ASHA American Speech-Language-Hearing Association

JSLHR

Tutorial

Preregistration: Practical Considerations for Speech, Language, and Hearing Research

Violet A. Brown[a] and **Julia F. Strand**[b]

[a] Department of Psychological & Brain Sciences, Washington University in St. Louis, MO [b] Department of Psychology, Carleton College, Northfield, MN

ARTICLE INFO

Article History:
Received May 31, 2022
Revision received July 18, 2022
Accepted August 27, 2022

Editor-in-Chief: Cara E. Stepp
Editor: Rachel M. Theodore

https://doi.org/10.1044/2022_JSLHR-22-00317

ABSTRACT

Purpose: In the last decade, psychology and other sciences have implemented numerous reforms to improve the robustness of our research, many of which are based on increasing transparency throughout the research process. Among these reforms is the practice of preregistration, in which researchers create a time-stamped and uneditable document before data collection that describes the methods of the study, how the data will be analyzed, the sample size, and many other decisions. The current article highlights the benefits of preregistration with a focus on the specific issues that speech, language, and hearing researchers are likely to encounter, and additionally provides a tutorial for writing preregistrations.
Conclusions: Although rates of preregistration have increased dramatically in recent years, the practice is still relatively uncommon in research on speech, language, and hearing. Low rates of adoption may be driven by a lack of understanding of the benefits of preregistration (either generally or for our discipline in particular) or uncertainty about how to proceed if it becomes necessary to deviate from the preregistered plan. Alternatively, researchers may see the benefits of preregistration but not know where to start, and gathering this information from a wide variety of sources is arduous and time consuming. This tutorial addresses each of these potential roadblocks to preregistration and equips readers with tools to facilitate writing preregistrations for research on speech, language, and hearing.
Supplemental Material: https://doi.org/10.23641/asha.21644843

When a researcher accesses an empirical article in the *Journal of Speech, Language, and Hearing Research*, their goal is likely to use it to inform the design of their own experiments and ensure that their work is well situated in the literature. A clinician's goal may be to optimize treatment for a client through rigorous evaluation of previous work, weighing the costs (e.g., time, effort, money) and benefits (e.g., improved quality of life for the client) of particular rehabilitation methods. Whether the goal is experimentation or rehabilitation, these outcomes are best achieved when the reader has access to detailed information about how the study was run. For example, if an effect only emerges in one of three versions of an experiment or only when one of many possible outlier exclusion criteria was employed, it is important that the reader is aware of that information so they can better understand the boundaries of the effect. Thus, as writers, we can increase the value of our research by reporting it as thoroughly and transparently as possible.

Despite the benefits of transparency, several barriers may hamper a researcher's ability to transparently report all features of an experiment. As a hypothetical example, imagine that a researcher analyzes a dataset and finds null results that do not support their hypothesis. However, upon further inspection, the researcher notices that several participants performed especially poorly on the task and discovers that excluding those participants produces results that are statistically significant and in line with their hypothesis. When the researcher writes the manuscript, they do not mention that the outcome of the study is contingent on excluding those participants (i.e., they only describe the second analysis). Although the decision

Correspondence to Violet A. Brown: violet.brown@wustl.edu. **Publisher Note:** This article is part of the Forum: Promoting Reproducibility for the Speech, Language, and Hearing Sciences. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

to exclude these participants may be perfectly justified (e.g., data from participants who fell asleep during the study are not informative), different researchers might make different decisions about what is reasonable, and in only reporting the second analysis, the researcher has obscured information about the specificity of the findings.

If the analysis that produced null results was the one the researcher had originally planned to conduct, why might they have omitted it from the manuscript? One possibility is that they simply forgot. That is, sufficient time may have passed between when the researcher ran the analyses and when they wrote the paper that they did not remember they had run the analysis both ways. Alternatively or in addition, the choice to only report the second analysis may be influenced by human biases in decision making such as confirmation bias (i.e., seeking out or interpreting information in ways that are consistent with our expectations; Nickerson, 1998). The incentive structure of academia may also deter the researcher from wholly transparent reporting: Statistically significant findings with unambiguous conclusions are most likely to get published (i.e., publication bias; Rosenthal, 1979), and success in academia depends on the individual's publication record. The researcher may therefore omit the first analysis from the manuscript to tell an uncomplicated story. Regardless of which of these factors were at play, readers lost access to information about the research because the results were reported selectively.

Given the barriers to transparency, what can speech, language, and hearing researchers do to facilitate transparent reporting and therefore ensure that our findings are maximally useful to the consumers of our work? One method of increasing research transparency is *preregistration*, whereby a researcher creates a time-stamped and uneditable document specifying their design and analysis plan before the study is run (Nosek et al., 2018). Not only does this provide researchers with a mechanism to document these decisions before they are forgotten, but it also reduces the potential for cognitive biases to influence decision making. If the exclusion criteria, analysis plan, and other research decisions are made before the results are known, it removes the possibility of study outcomes influencing those choices.

