The focus on the lack of replication in psychology has raised concerns regarding the reproducibility and validity of current published reports (Etz & Vandekerckhove, 2016; Lindsay, 2015; Open Science Collaboration, 2015; van Elk et al., 2015). Researchers have attributed this lack of replication to many things, including publication bias (Brannick, 2012), the potential for the over-reliance, abuse, or misunderstanding of p-values (Gigerenzer, 2004; Ioannidis, 2005; Simmons, Nelson, & Simonsohn, 2011), and researcher degreed of freedom (Simmons, Nelson, & Simonsohn, 2011), including selective removal or inclusion of datapoints.

In the years since the crisis began many recommendations have been made (Cumming, 2008; Simmons, Nelson, & Simonsohn, 2011; Asendorpf et al., 2013; Lakens, 2013; Benjamin et al., 2017; Lakens et al., 2018; Valentine, Buchanan, Scofield & Beauchamp, 2018; Maxwell, Lau, & Howard, 2015; Nosek, Spies, & Motyl, 2012), projects have been organized (Klein et al., 2014; Open Science Collaboration, 2015) and steps have been taken to assist in improving the reliability of psychological findings. While we have seen vast discussion of the problems and proposed solutions, research has yet to determine how this new culture of research has altered the specifics of reporting practices. Herein, we aim specifically to quantify the rates of reporting of outliers within psychology at two time points—in 2012, when the replication crisis was born (Pashler & Wagenmakers, 2012), and 5 years later, in 2017.

Outliers

Since Bernoulli first mentioned outliers in 1777, outliers, influential observations, or fringliers have been defined and redefined throughout the years (Beckman & Cook, 1983; Hodge & Austin, 2004; Munoz-Garcia, Moreno-Rebollo, Pascual-Acosta, 1990; Orr, Sackett, & Dubois, 1991; Osborne & Overbay, 2004), but for the purposes of this manuscript can be defined as, “an observation which being atypical and/or erroneous deviates decidedly from the general behavior of experimental data with respect to the criteria which is to be analyzed on it” (Munoz-Garcia, Moreno-Rebollo, Pascual-Acosta , 1990). And while there are a plethora of graphical and statistical options available to researchers to identify and describe these datapoints (i.e. visual depictions of data and residuals, the 3 standard deviation rule, Mahalanobis or Cook’s distance, leverage, etc.; Tabachnick & Fidell, 2007), this discussion of outliers rarely makes it to the page. Consider, for example, Orr, Sackett, and Dubois (1991), who inspected 100 Industrial/Organizational Psychology personnel studies and found no mention of outliers.

This is not to say that researchers are not taking these datapoints into consideration during their data cleaning and analysis plan, or even going so far as removing them, as work by LeBel and colleagues (2013) has reported that 11% of psychology researchers stated that they had not reported excluding participants for being outliers in their papers and additional work has suggested that more than a quarter decide whether to exclude data after looking at the impact of doing so (Fiedler & Schwarz, 2016). This, by definition, is a problem. The current zeitgeist of increased transparency and reproducibility applies not only to the manner in which data is collected, but also the various ways the data is transformed, cleaned, pared down, and analyzed. Therefore, it is just as important for a researcher to state how they identified outliers within their data, how the outliers were handled, and how this choice of handling impacted the estimates and conclusions of their analyses, as it is for them to report their sample size.

Reporting out outlier practices are vital because by keeping outliers in a dataset, analyses are more likely to have increased error variance (depending on sample size, Orr et al., 1991) and biased estimates (Osborne, & Overbay, 2004) as well as reduced effect size and power (Orr, Sackett, & Dubois, 1991; Osborne, & Overbay 2004) which can alter the results of the analysis and lead to falsely supporting (Type I error), denying a claim (Type II error), or failing to replicate previous work. Additionally, incorrect estimates of effect lead to misleading meta-analyses or sample size estimates for study planning. Beyond these effects on analyses and conclusions investigation and exploration of these outliers can be beneficial to scientists as a way to further improve our research and research practices. Further, outliers can be informative to researchers and to their research models as they can encourage the diagnosis, change, and evolution of a research model (Beckman & Cook, 1983). Taken together, these issues caused by not noting outliers can lead to furthering unwarranted avenues of research, ignoring important information, and creating erroneous theories, all of which serve to weaken the sciences, while making clear the presence or absence of outliers, how they were assessed, and how they were handled, can improve our transparency and replicability, help to strengthen our science.

Some may argue that use of the precious word limit dictated by journals to describe such acts as outlier identification and handling may be irrational. However, we contest that given the current availability and use of online supplements and appendixes, as well as the invent of the Open Science Framework (OSF; cite) which allows researchers the ability to upload any number of additional resources and supplements that can be easily referenced in their manuscripts in as little as 4 words: See supplement at osf.io/52mqw.

The social sciences have begun to take a careful look at the way we process and interpret data, as many famous experiments do not appear to replicate (Open Science Collaboration, 2015). The Open Science Foundation (OSF) was founded in 2013 to promote a transparent research process from formation of the hypotheses to completely reproducible papers (Nosek et al., 2015). This project examines the impact of the formation of OSF and changing research culture on the publication of information concerning data screening methods for outliers, as the impact of outliers can critically change the findings and interpretation of experiments.

In the wake of the reproducibility crisis and publications such as (many labs, osf’s rep paper), it is worthwhile to investigate changes that may have taken place.

Improvements that have come about in our new research culture:

Recommendations on how to be more transparent, journals and granting agencies are asking more about openness and transparency, online repositories like GitHub, OSF, and OpenDOAR allow for full disclosure of research material, data, and code. Registered replications including detailed data analysis plans are encouraged by many journals.

Within this manuscript we sought to identify the impact of this new research culture on the publication of information concerning data screening methods for outliers.

We chose to focus on outliers because they have been shown to increase error variance and bias estimates, reduce effect size and power, and ultimately change the conclusions of an analysis. Any one of these outcomes can lead to a failure to replicate. For example, if one researcher identifies a number of outliers within their dataset, but decides to allow the datapoints to remain in the analysis, this researcher may likely come to different conclusions than another individual who identified the same outliers, but chose to remove them.

Previous work has suggested that researchers have an ethical obligation to report removal of outliers (Rosenthal, 1994), but how often is this actually occurring? Previous work in blah has suggested reporting as low as 8%, and additional work has noted that 11% of psychology researchers surveyed stated that they had not reported excluding participants for being outliers in their papers (LeBel et al., 2013). However, we are unaware of any other research that has attempted to quantify the rates of reporting of outliers across psychology to date. Herein, we aim to do just this, across a variety of sub disciplines in psychology, and at two time points—before the beginning of the replication crisis and 5 years later—to identify not only how often we have been reporting outliers, but also how that rate of reporting may have changed with the new research culture introduced by the replicability crisis.

Can point to Pashler, Harold; Wagenmakers, Eric Jan (2012). "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?". *Perspectives on Psychological Science*. **7** (6): 528–530. [doi](https://en.wikipedia.org/wiki/Digital_object_identifier):[10.1177/1745691612465253](https://doi.org/10.1177%2F1745691612465253). [PMID](https://en.wikipedia.org/wiki/PubMed_Identifier) [26168108](https://www.ncbi.nlm.nih.gov/pubmed/26168108).

As a timestamp for the start of the replicability crisis.