A call for greater transparency in piloting

Isaac J. Handley-Miner[1*], Agata Bochynska[2], Brett Buttliere[3], Kai Li Chung[4], Mahmoud Elsherif[5], Christopher J. Graham[6], Neele Henriette Heiser[7], Hannah D. Lönneker[8], Bjørn Sætrevik[9], Yashvin Seetahul[10], Justin Sulik[11], Johanna Tomczak[12], Caroline Zygar-Hoffmann[13,14], Mary Beth Neff[2*]

[1]Boston College, U.S., [2]University of Oslo, Norway, [3]University of Warsaw, Poland, [4]University of Nottingham Malaysia, Malaysia, [5]University of Leicester, UK, [6]Royal College of Physicians of Edinburgh, UK, [7]University of Geneva, Switzerland, [8]University of Tübingen, Germany, [9]University of Bergen, Norway, [10]University of Innsbruck, Austria, [11]LMU Munich, Germany, [12]Independent Researcher, UK, [13]Charlotte Fresenius Hochschule, University of Psychology, Germany, [14]LMU Munich, Germany

[*]Correspondence should be addressed to Isaac J. Handley-Miner (email: isaac.handley-miner@bc.edu) and Mary Beth Neff (email: marybethneff@gmail.com)

Note: Middle authors are organized alphabetically; authorship order and title are subject to change

preprint link: https://doi.org/10.31234/osf.io/q95zn_v1

Abstract

A pilot study is a preliminary investigation conducted before a full-scale research study. Although piloting is a common and valuable part of the research process, pilot studies are not systematically reported (or even acknowledged) in the psychology literature. This paper argues that establishing norms and guidelines for regularly reporting pilot studies can improve many aspects of psychological research. Specifically, reporting pilot studies can help researchers avoid systematically excluding part of the research process from the scientific record, reveal selection biases in research designs that may affect the generalizability of findings, evaluate others' research more accurately, learn from other researchers' successes and failures, uphold ethical responsibilities, and improve the methodological practice of piloting. We conclude with three concrete suggestions for near-term changes to help pilot reporting become standard practice: scientific communities and institutions should develop simple pilot reporting templates; journals should establish clear pilot reporting guidelines for authors; and authors should acknowledge when pilot studies were run and share relevant materials and data. Current reporting standards favor conclusive and polished research narratives, but such narratives are the end product of a complex scientific process that should also be shared. A truly transparent science should spotlight processes, like piloting, that underpin the conclusions scientific papers draw.

Keywords: piloting, pilot study, pretest, preliminary study, feasibility study, reporting, open science, transparency

Introduction

Transparency is a core value of scientific research. It can increase the credibility of research by facilitating evaluation, reproduction, and replication of findings, while simultaneously allowing other researchers to learn from, reuse, and improve the products of research (Klein et al., 2018; Miguel et al., 2014; Munafò et al., 2017). Given these benefits, the open science movement has advocated for more accessible reporting and transparent information transfer to accelerate scientific progress (e.g., Nosek et al., 2022). The replication crisis, often referred to as part of the "credibility revolution," has further underscored the importance of maintaining and sharing detailed records of the methodological and statistical decisions made throughout the research process (Vazire, 2018). However, while transparency is encouraged for reporting final results, little attention is given to the complex, iterative nature of the research process, particularly regarding the insights gained from piloting.

Piloting—the practice of conducting preliminary investigations before a full-scale research study to ensure that the study works as expected[1]—is often a necessary part of the research process. Works as expected being the operative phrase: while this framing may suggest a singular goal, preliminary studies serve a variety of functions in practice, such as exploring and refining novel methodologies, assessing feasibility, verifying underlying assumptions, or identifying areas for improvement before committing to full-scale studies. Researchers also frequently use pilot studies to determine a method's efficacy as a way to justify and strengthen research claims (e.g., for stage 1 Registered Reports or applications for research funding) or to estimate effect size calculations for the final study[2] (though, for cautions regarding these uses,

[1] Terminology for this practice varies; other common terms for piloting include "preliminary study," "pre-test," "feasibility study," and "exploratory study." For a more comprehensive list, see Pilot Reporting Task Force (2024).
[2] Throughout this paper, "final study" refers to the study the pilot is intended to inform or support.

see Albers & Lakens, 2018 and National Center for Complementary and Integrative Health, n.d.).

While there may be disagreements about the boundaries of 'piloting,' this paper does not aim to standardize what should or should not count as piloting. Instead, we argue that the ambiguities surrounding piloting as a practice underscore the need for greater transparency. Regardless of how piloting is defined, pilot studies serve as a critical foundation for many research endeavors. Yet, in the psychological sciences, there are no common standards for conducting or reporting these preliminary studies. Since methodological choices informed by piloting can influence research outcomes, and pilot data is rarely or inconsistently reported, there is a clear need for greater transparency in piloting practices and more consistent reporting mechanisms.