Despite the benefits of preregistration (e.g., Nosek et al., 2018; Sarafoglou et al., 2022) and its widespread adoption throughout psychology and other disciplines (Nosek & Lindsay, 2018), the practice has not become standard in research on speech, language, and hearing.[1] Indeed, a recent survey among researchers in communication sciences and disorders (El Amin et al., 2022) found that although the desire to learn about preregistration was high, knowledge and implementation of preregistration were low (see also Alexander & Green, 2021). Respondents identified several barriers to preregistration (e.g., "lack of time"), but the most common response was "I don't know how to preregister my work" (El Amin et al., 2022, p. 11). The goal of the current article is therefore to help overcome these barriers by describing why preregistration is beneficial for speech, language, and hearing research (see Roettger, 2021, for similar arguments about linguistics research) and providing a guided tutorial—including a sample "gold standard" preregistration document—to make the process of preregistration more accessible to novices.

Why Preregistration Benefits Speech, Language, and Hearing Research

Researchers studying speech, language, and hearing (like all subdisciplines) have flexibility in how they design experiments and analyze data. Many of these choices (e.g., which covariates to include, which participants to exclude) do not have clear-cut "correct" answers, so researchers must make a selection from many reasonable options. This flexibility creates a "garden of forking paths" (Gelman & Loken, 2013), whereby each decision leads the researcher down a different path. Given the number of decisions that must be made along the way, it would not be surprising if the various paths led to different outcomes even with the same data and hypotheses being tested. Below, we provide several examples of these *researcher degrees of freedom*—choices that researchers must make throughout the research pipeline (Simmons et al., 2011)—that are particularly relevant for speech, language, and hearing researchers. These examples were selected because they demonstrate the feasibility of many available options throughout the decision tree and highlight the value of preregistration for keeping track of which decisions were made and when (see also Roettger, 2019).

In research relying on response time tasks, researchers often opt to exclude observations that are deemed outliers for being too slow or too fast. However, the criteria used to arrive at this determination vary considerably across experiments (see Simmons et al., 2011). In the *Journal of Speech, Language, and Hearing Research* alone, outliers have been defined as responses that fall more than 2 *SD*s beyond the mean response for each item (Krestar & McLennan, 2019); more than 3 *SD*s from a participant's mean response time for each task (Strand et al., 2018); more than 3 *SD*s from the mean response time for the block (Iuzzini-Seigel, 2021);

[1]Although the choice of whether to preregister is made by individuals, it can be encouraged by journals in their submission guidelines (e.g., journal badges; Kidwell et al., 2016) and may be required by granting agencies (e.g., clinical trials funded through the National Institutes of Health are required to be preregistered via clinicaltrials.gov).

Table 1. Covariates included in four different studies assessing the effect of clustering coefficient on spoken word recognition or word learning.

Covariate	Chan & Vitevitch (2009)	Altieri et al. (2010)	Goldstein & Vitevitch (2014)	Liben-Nowell et al. (2019)
Phonotactic probability	x	x	x	
Neighborhood density	x	x	x	x
File duration	x		x	
Familiarity	x	x		
Word frequency	x	x		x
Spread of the neighborhood	x			
Neighbors in a given position	x			
Neighborhood frequency	x	x		
Distribution of phonemes	x			
Semantic set size			x	
Biphone probability			x	
Stimulus onset			x	
Stimulus offset			x	
Stimulus duration			x	
Concreteness rating			x	
First-word associate strength			x	
Second-word associate strength			x	

Note. See studies for more details about what each covariate represents and how the variables were calculated/measured.

more than 3 *SD*s from the by-participant, by-block mean (Earle & Ullman, 2021); above or below a stable threshold (Schwartz et al., 2016); or for unspecified reasons. These extreme values are sometimes removed (as in the studies above) or sometimes replaced with a fixed value (e.g., values greater than 2 *SD*s above the mean replaced with the value 2 *SD*s above the mean; Smith, 2011).

In addition to flexibility in choosing which observations to exclude, researchers also have flexibility in the covariates included in statistical analyses. For instance, if a researcher is interested in how a particular lexical characteristic influences word identification, it is common practice to statistically control for the influence of other lexical variables. However, studies differ widely on the particular covariates they include. For example, in testing the effects of clustering coefficient (the proportion of phonological neighbors of a target word that are also neighbors of each other; Chan & Vitevitch, 2009), research teams have opted to statistically control for different combinations of at least 17 lexical covariates (see Table 1). Even after researchers have identified covariates that may be relevant, they also have flexibility in choosing how to operationalize them. For example, although three of the studies shown in Table 1 included word frequency as a covariate, two different measures of word frequency were used (Brysbaert & New, 2009; Kucera & Francis, 1967).

