at Open Science Framework (osf.io) or Figshare (figshare.org). After creating an account, they can attach licenses to shared data and material. Research that uses archival data should link to the (potentially) existing data, rather than archiving the material again. Note also that researchers should not publicly share copyright protected material.

STEP 5: Study preregistration

Currently, there are two ways to preregister a study. The first is to publish it in an online repository. Meyer (2018) provides an extensive list of the online available repositories, including the *databrary*, nyu.databrary.org, the *Harvard Dataverse* (dataverse.harvard.edu), and *Zenodo* (zenodo.org). Here, we focus on the Open Science Framework (OSF; osf.io) and *aspredicted* (aspredicted.org) websites. OSF provides support for almost any data type, allows the preregistration of studies by time-stamping when the preregistration was created, and enables users to license and assign a Digital Object Identifier (DOI) to the uploaded material. Notably, in OSF, the preregistration document is embargoed from public view for up to 4 years. As a result, the document could be made publicly available prior to the conclusion of a study (e.g., in case of longitudinal studies). Nevertheless, the limited embargo prevents the preregistration of multiple documents with potentially different hypotheses for a single study.

In *aspredicted* (aspredicted.org) researchers only need to answer nine questions about the study. Although the template of *aspredicted* encourages short preregistration documentation, this often results in less detailed formulation of research hypotheses and/or statistical analyses. Furthermore, the *aspredicted* website has adopted a “private forever” choice as a way to make the preregistration more appealing to researchers who may disagree with the eventual publication of their preregistration document. Allowing researchers to keep their preregistrations private makes *aspredicted* an informal registry

(Nosek et al., 2018). Another disadvantage of *aspredicted*, compared to OSF, is that it does not allow the uploading of other files than the preregistration document, such as material or data files. Our software provides links to the different preregistration websites.

A second way to preregister a study is by using registered reports (Nosek & Lakens, 2014). This new type of article allows the review of the introduction and methods *prior* to the data collection. If the registered report is accepted, the researchers need to run the proposed study in accordance with the accepted protocol. There are two key advantages of this publishing format. First, the acceptance of an article is mainly based on the importance of the research question and the appropriateness of the methodology, rather than the direction of the results (Easterbrook, Gopalan, Berlin, & Matthews, 1991; Sterling, 1959). Second, reviewers and editors could point to potential problems with the main design prior to the start of data collection. A curated list of journals that accept registered reports can be found at cos.io/rr. Currently, more than 150 journals offer this publishing format and this number is rapidly growing. Each journal has specific preregistration criteria; however, the steps suggested in this paper will cover most, if not all, of the requirements of the journals that offer registered reports.

STEP 6: Upload the data and the results report

In the case of original studies, the data can be uploaded online with the rest of the material. For secondary analyses, the researcher should ask for permission from the original researchers to link to the available data. Given the sensitivity of the data collected in psychopathology research (e.g., reports of past medication), there are ethical and legal constraints (Bonini, Eichler, Wathion, & Rasi, 2014; US Department of Health and Human Services, 2014) as to what can be shared (e.g., Gilmore, Kennedy, & Adolph, 2018; Joel, Eastwick, & Finkel, 2018; Klein et al., 2018; Knoppers, Harris, Budin-Ljøsne, & Dove, 2014; Meyer, 2018; Walsh et al., 2018), and general concerns about whether data sharing is beneficial or not (Houtkoop et al., 2018). Our software provides a suite of functions for anonymizing the available data (see Figure 4 and the tutorial on https://github.com/AngelosPsy/pssr_tutorial).

A report with the code used for each analysis, together with the corresponding results, is helpful in capturing the exact steps taken during the data analysis. Rmarkdown for R (Allaire et al., 2016) and Python Notebook (Kluyver et al., 2016) allow the users to see the code for running the analysis and the accompanying output. This goes beyond other click-based software where different files are created for the analyses and the results.

Nowadays it is also common that researchers release a *pre-print* of their publications online. Pre-prints are advanced versions of the manuscript that may be largely identical to the published paper. The American Psychological Association has designated PsyArXiv (psyarxiv.com) as the preferred service for publishing pre-prints. The advantage of publishing a pre-print is that authors may receive comments on their work before submitting their manuscript to a journal and can benefit this way from an extra round of reviews. However, not all publishers allow the online publication of preprints submitted to their journals. To check which publishers support pre-prints, authors can consult the Sherpa (sherpa.ac.uk) website.