This paper presents a case for improving the transparency of pilot studies. To build this case, we begin by outlining the benefits of piloting. We then discuss several potential issues and missed opportunities for evaluation and innovation that arise when pilot study information is not shared. Finally, we provide near-term recommendations for how researchers can build up these practices while the field develops guidelines and resources for making pilot studies more transparent.

The Importance of Pilot Studies

Piloting is invaluable to researchers. While a study's objectives, protocols, and equipment may appear sound, many issues only become apparent once data are collected. Without the preliminary checks that piloting provides, researchers risk uncovering problems after significant investment of time, money, and participant effort. To that end, piloting serves at least three core purposes. It allows researchers to ensure ethical oversight, test whether a study is feasible, and

confirm that a study is operating as intended. Below, we elaborate on these purposes to highlight the importance of pilot studies (see Table 1 for further examples); however, an exhaustive list or evaluation of piloting practices is beyond the scope of this paper.

Although researchers work hard to ensure ethical treatment of participants before running studies, it is impossible to anticipate every ethical issue that could arise during a study. Running pilot studies, even for studies that have already received ethics approval, helps researchers identify unintended behaviors and outcomes early, such as potential privacy breaches, unexpected participant distress, or issues related to cultural sensitivity. Detecting ethical issues at an early stage allows researchers to adjust procedures and improve participants' experience before exposing a full sample of participants to potential risks.

For many studies, it may be unclear whether the procedure or primary objective is feasible. For example, some studies may depend on participant behavior, data collection pipelines, or organizational partnerships that are difficult to control or predict. Will rats spontaneously engage with the stimulus designed for the study? Is this wearable capable of accurately tracking participants' sleep? Can this organization faithfully deliver the intervention to its employees? By testing for and resolving such feasibility issues before full-scale data collection, researchers can prevent wasting resources on infeasible study protocols. This makes pilot studies especially valuable for resource-intensive projects, such as field experiments, clinical studies, longitudinal research, and large-scale collaborations.

Because piloting often involves a full trial run of a study, it also allows researchers to assess whether each component of their study is operating as intended. For instance, piloting can guide adjustments to study instructions, length, and other features that affect participant comprehension and engagement. It can also uncover issues with measurement tools, such as

limited response variability (e.g., ceiling or floor effects), problematic trial order, or items that participants misinterpret. In some cases, piloting may reveal deeper concerns, such as experimental confounds or unequal attrition between conditions, which could compromise causal inferences. By identifying these issues early, piloting helps researchers ensure the validity and reliability of the final study.

Running pilots studies can also have several benefits beyond directly addressing ethical, feasibility, or operational issues. First, piloting can lead to novel insights about the research question or study design. For example, pilot participants might provide comments that cause researchers to include new measures or conduct different analyses. Additionally, pilot studies can serve as a training ground for researchers. Piloting offers a lower-stakes environment for researchers to familiarize themselves with complex protocols, unfamiliar methodologies, and new technologies. Finally, evidence of successful and ethical pilot studies is sometimes required by outside parties such as funders (e.g., National Institutes of Health, 2025), ethics review boards (e.g., Oates et al., 2021), and journals evaluating registered reports (e.g., Springer Nature Limited, n.d).

In summary, pilot studies are an invaluable tool for psychologists because pilots can stress-test every aspect of a study, from study design and recruitment to data collection and analysis (see Table 1 for examples; see also Hessels et al., 2025 for detailed examples in the context of eye-tracking). The versatility and broad applicability of piloting establish it as a central component of many psychologists' research pipelines. Given the importance and prevalence of piloting, we argue that information from these studies should be reported—because if piloting is what makes methods work, then not reporting those practices creates a critical gap in the research record.

Table 1. Example uses of pilot studies

Study component informed by pilot	Examples of specific elements pilots can assess	Purpose achieved by piloting		
		Is study ethical?	Is study feasible?	Does study operate as intended?
Study design	Did participants comprehend the instructions?			X
	Did unforeseen risk or discomfort arise for participants?	X		
	How long did the procedure take?		X	X
	How strong was the experimental manipulation?		X	X
	How were responses distributed on key variables?		X	X
	Did participants guess the purpose of the study?			X
	Were the materials appropriate for the cultural context?	X		X
	Are the measures valid?			X
Participant recruitment	Can sufficient data be collected?		X	
	What was the attrition rate?		X	X
	Why did participants drop out?	X	X	X
	What were the sample demographics?	X	X	X
Data collection	Were researchers sufficiently trained to carry out the protocol?			X
	Was randomization implemented correctly?			X
	Did digital elements display correctly?			X
	Were tools (e.g., MRI; eye-tracker) appropriately calibrated?			X
	Was the data stored securely?	X		X
Data analysis	How are the data structured?			X
	Do analysis scripts run properly?			X

	Do researchers have sufficient expertise to analyze the data?		X	
	Are the planned analyses appropriate given the observed variability in the data?		X	
	Can the data be sufficiently de-identified?	X		

Why Pilot Studies Should be Reported

Despite their value, pilot studies have long been described as "underdiscussed, underused, and underreported" (Prescott & Soeken, 1989, p. 60). Even when pilots are mentioned, researchers often supply vague references to having learned from the pilot study without specifying what was learned or how the study protocol was adjusted in light of the pilot (van Teijlingen & Hundley, 2002). Since pilot studies play a key role in refining methods and guiding decision-making, failing to report them transparently can result in knowledge loss and cause other researchers to repeat the same mistakes. In the following subsections, we outline key arguments for the transparent reporting of pilot studies (see Table 2 for a summary).