Beyond excluding outliers and selecting covariates, researchers must make decisions about sample size, method of analysis, whether and how to transform variables, and so on, and each of these decisions has the potential to substantially affect outcomes (Simmons et al., 2011). In a recent study, researchers provided multiple independent teams with the same dataset and asked them to test two hypotheses regarding whether scientists' genders and professional statuses affect verbosity during group meetings (Schweinsberg et al., 2021; see also Silberzahn et al., 2018). Crucially, choices about variable operationalization and data analysis were made independently by each research team. As a result, research teams differed substantially in both how they operationalized variables (e.g., "verbosity" could be quantified using the number of comments made or the number of words within comments) and how they conducted their analyses (e.g., the type of statistical tests run, covariates included). This analytic flexibility produced statistically significant findings in both directions as well as null effects. Thus, even seemingly straightforward research questions can produce a garden of forking paths, and researchers embarking upon different (perfectly reasonable) paths may come to different conclusions.[2]

These findings demonstrate the value in making experimental decisions before knowing how those decisions affect study outcomes. If conflicting conclusions can be derived from the same dataset, making these decisions and justifying them beforehand reduces bias and ensures that those decisions are driven by what the researcher thinks is most appropriate rather than how the decisions affect the results. Specifying decisions before data collection also reduces the likelihood that researchers will intentionally leverage the flexibility of the process by testing multiple combinations of choices until they find one that

[2]It is worth noting that the Schweinsberg et al. (2021) study was limited to analytic flexibility, but additional researcher degrees of freedom are also introduced through study design, stimulus selection, and so on.

is statistically significant (i.e., *p*-hack; see Wicherts et al., 2016). Preregistration therefore increases research transparency by clarifying the point in the research process at which each decision was made.

How to Preregister

A preregistration is a written document, finalized before data collection, that outlines the research plan. The key features of a preregistration are that it is time-stamped, uneditable, and includes a stable link that can be viewed by peer reviewers (during the review process) and readers (after the paper is published). Below, we catalog information that is particularly useful to consider when preregistering speech, language, and hearing research. It is worth noting that these suggestions, when implemented in full, serve as an example "gold standard" preregistration, but the level of detail included in a preregistration can vary. For example, a lab may opt to preregister the study procedure and intended sample size, but not the analysis plan. Although it is certainly desirable to specify as much detail as possible, preregistering any information about the study is useful, provided that the particular aspects of the study that were and were not preregistered are transparently reported in the manuscript. Researchers who are just beginning to implement preregistration may prefer to start by preregistering certain components of their experiment and add more detail to future preregistrations as they gain familiarity with the process (though see the "Addressing Deviations From the Preregistration" section below to help ease concerns about having to modify a preregistered plan).

The most common platform for preregistering research in psychology is the Open Science Framework (OSF; https://osf.io/). The OSF offers several formats for preregistration that differ in the number of questions and level of detail required, ranging from a single open-ended registration without a prompt to forms that contain a variety of specific prompts regarding experiment design and analysis.[3] Researchers within a team can collaboratively work on the preregistration on the OSF and receive a link to the preregistration document once it is submitted.

Critically, the OSF provides the option to "embargo" the preregistration—meaning that the document is completely hidden from everyone except the creator of the project and the people they choose to add as collaborators—until a date specified by the researcher. Embargoing ensures that only people involved in the project can access the preregistration before the paper is submitted, thereby alleviating concerns of being "scooped."[4] The preregistration can be unembargoed at any point, which is necessary before the paper is submitted for publication so it can be accessed by reviewers.

Here, we include a sample preregistration (see Supplemental Material S1) based on a previous study conducted by one of the authors (Brown et al., 2021) that has been modified for the purposes of this tutorial (the original, unedited preregistration can be accessed at https://osf.io/g2j94). This example uses the "AsPredicted" preregistration format (one of the several formats the OSF offers).[5] We opted to use the "AsPredicted" format because it provides some structure by asking targeted questions about the experimental design and analysis, but is short enough that it is still accessible for people with limited experience. It contains seven open-ended prompts (specifying hypotheses, dependent variables, conditions, analyses, outliers and exclusions, sample size, and a free-response section to include additional details) along with other questions about whether data collection is already underway (which is discouraged) and what type of study is being conducted (e.g., experiment or survey). Although the sample preregistration is specific to a particular research question ("How do different types of face masks affect speech intelligibility in young and older adults in various levels of background noise?"), we use it to highlight issues that are likely to arise in research in many related areas.

In the following sections, we walk through the headers that appear in most manuscripts reporting findings related to speech, language, and hearing research (e.g., Introduction, Participants, Procedure, Analysis) and describe some of the decisions that we recommend preregistering.[6] Note that preregistration documents are not typically formatted like manuscripts, but we categorize the decisions this way below so that our suggestions can flexibly be applied to any preregistration template.

[3]Although most preregistration options on OSF are plain text, there is also an option to create an "open-ended registration" in which researchers can upload a PDF that can contain code, images, tables, or other information beyond text.

