Promoting Transparent Reporting of Conflicts of Interests

and Statistical Analyses at the *Journal of Sex Research*

John K. Sakaluk, Statistical Consultant, *Journal of Sex Research*Department of Psychology

University of Victoria

Cynthia A. Graham, Editor-in-Chief, *Journal of Sex Research*
Department of Psychology

University of Southampton

 The publication process for scientific communication is iterative. Researchers first attempt to convince editors and reviewers of claims made about the variables under study; editors and reviewers then evaluate the reasonableness of these claims in light of the strength of the supporting evidence. Editors (with input from reviewers) then either decide that researchers’ claims: (1) are acceptable given the evidence, (2) require more evidence to evaluate; or (3) are unacceptable given the evidence. Ideally—and providing other aspects of the manuscript are satisfactory—manuscripts with strong supporting evidence should be published, those with good supporting evidence should be encouraged to improve and then reconsidered, and those with poor supporting evidence should be rejected. The integrity of this process, however, relies on authors transparently reporting aspects of their research process that ought to influence (for better, or for worse) how reviewers and editors would evaluate the evidential value of their article. When important details are omitted from article submissions, reviewers and editors will be less likely to make judicious appraisals of manuscripts. Transparent reporting by authors therefore helps to ensure a healthy, high-quality published literature of scientific findings.

 The *Journal of Sex Research* is updating its submission requirements to promote more transparent reporting in two areas of disclosure: *Conflicts of Interest* (COIs), and *Reporting of Statistical Analyses*. Our intent with this editorial is to outline the nature of these changes, and to articulate why we feel they are necessary to implement at this time.

**Promoting Transparent Reporting of COIs**

Potential conflicts of interest have been defined as “when an investigator’s relationship to an organization affects, or gives the appearance of affecting, his/her objectivity in the conduct of scholarly or scientific research” (Office of Research Integrity (ORI), 2017). While the most obvious examples of conflict of interest are financial in nature, for example, owning stock in pharmaceutical companies, conflicts of interest can occur as a result of many other less-recognized sources, such as personal relationships and academic competition (International Committee of Medical Journal Editors, 2010).

Though a conflict of interest does not guarantee that research findings would have been different than those obtained without such a conflict, the problem is that there is the *potential* for this to be the case (DeAngelis, 2000). It is also important to note that some conflicts of interest may be unavoidable and may not necessarily be “unethical” (ORI, 2017). Transparency in reporting possible conflicts of interest is nonetheless essential for at least two reasons. First, researchers themselves are not always best placed to recognize whether a conflict of interest exists in relation to their research. Second, it has been argued that disclosure of possible conflicts of interest are essential to maintain public trust in research (Johns, Barnes, & Florencio, 2003).

The advice from professional societies, organisations, and increasingly, from journals is for authors to disclose any possible conflicts of interest. While at *JSR* we have asked authors to disclose information about any funding they received for research and had a generic question about “conflicts of interest,” to date we have not provided any guidance to authors about what constitutes conflict of interest. We have remedied this now, with additional guidance on conflicts of interest and a request during the online submission process to provide details of any perceived conflicts of interest.

**Promoting Transparent Reporting of Statistical Analyses**

 Sexual science might be thought of as lying, in part, at the intersection of psychological and medical science. And for the last five years or so, both fields have been forced to grapple with the uncomfortable recognition that their research findings are not replicable (Open Science Collaboration, 2015; Nosek & Errington, 2017), populated by unreliable significant effects emerging at a rate that is too good to be true (Fanelli, 2010; Ioannidis, 2005).

There are many reasons why the replicability of a given body of research might leave something to be desired, ranging from systemic to individualized factors (Nosek & Bar-Anan, 2012; Nosek, Spies, & Motyl, 2012), and the use of devious to seemingly innocuous practices. For example, making simple reporting mistakes (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2016), running chronically underpowered studies (Maxwell, 2004; Schimmack, 2012), refusing to let others verify one’s findings (Wicherts, Bakker, & Molenaar, 2011; Wicherts, Borsboom, Kats, & Molenaar, 2006), or engaging in flagrant data fabrication (Fanelli, 2009; Stapel, 2014) are all very different behaviors morally, but they share a practical consequence in rendering the scientific record less trustworthy.