Table 2. Summary of arguments for why pilot studies should be reported

Reason to report pilots	**Brief explanation**
Avoid file-drawering parts of the research process	Unreported pilots fail to contribute to the overall body of knowledge on a particular topic or methodology. Additionally, if pilots are not reported by default, pilots with desirable results may be more likely to be shared.
Reveal selection biases that harm generalizability	During piloting, various features of the study design (measures, stimuli, manipulations, samples, etc.) are tested and selected. Study design features that lead to large/novel effects may be over-selected for during piloting, which can harm the generalizability of final study results. Reporting pilots can reveal whether such a selection bias occurs.

Facilitate research evaluation	Information about how study designs were refined, assumptions were tested, and logistical problems were addressed during piloting helps readers evaluate whether a study provides an informative test of a hypothesis.
Help other researchers	Knowing what was tried, what worked, what failed, and why design decisions were made can help other researchers avoid similar failures, understand scientific rationales for certain study design practices, and reuse/extend study designs.
Uphold ethical responsibilities for research reporting	Reporting pilot studies helps researchers meet their obligations to share findings with the public, reduce research waste, and demonstrate adherence to ethical guidelines for working with human and animal subjects.
Contribute to the improvement of pilot studies	Piloting is a core part of the research process, but methods for piloting can only be improved if the field can assess current practices and functions of piloting.

Reporting Pilots Avoids File-Drawering Parts of the Research Process

Only the end-product of research, rather than the research process itself, is typically shared with outside observers. Published journal articles often omit details about how study designs, procedures, or analyses were developed in favor of more concise reporting of final study findings. This lack of transparency of the research process is similar in nature to publication bias, or the tendency for certain types of studies to be more likely to be published than others (Ioannidis et al., 2014; Pennington, 2023).

Rosenthal (1979) described the “file-drawer problem” as a type of selective reporting where studies with “favorable” results are published while others are not. He argued that this could cause the published literature to reflect only a small fraction of studies actually conducted. If a substantial proportion of pilot studies remain unreported, regardless of whether the subsequent final study is published, this constitutes unreported research activities that fail to

contribute to the overall body of knowledge on a particular topic or methodology. In this way, the tendency to not report pilot studies can be thought of as a specific type of file-drawer problem.

Similar to the uncertainty surrounding the publication rates of final studies (Ensinck & Lakens, 2025; Lishner, 2022; Simonsohn et al., 2014), there is no definitive data on how often published final studies have unreported pilot studies. It could be that most published studies involved some form of piloting but do not report it, but it is also possible that many final studies involved no piloting. A recent survey of psychology researchers ($N = 135$), conducted by several of the authors of this paper, found that a majority of respondents said they pilot their studies at least 80% of the time (Pilot Reporting Task Force, 2024). However, researchers already invested in piloting practices may have been overrepresented in the study. Ultimately, current reporting practices hinder a clear assessment of the true prevalence of piloting.

For final studies, the file-drawer problem has been said to lead to an incomplete and biased knowledge base about a given topic. There can be systematic reasons for why some studies are file-drawered (Franco et al., 2014). If a study shows non-significant or unclear results, researchers or journals may be less willing to publish them. Further, researchers may be more likely to try to publish studies whose results support the researchers' prior work and hypotheses, or the majority view in the field. Such processes can lead to an underrepresentation of negative findings and an overestimation of effect-sizes in the published literature. This, in turn, can hinder the ability of meta-analyses to summarize the research in a given field.

The reporting of pilot studies may face pressures comparable to those causing the file-drawer problem. For example, pilot studies that support the conclusions of the final studies in the paper may be more likely to be reported. Making the reporting of pilot studies a norm in the field can help reduce this instantiation of the broader file-drawer problem. Furthermore, it

would clarify whether the reported studies are the first or n^{th} iteration of a particular investigation, providing a more complete and accurate picture of the research.

Reporting Pilots Can Reveal Selection Biases that Harm Generalizability

During the piloting process, researchers often test out different measures, manipulations, tasks, stimuli, samples, control conditions, experimental settings, and so forth to settle on a combination of study design parameters to best assess their question of interest. Many of these study design parameters can affect the results of the study (Almaatouq et al., 2024; Clark, 1973; Henrich et al., 2010; Yarkoni, 2022). For example, if a study were testing how watching short-form videos affected people's mood, the sample of videos used in the experiment could dramatically affect the results (e.g., videos of funny animals versus videos of political violence).