[4]It is worth noting, however, that "scooping" is far less common and has fewer adverse consequences on academic careers than people typically assume ("How to Deal with Being 'Scooped,'" 2016; "It's Not the End of the World," 2016; Laine, 2017; "Who's Afraid of Open Data," 2015).

[5]Note that it is also possible to preregister research using this format directly on AsPredicted.com. However, the OSF has several features that are appealing, such as adding data and stimuli to the same page after the experiment is run to consolidate all research materials.

[6]This tutorial is intended to be used for new empirical work, but it is also possible to write a preregistration for a secondary data analysis (see Weston et al., 2018). For example, a researcher may use an existing database of auditory lexical decision times (e.g., Tucker et al., 2019) to assess how a previously unexplored lexical variable affects response times. The data have already been collected in this case, but the researcher can still preregister an analysis plan as long as they have not analyzed the data (e.g., which items they will include, which covariates they will control for; see also Roettger, 2021).

Introduction

Because preregistrations are primarily concerned with methods and analyses, most information that appears in the introduction section of a paper does not appear in the preregistration. The only exception to this is that hypotheses or research questions—which often appear near the end of an introduction section in a manuscript—are often included in the preregistration.

Hypotheses

Preregistrations can include explicit, directional hypotheses about both main effects (e.g., "We hypothesize that older adults will have lower sentence identification scores...") and interactions ("...and be more affected by the presence of background noise than young adults, leading to an interaction between age group and noise level"). Although some findings in speech, language, and hearing research are robust and the hypotheses are clear even without preregistration (e.g., production of common or contextually expected words are phonetically reduced), many others are not. Thus, explicitly preregistering all hypotheses guards against *Hypothesizing After the Results are Known* (HARKing; Kerr, 1998), that is, reporting post hoc findings as if they had been predicted a priori. In addition to being an ethically dubious practice, a danger of HARKing is that if hypotheses are developed based on the results of a study, it becomes impossible to falsify the resulting theories. Adding to this concern, if the finding was a false positive, HARKing can turn those Type I errors into theory (see Kerr, 1998; Rubin, 2017).

There may also be cases in which a research team does not have a clear, directional hypothesis. The sample preregistration gives examples of how to report directional hypotheses (e.g., Lines 25–27) as well as hypothesized group differences without specific directional predictions (Lines 20–23). When a researcher does not have a directional hypothesis, it is perfectly acceptable to report why the question is interesting and what the results would mean without making a specific prediction (provided that the manuscript matches this style of reporting).

Method

Participants

The preregistration should specify the number of participants or observations that will be included in the final analysis. Note that this is different from specifying the number of participants run in the study, as an unpredictable number of participants or observations may be excluded from final analyses. In most cases, we recommend excluding any participants with usable data that were collected after the preregistered sample size was reached. For example, if you preregister a sample size of 96 participants but have usable data from 99, we suggest discarding data from the final three participants.[7] Although it may be argued that having more data is preferable, provided that your experiment is sufficiently powered, it is more straightforward to justify excluding those participants and sticking with your preregistered sample size than including them and justifying why you deviated from your preregistration.

It is worth noting, however, that removing "extra" participants may not be appropriate for experiments involving difficult-to-reach samples (e.g., cochlear implant patients) or those in which data collection is particularly time consuming and expensive (e.g., some functional magnetic resonance imaging research). Indeed, in these cases, the preregistered sample size may reflect the maximum feasible sample size rather than the sample size necessary to obtain a desired level of statistical power, so including data from additional participants is advantageous. In these cases, we urge researchers not to analyze the data until data collection is complete and to justify the choice explicitly in the manuscript (and in the preregistration if this scenario was anticipated). Furthermore, there may be ethical reasons to include all data even if the sample size is larger than what was preregistered. In clinical settings, participants may have made sacrifices such as forgoing other treatments to participate in the study, so including their data may be the more ethically justifiable choice (see Sajdak et al., 2013).

In addition to preregistering a participant sample size, researchers may also opt to specify their sample size in terms of the number of observations in each condition. There are some instances in which this may not be applicable (e.g., missing data are unlikely, participant means are entered into the analysis, the sample size represents the maximum feasible number of individuals that can be run in the study). However, if the power analysis is based on a certain number of observations in each condition and the actual number of observations per condition is less than the anticipated number, collecting data from the preregistered sample size will lead to an underpowered study. This may be particularly relevant in response time studies in which observations are only included in the final analysis if the associated response was correct. As one

[7]Note that although excluding three additional participants is not likely to affect study outcomes, having a consistent rule of excluding additional participants removes the temptation to analyze data multiple times and decide whether to include the additional participants based on how they affect the findings. Another option would be reporting the analyses including the additional participants, but also running all analyses excluding those participants to ensure that the decision to include them did not substantially affect study outcomes.

example, in one of our previous studies (Brown & Strand, 2019), we preregistered a sample size of 50 participants but noted that this number was based on the assumption that we would include response times from approximately 100 of the 136 trials for each participant (i.e., 5,000 observations per condition in this within-subjects study). Given that it is impossible to perfectly predict accuracy at any task, we noted in the preregistration that if we did not have at least 5,000 usable observations in each condition after collecting data from 50 participants, we would continue collecting data until we reached that threshold. This required collecting data from 53 participants, which was more than our preregistered sample size but ensured that our experiment was sufficiently powered and was consistent with our preregistration.