But while psychological (Nelson, Simmons, & Simonsohn, in press; Spellman, 2015) and medical scientists (Baker, 2016; Begley & Ellis, 2012) have debated—often heatedly—the extent of the replicability problem (e.g., Open Science Collaboration, 2015, *vs.* Gilbert, King, Pettigrew, & Wilson, 2016, *vs.* Etz & Vandekerckhove, 2016) and the merits of each solution *du jour* (e.g., Finkel, Eastwick, & Reis, 2017, *vs.* LeBel, Berger, Campbell, & Loving, in press; Simons, 2014, *vs.* Crandall & Sherman, 2016), sexual scientists have been notably silent on these matters (see Sakaluk, 2016a). However, for reasons that are not yet apparent, the past year or two seems to mark the end of this trend. In the last year, sexuality conferences have featured multiple replicability-related talks (e.g., Lalumière, 2016; Sakaluk, 2017); sexuality journals have begun piloting interventions to increase replicability (e.g., the *Canadian Journal of Human Sexuality* has been evaluating the use of *statcheck* in the peer-review process, Epskamp & Nuijten, 2016); and sexual scientists have conducted and published direct replications challenging classic findings (e.g., Balzarini, Dobson, Chin, & Campbell, 2016). The most recent example of the increased attention being paid to replicability in sexual science is the new statistical disclosure policies at the journal *Sexual Abuse*, which encourages authors to disclose practices “that can increase the risk of spurious significant findings”.[[1]](#footnote-1) *JSR* can now be counted among these recent efforts aspiring to increase transparency and replicability of sexual science findings.

 *JSR* is changing its submission requirements, *requiring* two new elements of statistical reporting, and *encouraging* two others, for all papers submitted using inferential statistical tests.[[2]](#footnote-2) The first new requirement—requiring authors to disclose all elements of analytic flexibility—is functionally new, and designed to curb increasing rates of Type-I (i.e., false-positive) errors of inference. The remaining three submission guidelines (one of which is also required) are likely practices that many already use, and are explicit guidelines of the 6th Edition of APA style (American Psychological Association, 2009), which the journal already adopts; we are now simply formalizing them. We describe each new reporting policy below, including a rationale for why *JSR* is adopting them.

**Required Reporting of Analytic Flexibility** One of the observations made during the earliest days of the so-called replicability crisis, is that social scientists regularly (ab)use flexibility when collecting, analyzing, and reporting on their data (John, Loewenstein, & Prelec, 2012). Some examples include collecting more data (58%) or stopping data collection early (22.5%) after researchers have seen whether a key effect is significant, and deciding to exclude cases after looking at the impact on key *p*-values (43.4%), and/or simply reporting only those studies that “worked” (i.e., yielded the desired significant effects; 50%). Critical observers have dubbed these practices, and others: *questionable research practices* (or “*QRPs.”* John et al., 2012), *“p-hacking”* or *researcher degrees of freedom* (Simmons, Nelson, & Simonsohn, 2012), and *the garden of forking paths* (Gelman & Loken, 2013).[[3]](#footnote-3)

A pivotal simulation paper by Simmons, Nelson, and Simonsohn (2012) makes clear the consequences of capitalizing on analytic flexibility: researchers inadvertently trick themselves into thinking they are conducting null-hypothesis significance tests of their effects at an acceptable Type-I error rate (i.e., ⍺ = .05), when in fact, these rates are *dramatically* increased (as high as ⍺ = .60, in some cases); *p*-values are effectively meaningless when *post hoc* analytic decisions are made (Dienes, 2008).What this means, in practicalities, is that a significant effect produced via (ab)use of *researcher degrees of freedom* should not be taken as seriously as a significant effect that was observed using principled data collection, analysis, and reporting practices that were decided *a priori*.[[4]](#footnote-4)

Though there may be good reasons, *a priori*, for collecting more data, reporting on some analyses and not others, and/or excluding some participants, once a researcher has seen how these decisions affect *p*-values, it is impossible distinguish between principled decision-making, and motivated reasoning and mindless data torturing. Researchers should therefore strive to avoid making these kinds of analytic decisions *post hoc* altogether. Conducting replications (direct and conceptual) and/or making use of pre-registration within the same multi-study paper are other recently discussed strategies that researchers can use to reduce the possibility of *p*-hacking (or the appearance of *p*-hacking, see Nosek, Ebersole, DeHaven, & Mellor, 2017; Sakaluk, 2016b).

In order to make it clearer to *JSR* editors, reviewers, and readers, when sample size, analytic, and reporting decisions have been made *a priori* vs. *post hoc*, we now require submitting authors to report for all studies, in the methods section(s): (1) how they determined sample size; (2) all data exclusions (if any); (3) all manipulations (if any); and (4) all measures used (see Simmons, Nelson, & Simonsohn, 2012). Submitting authors will also confirm that they have reported this information when submitting their papers in ScholarOne, the online submission portal.