As is the case across most scientific disciplines, psychologists are incentivized to publish statistically significant, novel results (Higginson & Munafò, 2016). Such a reward structure is likely to bias the process by which study design parameters are selected (Bear & Phillips, 2022; Fiedler, 2011). For example, in the hypothetical study mentioned above, short-form videos that lead to large, novel, or hypothesis-consistent effects may be more likely to be selected to use in the final study. To be clear, this need not be due to questionable research practices or self-serving motivations on behalf of the researcher. Because psychologists are trying to uncover valuable knowledge, they will be more likely to pursue effects that look strong or novel. If a psychologist tests a combination of study design parameters that just so happens to result in a large effect in their pilot, they might be more likely to scale up that combination of parameters into a final study; if, instead, that combination of study design parameters just so happens *not* to result in a clear effect, the psychologist might try to test their research question differently.

A selection process biasing study design parameters toward desirable effects could populate a scientific literature with findings that do not generalize when other combinations of study design parameters are employed. Although the original findings may indeed be confirmed in direct replications, they might be skewed towards a particular end of the distribution of possible effect sizes for a given research question. For example, imagine that stimuli for the hypothetical study on short-form videos were chosen through a piloting process that selected for strong effects. These selected stimuli would likely generate a larger effect than if a truly random sample of stimuli were tested. As such, it is possible that some approaches to piloting could cause the literature to contain many effects that are unrepresentative of the effect that would be expected if randomly selected combinations of study design parameters were tested.

Currently, the process by which study design parameters are selected during piloting remains opaque, as details about preliminary tests are rarely shared in the literature. Reporting pilot studies can provide insight into which combinations of study design parameters were tested before the final study and why certain parameters (e.g., stimuli, measures, manipulations) were selected. Without access to this information, it is difficult to know, for any given study or line of work, whether a selection bias is occurring. More generally, this information could be important for predicting the generalizability of final study results and could help researchers document under what circumstances a given effect holds.

Reporting Pilots Facilitates Research Evaluation

Scientific tests should be severe—that is, they should be capable of revealing when a hypothesis is incorrect (Claesen et al., 2022; Hofler et al., 2025; Lakatos, 1978; Mayo, 2018; Popper, 2005). The severity of a test reflects the likelihood that supportive evidence will emerge only if the hypothesis is true, rather than due to flaws in study design or measurement (Mayo,

2018; Meehl, 1990). If a hypothesis passes a test that is likely to detect its failure, researchers can be more confident that the result is informative and not driven by trivial or misleading factors.

Pilot studies can help researchers ensure that a study is set up to adequately test its hypotheses. Before collecting confirmatory data, pilots can check whether measures, procedures, and assumptions are operating as intended (as discussed earlier). Even the most robust statistical analysis cannot compensate for problems like poor measurement, confusing instructions, or overlooked confounds (Meehl, 1978; Mayo, 2018). Without pilot reporting, it is often unclear whether (or which) mitigation strategies were used to assess the severity of a given method. Making the piloting process transparent allows others to see how the study was refined to eliminate avoidable errors and build a stronger foundation for the final test.

For example, imagine a series of pilot studies were run to assess the clarity and reliability of a novel self-report measure. An initial pilot identified ambiguous item wording, but, after researchers revised the phrasing, subsequent pilots revealed that the revised item better mapped onto the construct of interest. Reporting such pilot findings would help demonstrate that the final measure aligns more closely with the intended construct (Flake, Pek, & Hehman, 2017; Flake & Fried, 2020). Generally, reporting piloting insights like these helps readers judge whether the final study's design truly tests the hypothesis under severe conditions.

To be clear, we are not arguing that piloting should be used to ensure a method produces a specific effect. Rather, we are arguing that reporting the piloting process helps the scientific community evaluate whether researchers have ensured that an effect is not attributable to design flaws or confounds. When it is clear which decisions were shaped by piloting, readers can make more informed judgments of a study's rationale and execution, preventing unwarranted pessimism or optimism regarding its potential contributions.

Reporting Pilots Helps Other Researchers

Science is a fundamentally collective exercise (Nature, 2021), where the successes and failures of individual researchers serve as valuable information for the wider community. Because individual researchers build on each other's work, consistent and transparent reporting of pilots is essential for creating a more balanced record of the research process.

At the most basic level, it is useful for the broader scientific community to know what was tried, what worked, and what failed. Sharing this information can accelerate progress by sparing others from repeating the same mistakes and guiding them toward effective approaches from the outset. Such insights may be especially valuable for research groups who lack the funds or personnel for extensive piloting (Puthillam et al., 2024).