Stimuli

Many decisions about speech, language, and hearing research (e.g., which participants to exclude) may be made either before the study starts or after it ends, and preregistration helps reveal when these decisions were made. Decisions about which stimuli to include, however, typically cannot be altered after data are collected, so information about which stimuli were presented is a choice that must be made ahead of time. Thus, the main purpose of describing the stimuli in a preregistration document is to make it clear that what you report in the paper is a full representation of all the stimuli that were used (e.g., a study that collected data at four signal-to-noise ratios reported all four). Although it should be standard practice (even in the absence of preregistration) to report all conditions that were run or at least justify the decision to omit conditions in the paper, the presence of a preregistration document helps alert readers to conditions that were run but not reported, thereby increasing transparency about the research process.

Procedure

Preregistrations typically detail the tasks participants will complete and the order in which they will complete them. We recommend including sufficient detail that an independent research team could conduct your experiment based on the description of the methods in the preregistration (see Lines 64–76 in the sample preregistration for an example of the level of detail recommended). For example, rather than simply saying that participants will complete a spoken word identification task in noise, it would be preferable to report the type of words, type and level of background noise, number of trials per condition, whether the conditions were blocked (and counterbalanced) or intermixed, how participants responded to the task (e.g., typed vs. verbal responses), and so on. It may also be important to describe any catch trials or attention-check procedures and how frequently they occurred. For online auditory research, any procedures to ensure that participants were wearing headphones should also be described in the preregistration when appropriate (Eerola et al., 2021; Woods et al., 2017). As with information about stimuli, one of the goals of detailing the procedures in the preregistration is to make it clear that the final paper reports all conditions and procedures that were included in the experiment, which deters selective reporting of conditions or procedures and therefore gives a complete picture of how the work was conducted.

Another advantage of preregistering the methods is that methodological details are documented immediately after they are made, thereby more closely linking the writing phase of the research process to the design phase. Although preregistration certainly increases workload on the front end of a project, it streamlines the writing process down the line because methodological details that may have been forgotten (e.g., "What was the interstimulus interval? Was there a fixation cross or just a blank screen?") have already been documented. Thus, the preregistration can easily be modified to become the "Method" section of a manuscript.

Results

Exclusion Criteria

It is common to exclude observations (e.g., individual response time trials) when they are thought to be too unrealistically short or long to represent the phenomenon of interest. For example, completing a word/nonword judgment for spoken words typically takes less than 1 s, so responses that are considerably longer (e.g., 5 s) are likely not reflecting the process of searching for and accessing a word in the mental lexicon. Similarly, word/nonword judgments cannot possibly occur in 20 ms, so extremely short response times like these are erroneous. Excluding extreme observations may therefore allow researchers to better understand the mechanisms underlying the processes of interest by reducing noise in the dataset. We therefore recommend preregistering details about what constitutes an outlier (when applicable).

As one example from our work, given that response times tend to be skewed, we typically exclude response time trials that are more than 3 median absolute deviations from a participant's median response time in that condition (see Leys et al., 2013). For example, if a participant's median response time in a particular condition is 700 ms and the median absolute deviation is 200 ms, response times below 100 ms and above 1,300 ms would be excluded for that participant in that condition. If defining exclusion criteria seems like a daunting task, it may be helpful to analyze or visualize previous datasets to gain insight into the kinds of exclusions that may be important to specify.

In addition to excluding individual trials, it may also be useful to plan to exclude all data from participants who meet certain criteria. For example, a participant who types "asdf" on 90% of trials may not be providing quality data on the 10% of trials they do respond to. In our work, we typically preregister excluding participants who disclosed completing the study incorrectly (e.g., after completing a study on audiovisual speech perception they mention to the experimenter that they kept their eyes closed or looked away from the screen to focus on listening), encountered technical difficulties (e.g., the computer crashed during the experiment), or failed attentional checks (e.g., did not respond to some proportion of periodic "catch" trials). In addition to these exclusion criteria—which are made without regard to how the participant performed on the task—researchers may also opt to exclude participants based on poor performance on a task. For example, we often preregister a plan to exclude participants whose mean speech identification accuracy is more than 3 *SD*s below the mean (across participants) in that condition (see Lines 151–153 in the sample preregistration).