Discussion

We have presented six steps towards the preregistration of psychopathology studies and the public sharing of the data and material. The preregistration of original and/or secondary studies allows researchers to take credit for their predictions and remove potential criticism of post-hoc hypothesizing. Given the recent confidence crisis in psychology research, there is an urgent need to enhance the replicability and reproducibility of research. As we have argued above, the preregistration of a study can help greatly in this direction. The open sharing of data and research material could also accelerate follow-up research, as the available material can help in the exact replication of a study and inspire follow-up studies. All the above could potentially lead to the quicker answering of the main research questions about etiology, maintenance, and treatment of psychopathology.

To further assist with following the steps above, we have created a free and easy-to-use software that requires no prior knowledge of a programming language. It runs on the researcher's local server and it automatically time-indexes all study material whenever they are accessed. This is an advantage over web-based projects that allow the time-stamping of a study's documents only when the user is online.

Common critiques of preregistration include that a) it limits creativity in coming up with research questions/hypotheses, b) is difficult when large-scale studies are conducted, and that c) could allow other research labs to "steal" the study idea. We think that none of these arguments are actual threats to a study. Creativity and exploration are allowed in a preregistered study as long as the relevant hypotheses are defined as exploratory. In case of large-scale studies, researchers eventually need to spend time thinking about their research questions and data analyses approach. We encourage that this is done before seeing the data as this will provide more unbiased hypotheses. Lastly, potential "scooping" can be prevented by keeping a preregistration document private until the study is completed. All in all, we believe that there is little reason not to preregister a study given all the mentioned advantages.

Preregistration requires effort. Incentives for following these steps could be established in journals and grant committees. Recently, some journals have used badges for acknowledging open practices (Kidwell et al., 2016). Similar badges could be adopted for following all the steps described above. Alternatively, authors could simply add a sentence in their manuscript acknowledging that they have followed the suggested steps. For sponsored research, it could be useful to add a new section to grant applications on whether the applicant(s) will follow standard open practices.

To conclude, the preregistration of studies and sharing of data/material provides vast benefits to researchers and the community. By following the suggested steps, psychopathology research will be able to provide faster and more correct answers to the key questions of the field. This is important for patient care as well as for benefiting society by reducing the related economic costs.

References

Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., . . . Arslan, R. (2016). Rmarkdown: Dynamic documents for R. *R Package Version*, *1*, 9010.

Armitage, P., McPherson, C., & Rowe, B. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society. Series A (General)*, 235–244. doi:10.2307/2343787

Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, *9*, 67–103. doi:10.1111/j.1539-6053.2009.01036.x

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . . others. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6. doi:10.1038/s41562-017-0189-z

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 317–335.

Bernardo, J. M., & Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review*, *70*, 351–372. doi:10.1111/j.1751-5823.2002.tb00175.x

Bonini, S., Eichler, H.-G., Wathion, N., & Rasi, G. (2014). Transparency and the european medicines agency?sharing of clinical trial data. *New England Journal of Medicine*, *371*, 2452–2455. doi:10.1056/NEJMp1409464

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224. doi:10.1016/j.jesp.2013.10.005

Bryan, J. (2018). Excuse me, do you have a moment to talk about version control? *The American Statistician*, *72*, 20–27. doi:10.1080/00031305.2017.1399928

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd. Hillsdale, NJ: erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033- 2909.112.1.155

Comer, J. S., & Kendall, P. C. (2013). *The Oxford handbook of research strategies for clinical psychology*. Oxford University Press.

De Groot, A. (2014). The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta Psychologica*, *148*, 188–194. doi:10.1016/j.actpsy.2014.02.001

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. doi:10.1016/j.jmp.2015.10.003

Dodge, H. F., & Romig, H. (1929). A method of sampling inspection. *Bell Labs*

Technical Journal, *8*, 613–631. doi:10.1002/j.1538-7305.1929.tb01240.x

Easterbrook, P. J., Gopalan, R., Berlin, J., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*, 867–872. doi:10.1016/0140-6736(91)90201-Y

Gandrud, C. (2016). *Reproducible research with R and R studio*. Chapman; Hall/CRC.

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biological Psychology*, *92*, 342–352.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Retrieved from Http://Www.stat.columbia.edu/ Gelman/Research/ Unpublished/P_hacking.pdf*.