Beyond documenting *what* was done, detailed accounts of the reasons behind piloting-based decisions can help others understand *why* certain things were done. Without knowing what other researchers learned during the piloting process, one is left to wonder whether a particular feature of the design was pivotal or merely an arbitrary choice. For example, a researcher might select one personality scale over another because they found in piloting that it resulted in fewer missing responses and less confusion over item wording. Without transparent reporting of such pilot findings, readers cannot distinguish this informed decision from a case where no alternatives were considered and a scale was chosen merely out of convenience or tradition. In both cases, the final paper would simply report which scale was used, despite fundamentally different decision-making processes. Consequently, a potential knock-on effect of *not* reporting decisions based on piloting is that arbitrary conventions (i.e., researchers using a particular approach simply because a previous researcher used that approach) may persevere in place of scientifically well-justified ones.

Sharing learnings from pilot studies can also be particularly useful for researchers seeking to replicate or extend existing work. Even direct replications that strive to re-run a study as faithfully as possible to the original study will inevitably alter some features of the study (e.g., the time of day at which data was collected; the university name listed in the consent form; the physical demeanor of a confederate). Results from piloting could reveal features that matter for the efficacy of an experimental manipulation. Replicators can then ensure they do not inadvertently alter any features that happen to be important. Similarly, those seeking to extend existing findings by, say, identifying moderators or boundary conditions might find pilot results useful for selecting possible variables to manipulate.

Reporting Pilots Upholds Ethical Responsibilities for Research Reporting

Most scientists rely on public funding, either directly through government grants or indirectly through institutions that depend on government support and tax incentives to maintain research infrastructure. Given this reliance, many scientists feel they have an ethical obligation to share their findings with the public, which serves as a driving ethos behind open access movements (e.g., Kimbrough & Gasaway, 2020; Suber, 2003). Arguably, this obligation should hold equally for pilot studies as for final studies.

Reporting pilot studies helps researchers fulfill their responsibility as stewards of public funding by reducing research waste. As discussed earlier, sharing insights from pilot studies helps others refine their methods and avoid similar failures. This transparency, in turn, helps conserve valuable resources and ensures public investments are used effectively. Furthermore, making piloting insights available for other researchers respects participants' contributions by maximizing the scientific value derived from their involvement.

In areas where research poses inherent risks to human or animal participants, reducing unnecessary failures is also a matter of harm reduction. Every failed or redundant study exposes participants to risks that might have been avoided had prior piloting work been shared. This applies even to low-risk studies, where sub-optimal research resulting from missing pilot reporting can waste participants' time and patience and carries an opportunity cost. Therefore, transparent pilot reporting can reduce the cumulative risk, broadly defined, faced by research participants across studies.

Furthermore, there is an ethical imperative to maintain scrutiny over how research is conducted. Just as with final studies, researchers must ensure that pilot study participants are not exposed to unnecessary risks, that potential benefits are clearly communicated, and that informed consent is properly obtained (Khan et al., 2021; Thabane et al., 2010). Poorly documented or unpublished pilot studies risk evading this ethical oversight (Sim, 2021), as the field can only ensure ethical guidelines are followed for studies that are reported.

Reporting Pilots Contributes to the Improvement of Pilot Studies

Psychology can only improve piloting practices if pilot studies are transparently reported. As with final studies, sharing information about pilot studies allows researchers to distinguish between the norms and functions of conducting them. For instance, it helps clarify whether certain practices are followed out of tradition or because they are specifically tailored to the needs of a given study. Moreover, there are likely more effective and less effective piloting practices, which may vary across different study types (Hessels et al., 2025). However, it is difficult to identify and improve those practices if they are not made public. Sharing piloting techniques therefore allows researchers to benefit from other scientists' piloting innovations.

Approaches to piloting likely differ between subfields. For example, some clinical subfields with hard-to-reach samples prone to attrition might have different piloting needs from some cognitive subfields where research can be conducted with student samples but where technical checks are needed to guarantee proper functioning of equipment. However, this diversity of approaches remains largely invisible to researchers. Without systematic reporting across subfields, not only is it difficult to learn potentially valuable approaches developed elsewhere (which is particularly important given the rise of interdisciplinary research), but also more visible subfields may inadvertently establish inappropriate standards for areas with fundamentally different needs. Transparent reporting of current piloting practices would therefore allow researchers to assess the status quo and discuss whether and how to improve piloting practices both within and across subfields.

Increased reporting of piloting could also raise awareness around the general value of piloting. The current lack of piloting transparency could mean that researchers' knowledge of piloting is subject to idiosyncratic experience in individual labs or research environments. As mentioned above, this can result in a limited or inconsistent understanding of why certain piloting methods are used or how they can be adapted and improved across subfields. However, it could also mean that many researchers, without such experiences, are unaware that piloting is an available practice or even that others are piloting their studies at all. If reporting piloting were normative, it might help researchers realize the importance of piloting and/or better understand how to apply it in their own work.