Note, however, that in the context of speech identification in noise studies, when an experiment includes an easy condition in which accuracy is near ceiling and standard deviations are small, this can produce a cutoff for exclusion that is unreasonably high. For example, if mean accuracy is 98% and the standard deviation across participants is 0.5%, participants with accuracies below 96.5% would be excluded based on a 3-*SD* exclusion criterion. However, few would argue that an individual completing a speech identification task with 96.5% accuracy should be excluded because of "poor accuracy." We therefore recommend that researchers include a note in their preregistrations to protect against unnecessarily excluding individuals from analyses. In our preregistrations, we often explicitly state that the 3-*SD* cutoff does not apply if a participant has over 90% accuracy in that condition (see Lines 153–158 in the sample preregistration). As mentioned above, preregistering a final sample size (rather than the number of participants run) means that the participants who are excluded from the analysis are replaced.

Correcting Typos

In our area of research (speech perception), it is common to correct or clean participant responses on open-set tasks. For example, a target word like "thief" often has artificially low accuracy because many participants forget "*i* before *e* except after *c*." It is therefore helpful to outline rules for correcting typos that can be applied blind to condition, and include additional description of whether inflected forms (e.g., pluralizations), homophones, or other deviations from the target will be counted as correct (see sample preregistration, Lines 44–52).

Analyses

Depending on the complexity of the analytical plan, this section of the preregistration is often the longest and most detailed. Indeed, this is the point in the research pipeline at which researchers discover whether their hypotheses were supported and may therefore be especially susceptible to embarking upon several forking paths (i.e., tweaking their analysis plan) if all analytical decisions have not been outlined in a preregistration.[8] Even in cases in which the analytic procedure seems straightforward, researchers must decide which particular comparisons to conduct, how they will infer statistical significance (e.g., likelihood ratio vs. Wald test), whether they will correct for multiple comparisons, whether variables will be analyzed continuously or categorically, which covariates to include, and so on.

As the analysis becomes more complex, the number of researcher degrees of freedom increases dramatically. One specific example involves linear mixed-effects models, which are being implemented with increasing frequency in speech, language, and hearing research. If a researcher plans to analyze their data with these models, it is helpful to specify not only the outcome and predictors, but also the precise random effects structure that will be modeled, and to justify these decisions in the preregistration (see Lines 87–89 in the sample preregistration).

An additional consideration regarding mixed-effects modeling is what steps will be taken if models fail to converge or produce singular fits. Despite recommendations not to arbitrarily remove theoretically important random effects from the model specification in cases of nonconvergence (Barr et al., 2013; Brown, 2021), this practice is common. In some cases, the data may not support a complex random effects structure and it may be necessary to remove those random effects, but the steps leading to their omission should be described and justified in the preregistration to limit these researcher degrees of freedom to the extent possible (see Lines 94–105 in the sample preregistration). For example, researchers could note that the random slope that contributes the least to the total variance will be removed if necessary, but only as a last resort if adjusting control parameters does not facilitate convergence, and only if likelihood ratio tests indicate that removing the random effect is justified (Matuschek et al., 2017; see Roettger, 2021, for another example of how to address the possibility of nonconvergence in a preregistration).

[8]This *p*-hacking may not be malicious—when a researcher has a theory predicting a particular pattern of results, it is easier to convince oneself that analytical decisions are justified when they lead to expected rather than unexpected results. This is why it is important to explicitly state which decisions were made beforehand and which were made after seeing the data, a practice that is facilitated by preregistration.

In addition to limiting researcher degrees of freedom by specifying the analyses ahead of time, being specific in the preregistration about which particular analyses you plan to conduct can also have benefits during the review process. Indeed, having a preregistered analysis plan can help authors counter reviewer suggestions to include additional analyses that the authors believe are outside the scope of the project. In our own revisions, we have replied to reviewer suggestions for additional analyses by explaining that we prefer not to deviate from our preregistered plan and noting that interested readers may feel free to run additional analyses on our data (a benefit of publicly sharing data and code). Similarly, having a time-stamped document clearly describing all experimental conditions protects authors in situations in which a reviewer requests that an experiment be omitted from a paper.

Addressing Deviations From the Preregistration

Preregistration has been described as "a plan, not a prison" (DeHaven, 2017), and this point warrants emphasis: If circumstances change and it becomes necessary to deviate from the preregistered plan, those changes simply need to be transparently described in the manuscript. Preregistrations are not binding contracts; they are documents that describe the intended plan for a research project and clarify when decisions were made to keep researchers honest with themselves and with others.

Researchers may discover issues that require deviating from the preregistered plan during data collection. For example, they may identify a problem with the study that necessitates stopping it, making a methodological change, and then rerunning it. Indeed, even with pilot testing, a task may produce better or worse performance than expected, and if performance is at ceiling or floor, it may be impossible for the effects of interest to emerge. In this case, once the new level of difficulty is determined, the researcher can simply create a new preregistration within the same OSF project that links to the original preregistration and describes the change (see https://osf.io/8rejp for an example of this kind of updated preregistration from our lab).

It may also be necessary to deviate from a preregistration during data analysis and reporting (i.e., after data have been collected). As one example, in a recent paper from our lab (Brown, Fox, & Strand, 2022), we reported data from 135 participants in the manuscript despite preregistering that our final analysis would include data from 136 participants. COVID-related campus closures made it impossible for us to collect additional data, so we simply noted and justified this deviation from the preregistration in the "Participants" section of the manuscript.