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*, *1*, 121–130.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*, 1–6. doi:10.1126/scitranslmed.aaf5027

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*, 70–85.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Joel, S., Eastwick, P. W., & Finkel, E. J. (2018). Open sharing of data on close relationships and other sensitive social psychological topics: Challenges, tools, and future directions. *Advances in Methods and Practices in Psychological Science*, *1*, 86–94. doi:10.1177/2515245917744281

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217. doi:10.1207/s15327957pspr0203_4

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., . . . others. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology*, *14*, e1002456. doi:10.1371/journal.pbio.1002456

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., . . . others. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*, 1–15. doi:10.1525/collabra.158

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., . . . others. (2016). Jupyter notebooks-a publishing format for reproducible computational workflows. In *ELPUB* (pp. 87–90).

Knoppers, B. M., Harris, J. R., Budin-Ljøsne, I., & Dove, E. S. (2014). A human rights approach to an international code of conduct for genomic and clinical data sharing. *Human Genetics*, *133*, 895–903. doi:10.1007/s00439-014-1432-6

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312. doi:10.1177/1745691611406925

Krypotos, A.-M., Blanken, T. F., Arnaudova, I., Matzke, D., & Beckers, T. (2017). A primer on Bayesian analysis for experimental psychopathologists. *Journal of Experimental Psychopathology*, *8*, 140–157. doi:10.5127/jep.057316

Krypotos, A.-M., Klugkist, I., & Engelhard, I. M. (2017). Bayesian hypothesis testing for human threat conditioning research: An introduction and the condir R package. *European Journal of Psychotraumatology*, *8*. doi:10.1080/20008198.2017.1314782

Lau-Zhu, A., Holmes, E.A., & Porcheret, K. (2018). Intrusive memories of trauma in the laboratory: Methodological developments and future directions. *Current Behavioral Neuroscience Reports*, *5*, 61–71. doi: 10.1007/s40473-018-0141-1

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*, 701–710. doi:10.1002/ejsp.2023

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*, 355–362. doi:10.1177/1948550617697177

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., . . . others. (2018). Justify your alpha. *Nature Human Behaviour*, *2*, 168. doi:10.1038/s41562-018-0311-x

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*, 626–629. doi:10.1016/j.jpsychires.2010.10.008

Marks, D. F., & Yardley, L. (2004). *Research methods for clinical and health psychology*. Sage.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*. doi:10.1146/annurev.psych.59.103006.093735

McClelland, G. H. (2000). Increasing statistical power without increasing sample size.

American Psychologist, *55*, 963–964.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN, US: University of Minnesota Press. doi:10.1037/11281-000

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, *1*, 131–144. doi:10.1177/2515245917747656

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269.

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*. Hogrefe Publishing.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631. doi:10.1177/1745691612459058

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349* (6251), aac4716. doi:10.1126/science.aac4716

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*, 301–308. doi:10.3758/s13423-014-0595-4

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*, 128–142. doi:10.3758/s13423-017-1230-y

Shafer, G. (1982). Lindley's paradox. *Journal of the American Statistical Association*, *77*, 325–334.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*, 337-356.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance-or vice versa. *Journal of the American Statistical Association*, *54*, 30–34. doi:10.1080/01621459.1959.10501497

Świątkowski, W., & Dompnier, B. (2017). Replicability crisis in social psychology: Looking at the past to find new pathways for the future. *International Review of Social Psychology*, *30*, 111–124. doi:10.5334/irsp.66

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., . . . Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology*, *10*, 1. doi:10.1186/1471-2288-10-1

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/10694-000

Toffolo, M. B., van den Hout, M. A., Hooge, I. T., Engelhard, I. M., & Cath, D. C.

(2013). Mild uncertainty promotes checking behavior in subclinical obsessive-compulsive disorder. *Clinical Psychological Science*, *1*, 103–109. doi:10.1177/2167702612472487

US Department of Health and Human Services. (2014). Federal policy for the protection of human subjects ("common rule"). *DHHS Version*, *45*.

van den Hout, M. A., Engelhard, I. M., & McNally, R. J. (2017). Thoughts on experimental psychopathology. *Psycopathology Review*, *4*, 141–154. doi:10.5127/pr.045115

van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology? A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. doi:10.1016/j.jesp.2016.03.004

Vuorre, M., & Curley, J. P. (2018). Curating research assets: A tutorial on the git version control system. *Advances in Methods and Practices in Psychological Science*, *1*, 219–236. doi:10.1177/2515245918754826

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values.