Challenges and opportunities for making transparent pilot reporting normative

Why pilots are rarely reported

Despite the many benefits that can come from reporting pilot studies, it is no surprise that pilots are rarely reported because the field does not provide guidance on whether or how to report them. The APA publication manual (7th edition) contains a single mention of 'pilot' or 'piloting' in its 446 pages, and offers no instructions about whether or how to report pilots (American Psychological Association, 2019, p. 100). Since many curricula on empirical psychological research are based on the APA publication manual, piloting practices may be overlooked in the formal education of psychologists. Further, the vast majority of psychology journals' author submission guidelines say nothing about pilot studies (although a few mention pilot studies in the context of registered report-style articles; see PLOS, n.d.; Springer Nature Limited, n.d.). There are also almost no articles in psychology dedicated to discussing piloting and whether/how it should be reported (cf. Hessels et al., 2025). Given how central piloting is to many psychology researchers' workflows, this is a critical oversight.

Beyond the lack of explicit guidance, authors may have various concerns about reporting pilot studies. The previously mentioned survey on psychologists' piloting practices asked respondents, "What challenges could researchers face when including information about pilot studies in publications?" (Pilot Reporting Task Force, 2024, p. 18). Some general themes from researchers' responses included fearing negative reactions about the quality of their studies, adhering to journal word count constraints, a dearth of reporting guidelines, having insufficient documentation of their own pilots, and the additional time reporting could take.

Despite the barriers they anticipated, a majority of the respondents believed that many details about pilot studies would be helpful to include in publications. These details included an acknowledgment that pilots were run, sample sizes of pilots, procedures of pilots, and descriptions and explanations of any differences between pilots and final studies (Pilot Reporting Task Force, 2024). Some of the most common reasons respondents thought this information should be reported were that it could improve the assessment of final studies, meet ethical considerations around transparency, and advance knowledge and research efficiency. These survey responses suggest that, even without explicit exposure to the arguments made earlier in this paper, many psychologists already support the idea of reporting details from pilot studies. This raises the question: how can our field encourage a behavior that many researchers support in the abstract, but for which they hold practical concerns?

How to promote pilot reporting

The tension between theoretical support and practical concern is not unique to pilot reporting. It echoes other initiatives, such as data sharing and preregistration, where community norms had to shift in tandem with infrastructure to foster a more transparent and open research culture (see Nosek, 2019, for further discussion). Our primary aim in this article is to articulate the value of reporting pilot studies. However, we recognize that this is distinct from addressing the practical challenges of how such reporting should be done. While piloting information may be important for the proper evaluation of science, the mechanism for disseminating this information requires investment.

We have mentioned that current scientific publishing focuses on the product of the research, which tends to reward conclusive and polished narratives. By contrast, piloting is often

dynamic, iterative, and exploratory. Thus, pilot reporting may demand new systems better suited to capturing this complexity (especially within traditional article formats).

This also raises the question of what constitutes "reporting." Although our arguments about the benefits of pilot reporting may allude to a certain depth of reporting, we do not intend to put the cart before the horse and preach standard quality for a mechanism that does not even exist. Rather than stipulating prescriptive expectations, we call for a forward momentum that recognizes that defining what should be reported is inherently intertwined with figuring out how it can be done. Still, we understand it would be unsatisfying to raise these issues without offering any actionable steps in the interim. Therefore, we offer three preliminary recommendations for supporting more widespread and meaningful reporting of pilot studies, even as broader systems for doing so continue to develop.

The first suggestion is for the field to develop user-friendly templates for reporting information from pilot studies (whether as standalone reports or more concise companions for manuscript appendices). The Pilot Reporting Taskforce (https://pilotreportingtf.github.io/) is currently developing such templates. However, it is important that other researchers also innovate in this space given the variability of practices across subfields and methodologies.

The second suggestion is that journals outline expectations for pilot study reporting in their author submission guidelines. As Collins (2025) argued, decades of reporting guideline development in biomedicine was not enough to ensure such guidelines were used. Without workflow integration and aligned incentives from journals, even well-tested reporting checklists remain under-utilized. Having clear guidance from journals can also reassure authors that reporting piloting information will not be penalized and help authors decide what to include within the journal's constraints.

The third suggestion is for authors to start consistently reporting bare-bones information about their pilot studies. Inspired by the "21 word solution" to disclosing details about data collection and analysis (Simmons et al., 2012), we suggest that authors include a pilot transparency statement that specifies which details about pilot studies will be shared. We recommend that authors acknowledge all pilot studies (or note their absence), briefly describe any revisions to the study that piloting informed, and publicly share the study materials (e.g., instructions, procedure notes, stimuli) and de-identified data (when possible given ethical considerations) from their pilots. A hypothetical example of this reporting approach is illustrated in Box 1.

Box 1. Hypothetical example of how to employ a pilot transparency statement and share corresponding details from pilot studies

Throughout this paper, we report any pilot studies or explicitly note their absence, describe any pilot-informed revisions, and provide corresponding materials and data.