Deviations from the preregistration may also come in the form of analytical decisions. For example, in one of our studies, we preregistered that we would remove response times slower than 2,000 ms (the task was quite simple, so response times this slow were deemed erroneous). However, we realized that because one of the two conditions was expected to result in slower response times, using the same cutoff in both conditions would systematically remove more (slow) response times from the harder condition, which would not only bias results but also make it more difficult to detect the effect of interest (see Brown & Strand, 2019). This realization was made while data collection was underway (i.e., before analyzing any data) after learning about a more appropriate method for removing response time outliers (Leys et al., 2013), so we simply noted this deviation from the preregistration in the manuscript. In cases like these, it may be advisable to specify in the manuscript whether the outcome is contingent on the decision (i.e., conduct the analyses both ways and report whether the results were consistent). Note, however, that this does not mean that a researcher *must* conduct all preregistered analyses; indeed, it may be that a preregistration included an analysis that upon further consideration is not appropriate given the data and research question. The goal of preregistrations is to indicate *when* decisions are made, not to dictate *which* decisions are made.

Similarly, having preregistered an analysis plan does not preclude conducting exploratory analyses in addition to the preregistered ones. In other words, preregistrations "do not tie researchers' hands, but merely uncover readers' eyes" (Nelson et al., 2018, p. 519). It is perfectly acceptable to present exploratory analyses (i.e., any analyses beyond what you preregistered) in the manuscript, as long as it is clear which analyses were confirmatory and which were exploratory. If additional analyses are relatively straightforward, it may be sufficient to simply describe the exploratory analyses along with the preregistered analyses in the "Results" section of the manuscript and note which analyses were exploratory (e.g., "An exploratory analysis revealed that. . ."). If there are several exploratory analyses or detailing the analyses would be cumbersome in the context of the preregistered analyses, it may be more appropriate to include subheaders in the manuscript to clearly distinguish the preregistered analyses from the exploratory analyses (see Brown, Dillman-Hasso, et al., 2022, for an example of one of our papers that uses both approaches). There is no wrong way to report exploratory analyses, provided that they are explicitly described as such.

Conclusions

In the last decade, many reforms have been implemented to increase transparency about the research process. Preregistration provides a powerful tool for increasing transparency by reducing researcher degrees of freedom, counteracting cognitive biases, and overcoming the

fallibility of human memory. Of course, preregistration is not a panacea for low rates of replicability; it is possible to preregister a study with major methodological flaws, and preregistered studies with null effects may still be difficult to publish given publication bias (Rosenthal, 1979). However, along with more transparent reporting in manuscripts; sharing data, code, and materials; incentivizing replication attempts (Simons, 2014); and publishing under the Registered Report format (Chambers, 2013), preregistration can help build a more robust and replicable literature in speech, language, and hearing research.

Data Availability Statement

There are no data associated with this tutorial.

Acknowledgments

This work was funded by the National Science Foundation through a graduate research fellowship awarded to Violet Brown (DGE-1745038) and the National Institute on Deafness and Other Communication Disorders via a grant to Julia Strand (R15-DC018114). The authors are grateful to Emily Hazlett for feedback and additional support.

References

Alexander, J., & Green, J. A. (2021). Publication bias and evidential value in speech, language, and hearing research. *OSF Preprints*. https://doi.org/10.31219/osf.io/vncbs

Altieri, N., Gruenenfelder, T., & Pisoni, D. B. (2010). Clustering coefficients of lexical neighborhoods: Does neighborhood structure matter in spoken word recognition? *The Mental Lexicon, 5*(1), 1–21. https://doi.org/10.1075/ml.5.1.01alt

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1), 251524592096035. https://doi.org/10.1177/2515245920960351

Brown, V. A., Dillman-Hasso, N. H., Li, Z., Ray, L., Mamantov, E., Van Engen, K. J., & Strand, J. F. (2022). Revisiting the target–masker linguistic similarity hypothesis. *Attention, Perception & Psychophysics, 84*(5), 1772–1787. https://doi.org/10.3758/s13414-022-02486-3

Brown, V. A., Fox, N. P., & Strand, J. F. (2022). "Where are the. . .fixations?": Grammatical number cues guide anticipatory fixations to upcoming referents and reduce lexical competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(5), 643–657. https://doi.org/10.1037/xlm0001019

Brown, V. A., & Strand, J. F. (2019). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition, 2*(1), 44. https://doi.org/10.5334/joc.89

Brown, V. A., Van Engen, K. J., & Peelle, J. E. (2021). Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults. *Cognitive Research: Principles & Implications, 6*(1), 49. https://doi.org/10.31234/osf.io/7waj3

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Chambers, C. D. (2013). Registered reports: A new publishing initiative at cortex. *Cortex, 49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1934–1949. https://doi.org/10.1037/a0016902