Sample size determination may be new ground for some sexual scientists. Open-source resources for conducting *a priori* power analyses,[[5]](#footnote-5) which inform researchers how many participants they need to detect an effect of an expected size in a particular statistical design are becoming more widely available and accessible (e.g., G\*Power, Faul, Erdfelder, Lang, & Buchner, 2007, PANGEA, Westfall, 2016; and the bias-correcting power-analysis apps supplied by Maxwell, Delaney, & Kelley, in press; see also Judd, West, & Kenny, 2017). Researchers struggling to determine an appropriate effect size to use in these calculations might calibrate their power analyses to the average effect size in their area of interest (e.g., *r* = .21 in social psychology, Richard, Bond, & Stokes-Zoota, 2003), or simply use the smallest effect size that they would deem worthy of their attention, effort, and resources (Lakens, 2014).[[6]](#footnote-6)

Otherwise, meeting our new statistical disclosure requirement at *JSR* should be relatively straightforward: if you determined your sample size ahead of time, included all data, manipulations, and measures, succinctly say so in the manuscript, and choose the appropriate checklist selection at submission. And if you did not determine your sample size ahead of time, and/or included all data, manipulations, and measures, say so, and why. Disclosure of this kind is crucial for readers to be able to assign an appropriate level of confidence (or skepticism) to the effects reported in a paper.

**Required Main-Text Reporting of Inferential Tests in Full APA Format** *JSR* also now requires authors to report the results of inferential statistical tests in the main body of the paper using *full* APA 6th Edition format (American Psychological Association 2009, e.g., *F* (2, 132) = 4.47, *p* = .01, *t* (198) = 2.04, *p* = .04, χ2 (3) = 3.35, *p* = .34, *z* = 2.77, *p* = .01). Complete reporting of this sort is necessary in order for readers to ensure that authors have correctly reported test statistic values, degrees of freedom, and *p*-values (these values must “add up”). Studies using human eye-ball coding (Bakker & Wicherts, 2011) and automated computer-checks (Nuijten et al., 2016) suggest that statistical reporting errors (including those that affect declarations of statistical significance) are disturbingly common. Though *JSR* is not at this time considering implementing any sort of formal check of accurate statistical reporting (e.g., Sakaluk, Williams, & Biernat, 2014), we think it important to ensure that the statistical reporting in *JSR* articles can be verified in this way by interested parties. For the meantime, researchers wishing to verify the accuracy of their own statistical reporting can do so easily using online apps like *p-checker* (<http://shinyapps.org/apps/p-checker/>, Schönbrodt, 2015) and *statcheck* (<http://statcheck.io>, Epskamp, Nuijten & Rife, 2016).

**Encouraged Reporting of Confidence Intervals and Effect Sizes**

 *JSR* strongly encourages researchers to include accompanying confidence intervals and effect sizes for all statistical tests. Confidence intervals are intervals for a point-estimate (e.g., a mean, mean-difference, correlation coefficient, regression slope, or odds-ratio) that will, in the long run, contain the population parameter for the point-estimate a certain percentage of the time (e.g., 95% of 95% confidence intervals, with infinite random sampling of a population, will contain the population parameter).[[7]](#footnote-7) Effect sizes, alternatively, provide a sense of the magnitude of a given effect (independent of sample size or other influences of estimation precision), often in a standardized metric (e.g., zero-order correlation, standardized mean difference, odds-ratio).

Though they convey different information, confidence intervals and effect sizes both help to provide more context and meaning to a statistical test, transcending what is conveyed by *p*-values alone.[[8]](#footnote-8) A statistical test might be significant, for example, but an accompanying 95% confidence interval might be so broad, and nearly encompassing the null value (e.g., 95% CI: .001, 2.45, for a mean-difference on a 5-point rating scale), suggesting that the effect under consideration was not examined with a sufficient level of precision (Cumming, 2014). Or alternatively, an effect might be statistically reliable, but small (e.g., *d* = 0.08); a reader would then need to think critically about whether such a small effect was theoretically or applicably meaningful, or trivial (e.g., Hyde, 2005; Prentice & Miller, 1992).