Study 1

Pilot studies

Study 1 was informed by three pilot studies. Pilot Study A (N=12) assessed whether participants understood the study instructions. Based on participants' comments and performance on comprehension checks, we revised the instructions to include an example. Pilot Study B (N=32) assessed the suitability of our stimuli. We selected 16 vignettes (of the 24 tested) to use in the final study based on the following three criteria: (1) no participant indicated that the vignette was hard to understand; (2) no participant was familiar with the event described in the vignette; (3) the vignette received an average disgust rating of at least 4 (on a 7-pt scale). We excluded our primary dependent variable from this pilot so as to avoid biasing our stimuli selection process. Pilot Study C (N=5) was a final check to ensure our study operated as intended; no revisions were made to the study based on

Pilot Study C. The materials and de-identified data for these three pilots are available in our paper's OSF repository at [*LINK*].

While this level of detail may fall short of what some consider ideal, we think it is a good first step for several reasons. First, it is simple. Acknowledging any pilots that were run, describing how pilots informed the final study, and sharing materials and de-identified data from pilots often requires minimal effort. Second, this practice is unlikely to be objectionable to journal editors and reviewers since it occupies very little manuscript space, does not detract from the manuscript's final studies, and does not challenge journal reporting norms. Finally, acknowledging when pilots were run gives reviewers the opportunity to request more information if they think those pilots would be valuable to report in more detail. To be clear, we consider this approach a stopgap while journals and other institutions develop policies and infrastructure to make pilot reporting easy. The field should begin debating the ideal forms of pilot reporting, but bare-bones reporting will progress those initiatives in the meantime.

Conclusion

Transparent science should share more than final outcomes; it should also share the steps that were integral to arriving at them. One often overlooked step is piloting, an indispensable tool for psychological research that supports nearly every stage of the research process. Many research outcomes are shaped by decisions made during piloting. Yet, pilot studies are not systematically reported, or even acknowledged, in the psychology literature.

Despite the benefits of reporting pilot studies, it comes as little surprise that this practice is rare. Psychologists are not incentivized to report pilots, norms around reporting pilots do not

exist, and guidelines or tools for reporting pilots are lacking. Researchers, online repositories, journals, and funders should work to make reporting pilots possible and easy (Nosek, 2019). While these systems are still developing, researchers can begin to bridge this gap by adopting small, actionable practices that make piloting more visible, such as including brief pilot transparency statements in manuscripts.

Finally, although it is beyond the scope of this paper, the field must also engage in broader discussions about the general practice of piloting, and other preliminary research practices, including potential pitfalls and best practices among subfields (e.g., Hessels et al., 2025). Given how common piloting is in psychology, strikingly little has been written about the practice in the psychology literature.

Science is an iterative process that continually develops over time, and it is crucial that this process is made transparent for other scientists to learn from, critique, and repurpose. Reporting pilot studies can bring our field one step closer to that ideal.

Acknowledgements

We thank Shannon Barrios, Rachel Hayes-Harb, Jade Pickering, and the other members of the Pilot Reporting Task Force for their valuable contributions to the early conceptualization of this work. We would also like to thank the attendees at our Society for the Improvement of Psychological Science (SIPS) unconference and hackathon sessions for their continued interest and support. Finally, we thank the Berkeley Initiative for Transparency in the Social Sciences (BITSS) for a Catalyst Grant that supported the early stages of this project.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2024). Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences. *Behavioral and Brain Sciences*, *47*, e33. https://doi.org/10.1017/S0140525X22002874

American Psychological Association (2019). *Publication manual of the American Psychological Association, seventh edition*.

Bear, A., & Phillips, J. (2022). Random effects won't solve the problem of generalizability. *Behavioral and Brain Sciences*, *45*, e3. https://doi.org/10.1017/S0140525X2100011X

Claesen, A., Lakens, D., Vanpaemel, W., & van Dongen, N. (2022). *Severity and crises in science: Are we getting it right when we're right and wrong when we're wrong?* PsyArXiv. https://doi.org/10.31234/osf.io/ekhc8

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Collins, G. S. (2025). Innovative solutions are needed to overcome implementation barriers to using reporting guidelines. *BMJ*, *389*, r718. https://doi.org/10.1136/bmj.r718

Ensinck, E. N. F., & Lakens, D. (2025). An inception-cohort study quantifying how many registered studies are publicly shared. *Advances in Methods and Practices in*

Psychological Science, *8*(1), 25152459241296031. https://doi.org/10.1177/25152459241296031

Fiedler, K. (2011). Voodoo correlations are everywhere—Not only in neuroscience. *Perspectives on Psychological Science*, *6*(2), 163–171. https://doi.org/10.1177/1745691611400237

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research. *Social Psychological and Personality Science*, *8*(4). https://doi.org/10.1177/1948550617693063

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., Holleman, G. A., & Hooge, I. T. C. (2025). The fundamentals of eye tracking part 5: The importance of piloting. *Behavior Research Methods*, *57*(8), 216. https://doi.org/10.3758/s13428-025-02737-9