Psychonomic Bulletin & Review, *14*, 779–804.

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven selfish reasons for preregistration. *APS Observer*, *29* (9).

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189. doi:10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638. doi:10.1177/1745691612463078

Walsh, C. G., Xia, W., Li, M., Denny, J. C., Harris, P. A., & Malin, B. A. (2018). Enabling open-science initiatives in clinical psychology and psychiatry without sacrificing patients' privacy: Current practices and future challenges. *Advances in Methods and Practices in Psychological Science*, *1*, 104–114. doi:10.1177/2515245917749652

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology an empirical comparison using 855 - tests. *Perspectives on Psychological Science*, *6*, 291–298. doi:10.1177/1745691611406923

Wolitzky-Taylor, K. B., Arch, J. J., Rosenfield, D., & Craske, M. G. (2012). Moderators and non-specific predictors of treatment outcome for anxiety disorders: A comparison of cognitive behavioral therapy to acceptance and commitment therapy. *Journal of Consulting and Clinical Psychology*, *80*, 786–799. doi:10.1037/a0029418

Figure 1. The ‘Create project’ tab of *pss*. The user can create a project by providing an informative name in the corresponding box (see below the “Project name” time) and by clicking on the “Create new project” button. For more details, please see the corresponding online tutorial (https://github.com/AngelosPsy/pssr_tutorial).

Figure 2. The 'Record changes' tab of *pss*. Here, the user should provide a name and an email address. After that, the user can see the changes that were made on each folder by clicking on the 'Track Changes' button. By providing a name and clicking on 'Timestamp changes', the user has timestamped all changes in the project. A list of changes is provided in the 'Version Control' tab. For more details, please see the corresponding online tutorial (https://github.com/AngelosPsy/pssr_tutorial).

Figure 3. The ‘Preregistration’ tab of *pss*. The user needs to type in a name and then select the template that will be used. Once the template is selected and the name of the project is given, a new window will appear that allows the user to write up the preregistration document. After finishing the write up of the document, the user may render the document by pressing the ‘render’ button. This will create a pdf document of the preregistration document. For more details, please see the corresponding online tutorial (https://github.com/AngelosPsy/pssr_tutorial).

Figure 4. The 'Anonymize data' tab of *pss.* The user needs to upload a data set, one at a time, and choose the column with the data to be anonymized (here the 'x' column). Afterwards, the software will create a new copy of the data with now the columns filled in with random numbers (default option). For more details, please see the corresponding online tutorial (https://github.com/AngelosPsy/pssr_tutorial).

Appendix

Steps checklist

Step 1 : Determination of research questions and predictions

– [ ] Confirmatory hypotheses.

– [ ] Predictions.

– [ ] Exploratory hypotheses.

– [ ] Predictions.

Step 2 : Determine the methods and statistical plan before data collection

– [ ] *Methods.*

> [ ] Stimuli.

> [ ] Procedure.

> [ ] Protocol.

> [ ] Dependent variable(s).

> [ ] Independent variable(s).

– [ ] *Statistical analyses.*

> [ ] Dependent variable(s) by name.

> [ ] Independent variable(s) by name.

> [ ] Type of statistical test to be used.

> [ ] Data reduction.

> [ ] In case of frequentist analyses: determine alpha level, power, expected effect.

> [ ] In case of Bayesian analyses: determine prior distributions, define expected level of strong evidence.

> [ ] In case of model selection: determine model parameters, comparison criteria and if applicable prior distributions.

> [ ] In case of correlational analyses: determine predicted correlation coefficient, alpha level, power.

> [ ] Creation of analysis scripts.

Step 3: Material collection

– [ ] Information brochure, informed consent.

– [ ] Study protocol.

– [ ] Experimental task and/or questionnaires.

– [ ] Licensing of all material.

Step 4: Pilot study

– [ ] Run pilot study

– [ ] Modifications in the current protocol

Step 5: Study's pre-registration

– [ ] Pre-registration on an website (e.g., osf.io, aspredicted.org).

– [ ] Time stamp of the preregistration project.

Step 6: Upload data and results report

– [ ] Anonymization of all data.