DeHaven, A. (2017). *Preregistration: A plan, not a prison*. Center for Open Science. https://cos.io/blog/preregistration-plan-not-prison/

Earle, F. S., & Ullman, M. T. (2021). Deficits of learning in procedural memory and consolidation in declarative memory in adults with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 64*(2), 531–541. https://doi.org/10.1044/2020_JSLHR-20-00292

Eerola, T., Armitage, J., Lavan, N., & Knight, S. (2021). Online data collection in auditory perception and cognition research: Recruitment, testing, data quality and ethical considerations. *Auditory Perception & Cognition, 4*(3–4), 251–280. https://doi.org/10.1080/25742442.2021.2007718

El Amin, M., Borders, J. C., Long, H. L., Keller, M. A., & Kearney, E. (2022). *Open Science practices in communication sciences and disorders: A survey*. https://osf.io/jwba3/

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*. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition. *Frontiers in Psychology, 5,* 1307. https://doi.org/10.3389/fpsyg.2014.01307

How to deal with being "scooped": The vast majority of science is a process of derivative, incremental advance. (2016). Impact of Social Sciences. https://blogs.lse.ac.uk/impactofsocialsciences/2016/04/19/so-youve-been-scooped/

It's not the end of the world if your research gets "scooped". (2016). Times Higher Education (THE). https://www.timeshighereducation.com/blog/its-not-end-world-if-your-research-gets-scooped

Iuzzini-Seigel, J. (2021). Procedural learning, grammar, and motor skills in children with childhood apraxia of speech, speech sound disorder, and typically developing speech. *Journal of Speech, Language, and Hearing Research, 64*(4), 1081–1103. https://doi.org/10.1044/2020_JSLHR-20-00581

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

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency.

PLOS Biology, 14(5), Article e1002456. https://doi.org/10.1371/journal.pbio.1002456

Krestar, M. L., & McLennan, C. T. (2019). Responses to semantically neutral words in varying emotional intonations. *Journal of Speech, Language, and Hearing Research, 62*(3), 733–744. https://doi.org/10.1044/2018_JSLHR-H-17-0428

Kucera, H., & Francis, N. (1967). *Frequency and word association norms*. Brown University Press.

Laine, H. (2017). Afraid of scooping—Case study on researcher strategies against fear of scooping in the context of open science. *Data Science Journal, 16,* 29. https://doi.org/10.5334/dsj-2017-029

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liben-Nowell, D., Strand, J., Sharp, A., Wexler, T., & Woods, K. (2019). The danger of testing by selecting controlled subsets, with applications to spoken-word recognition. *Journal of Cognition, 2*(1), 2. https://doi.org/10.5334/joc.51

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lindsay, S. D. (2018). *Preregistration becoming the norm in psychological science*. Association for Psychological Science. https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology, 10*(1), 1. https://doi.org/10.5334/labphon.147

Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics, 59*(5), 1227–1249. https://doi.org/10.1515/ling-2019-0048

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

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

Sajdak, R., Trembath, L., & Thomas, K. S. (2013). The importance of standard operating procedures in clinical trials. *Journal of Nuclear Medicine Technology, 41*(3), 231–233. https://doi.org/10.2967/jnmt.113.121467

Sarafoglou, A., Kovacs, M., Bakos, B., Wagenmakers, E.-J., & Aczel, B. (2022). A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science, 9*(7), 211997. https://doi.org/10.1098/rsos.211997

Schwartz, R. G., Hestvik, A., Seiger-Gardner, L., & Almodovar, D. (2016). Processing binding relations in specific language impairment. *Journal of Speech, Language, and Hearing Research, 59*(6), 1384–1394. https://doi.org/10.1044/2016_JSLHR-L-15-0107

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson, E., Snellman, K., Amy Sommer, S., Otner, S. M. G., Robinson, D., . . . Luis Uhlmann, E. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes, 165,* 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Smith, P. A. (2011). Attention, working memory, and grammaticality judgment in typical young adults. *Journal of Speech, Language, and Hearing Research, 54*(3), 918–931. https://doi.org/10.1044/1092-4388(2010/10-0009)

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research, 61*(6), 1463–1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior Research Methods, 51*(3), 1187–1204. https://doi.org/10.3758/s13428-018-1056-1

Weston, S. J., Mellor, D. T., Bakker, M., van den Akker, O., Campbell, L., Ritchie, S. J., Chopik, W. J., Damian, R. I., Kosie, J., & Soderberg, C. K. (2018). *Secondary data preregistration*. https://osf.io/x4gzt/

Who's afraid of Open Data: Scientists' objections to data sharing don't stand up to scrutiny. (2015). *Impact of Social Sciences*. https://blogs.lse.ac.uk/impactofsocialsciences/2015/12/16/whosafraid-of-open-data-dorothy-bishop/

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7,* 1832. https://doi.org/10.3389/fpsyg.2016.01832

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics, 79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2