Higginson, A. D., & Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology*, *14*(11), e2000995. https://doi.org/10.1371/journal.pbio.2000995

Höfler, M., Kräplin, A., Elsherif, M. M., Schepke, M., Montefinese, M., Seetahul, Y., Sætrevik, B., Peikert, A., Varga, M.A., & Wallrich, L. (2025). Evaluate what is claimed to be confirmed: Initial version of a falsification assessment form (FAF). *PsyArXiv*. https://doi.org/10.31234/osf.io/eck8y_v1

Ioannidis, J. P. A., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences*, *18*(5), 235–241. https://doi.org/10.1016/j.tics.2014.02.010

Khan, M. I. U., Mbuagbaw, L., Holek, M., Bdair, F., Durrani, Z. H., Mellor, K., Eddy, S., Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Lancaster, G. A., & Thabane, L. (2021). Transparency of informed consent in pilot and feasibility studies is inadequate: A single-center quality assurance study. *Pilot and Feasibility Studies*, *7*(1), 96. https://doi.org/10.1186/s40814-021-00828-w

Kimbrough, J., & Gasaway, L. (2020). Publication of government-funded research, open access, and the public interest. *Vanderbilt Journal of Entertainment and Technology Law*, *18*(2), 267–302.

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1), 20. https://doi.org/10.1525/collabra.158

Lakatos, I. (1978). *The methodology of scientific research programmes* (J. Worrall & G. Currie, Eds.; Vol. 1). Cambridge University Press. https://doi.org/10.1017/CBO9780511621123

Lishner, D. A. (2022). Sorting the file drawer: A typology for describing unpublished studies. *Perspectives on Psychological Science*, *17*(1), 252–269. https://doi.org/10.1177/1745691620979831

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. https://doi.org/10.1037/0022-006X.46.4.806

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, *343*(6166), 30–31. https://doi.org/10.1126/science.1245317

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

National Center for Complementary and Integrative Health (n.d.). *Pilot studies: Common uses and misuses*. Retrieved August 4, 2025, from https://www.nccih.nih.gov/grants/pilot-studies-common-uses-and-misuses

National Institutes of Health (2025). *Research instructions for NIH and other PHS agencies*. U.S. Department of Health and Human Services. https://grants.nih.gov/grants/how-to-apply-application-guide/forms-i/research-forms-i.pdf

Nature (2021). Research collaborations bring big rewards: The world needs more. *Nature*, *594*(7863), 301–302. https://doi.org/10.1038/d41586-021-01581-z

Nosek, B. (2019). *Strategy for culture change*. Center for Open Science. https://www.cos.io/blog/strategy-for-culture-change

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Oates, J., Carpenter, D., Fisher, M., Goodson, S., Hannah, B., Kwiatkowski, R., Prutton, K., Reeves, D., & Wainwright, T. (2021). *BPS code of human research ethics*. British Psychological Society. 10.53841/bpsrep.2021.inf180

Pennington, C. (2023). *A student's guide to open science: Using the replication crisis to reform psychology*. McGraw-Hill Education (UK).

Pilot Reporting Task Force. (2024). *Piloting practices across psychological sub-disciplines*. OSF. https://doi.org/10.17605/OSF.IO/3QDY2

PLOS (n.d.). *Submission guidelines*. PLOS One. Retrieved August 4, 2025, from https://journals.plos.org/plosone/s/submission-guidelines

Popper, K. (2005). *The logic of scientific discovery* (2nd ed.). Routledge. https://doi.org/10.4324/9780203994627

Prescott, P. A., & Soeken, K. L. (1989). The potential uses of pilot work. *Nursing Research*, *38*(1). https://doi.org/10.1097/00006199-198901000-00015

Puthillam, A., Montilla Doble, L. J., Delos Santos, J. J. I., Elsherif, M. M., Steltenpohl, C. N., Moreau, D., Pownall, M., Silverstein, P., Anand-Vembar, S., & Kapoor, H. (2024). Guidelines to improve internationalization in the psychological sciences. *Social and Personality Psychology Compass*, *18*(1), e12847. https://doi.org/10.1111/spc3.12847

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

Sim, J. (2021). Distinctive aspects of consent in pilot and feasibility studies. *Journal of Evaluation in Clinical Practice*, *27*(3), 657–664. https://doi.org/10.1111/jep.13556

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 word solution*. Social Science Research Network. https://doi.org/10.2139/ssrn.2160588

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Springer Nature Limited (n.d.). *Submission guidelines—Registered reports*. Nature Human Behaviour. Retrieved June 7, 2025, from https://www.nature.com/nathumbehav/submission-guidelines/registeredreports

Suber, P. (2003). The taxpayer argument for open access. *SPARC Open Access Newsletter*. http://dash.harvard.edu/handle/1/4725013

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

van Teijlingen, E. R., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, *16*(40), 33–36. https://doi.org/10.7748/ns2002.06.16.40.33.c3214

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685