Meeting the encouraged standard of reporting confidence intervals and effect size measures should, thankfully, be relatively straightforward. Confidence intervals for estimates of effects can often be produced using statistical software with the simple click of an option box, or the inclusion of a handful of characters of code. Producing measures of effect size can sometimes be a bit more onerous. Whereas effect size measures for associations (e.g., *r*, *R*2, β, and ORs) are typically outputted automatically (or near-automatically) by software, effect size measures for group comparisons are either inconsistently available in statistical software (e.g., Cohen’s *d*), or when available, are less intuitive (e.g., *η2*, and *ηp2*). Lakens (2013) published resources designed to help researchers better understand and easily compute effect size measures for statistical analyses of group comparisons (i.e., *t*-tests and *ANOVA*s).

**Encouraged Reporting of Group Sizes, Means, Standard Deviations, and Zero-Order Correlations**

Finally, JSR encourages the reporting of means and standard deviations for all group comparisons (alongside sample sizes per group). In the event that group comparisons were not the focus of a study, JSR encourages the reporting of zero-order correlations for all focal study variables. This reporting standard, also encouraged by the APA (American Psychological Association, 2009), comes in light of the increasing recognition that a singleton study, no matter how skillfully conducted and analyzed, can never yield the sole and final word on a given effect. The generation of knowledge is instead gradual, produced through cumulative science consisting of multiple direct and conceptual replications, as well as extensions and boundary tests, of a given effect. And though there are a number of methods for reviewing related literature (Card, 2012), meta-analysis remains one of the most effective for evaluating the state of knowledge generated by a given area of research (Chan & Arvey, 2012), despite its limitations.[[9]](#footnote-9)

Results from articles published in *JSR* cannot contribute to meta-analyses (and therefore cumulative science) if meta-analytic researchers cannot easily identify and extract (or calculate, if necessary) effect sizes that fall within the scope of their syntheses. Though the new *JSR* submission policy of encouraging reports of effect sizes should help towards this aim, a given paper may describe data relevant to an ongoing meta-analytic review that was not the primary focus of the original paper; the needed effect size measure may therefore not be reported in the body of the main text. It is therefore imperative that the necessary data are available for aspiring meta-analysts, without requiring the ongoing assistance of the original authors who are, for the most part, unresponsive or uncooperative to such requests (Wicherts et al., 2006).

Researchers submitting manuscripts to *JSR* can report the necessary grouped descriptive statistics or zero-order correlations in a table, when appropriate, in light of the purpose of the manuscript and space is available. If there are space constraints in the main article, JSR now publishes online supplementary tables that are linked to an article. Researchers can also make use of the free services available for sharing research materials, like the Open Science Framework (http://osf.io), which could host supplemental tables that meta-analysts could easily access. Meeting this encouraged reporting standard will not only help those conducting meta-analyses, but the original researchers too; meta-analysts will inevitably cite the papers contributing estimates to their syntheses, which will increase the impact metrics of the original researchers.

**Conclusions**

Norms for conducting, interpreting, and reporting statistical results are changing in social and medical science fields, in light of widespread—and legitimate—concerns about the trustworthiness of empirical findings. The field of sexual science has largely avoided the debates surrounding these changes, but is starting to recognize and rise to the challenges of fostering a replicable scientific literature. Undisclosed conflicts of interest and capitalizing on flexibility of data analysis are two primary mechanisms through which less trustworthy effects can populate the sexual science literature. We therefore believe that the updated *JSR* policies promoting transparent disclosure of these features of research will serve as a small, but important step towards ensuring the long-term replicability of sexuality research published in *JSR*.

**References**

American Psychological Association (2009). *Publication manual of the American Psychological Association* (6th edition). Washington, DC.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Retrieved from <http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>.

Bakker, M., & Wicherts, J. M. (2011). The (mis) reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*, 666-678.

Balzarini, R. N., Dobson, K., Chin, K., & Campbell, L. (2017). Does exposure to erotica reduce attraction and love for romantic partners in men? Independent replications of Kenrick, Gutierres, and Goldberg (1989) study 2. *Journal of Experimental Social Psychology*, *70*, 191-197.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531-533.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*, 389.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017, September 26). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved from psyarxiv.com/9h3nu

Chan, M. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science*, *7*, 79-92.

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93-99.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7-29.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Hampshire, England: Palgrave Macmillan.

DeAngelis, C. D. (2000). Conflict of interest and the public trust. *JAMA*, *284*, 2237-2238.

Epskamp, S., & Nuijten, M. B. (2016). Statcheck: Extract statistics from articles and recompute p values. R package version 1.2.2. [https://CRAN.R-project.org/package=statcheck](https://CRAN.R-project.org/package%3Dstatcheck)

Epskamp, S., Nuijten, M. B., & Rife, S. (2016). *statcheck on the web*. http://statcheck.io

Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLoS ONE*, *11*(2), e0149794.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS ONE*, *4*(5), e5738.

Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PloS ONE*, *5*(4, e10068.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (in press). Replicability and other features of a high-quality science: Toward a balanced and empirical approach. *Journal of Personality and Social Psychology*.

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

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, *351*(6277), 1037-1037.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581.

International Committee of Medical Journal Editors (2010). Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. *Journal of Pharmacology & Pharmacotherapeutics*, *1*, 42-58.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524-532.

Johns, M. E., Barnes, M., & Florencio, P. S. (2003). Restoring balance to industry-academia relationships in an era of institutional financial conflicts of interest: Promoting research while maintaining trust. *JAMA*, *289*, 741-746.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lakens, D. (2014). Performing high‐powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*, 701-710.

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, *4*, 24.

Lalumière, M. (September, 2016). *Why I no longer care about p values, or the notion of “statistical significance”: A user’s perspective*. Paper presented at the Canadian Sex Research Forum, Quebec City, QC.

LeBel, E. P., Berger, D., Campbell, L., & Loving, T. J. (in press). Falsifiability is not optional. *Journal of Personality in Social Psychology*.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*, 147-163. [http://dx.doi.org/10.1037/1082-989X.9.2.147](http://psycnet.apa.org/doi/10.1037/1082-989X.9.2.147)

Maxwell, S. E., Delaney, H. D., & Kelley, K. (in press). *Designing experiments and analyzing data: A model comparison perspective.* New York, NY: Routledge. Apps available at https://designingexperiments.com/shiny-r-web-apps/

Nelson, L. D., Simmons, J., & Simonsohn, U. (in press). Psychology's renaissance. *Annual Review of Psychology*.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*, 217-243.

Nosek, B. A., & Errington, T. M. (2017). Reproducibility in cancer biology: Making sense of replications. *Elife*, *6*, e23383.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2017, August 24). The preregistration revolution. Retrieved from osf.io/2dxu5

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

Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*, 1205-1226

Office of Research Integrity U.S. Department of Health and Human Services. *A brief overview on conflict of interests*. Available at: https://ori.hhs.gov/plagiarism-35. Accessed 25 September, 2017.

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

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160-164.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of sSocial psychology quantitatively described. *Review of General Psychology, 7*, 331-363. [http://dx.doi.org/10.1037/1089-2680.7.4.331](http://psycnet.apa.org/doi/10.1037/1089-2680.7.4.331)

Sakaluk, J. K. (January, 2017). *What the replication crisis means for sexual science—and why sexual scientists should care.* Paper presented at the Sexuality Preconference of the Society for Personality and Social Psychology, San Antonio, Texas.

Sakaluk, J. K. (2016a). Promoting replicable sexual science: A methodological review and call for metascience. *The Canadian Journal of Human Sexuality*, *25*, 1-8.

Sakaluk, J. K. (2016b). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology*, *66*, 47-54.

Sakaluk, J., Williams, A., & Biernat, M. (2014). Analytic review as a solution to the misreporting of statistical results in psychological science. *Perspectives on Psychological Science*, *9*, 652-660.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. Psychological Methods, 17, 551-566. [http://dx.doi.org/10.1037/a0029487](http://psycnet.apa.org/doi/10.1037/a0029487)

Schönbrodt, F. D. (2015). *p-checker: One-for-all p-value analyzer.* Retrieved from http://shinyapps.org/apps/p-checker/.

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *Dialogue, The Official Newsletter of the Society for Personality and Social Psychology*, *26*(2), 4-7.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76-80.

Spellman, B. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, *10*, 886-899.

Stapel, D. (2014). Faking Science: A True Story of Academic Fraud. *Translated by Nicholas JL Brown.[Jan 14, 2015]*.

Westfall, J. (2016). PANGEA: Power ANalysis for GEneral Anova designs. *Unpublished manuscript.* Available at http://jakewestfall.org/publications/pangea.pdf.

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS ONE*, *6*(11), e26828.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*, 726-728. [http://dx.doi.org/10.1037/0003-066X.61.7.726](http://psycnet.apa.org/doi/10.1037/0003-066X.61.7.726)

1. We credit Dr. Michael Seto sharing this development on Twitter, prompting discussions that have led to *JSR* changing its own submission guidelines. [↑](#footnote-ref-1)
2. These new policies are not relevant for papers based on qualitative research. [↑](#footnote-ref-2)
3. The first author took an informal survey of attendees during a presentation (Sakaluk, 2017) who reported ever having engaged in any of these practices during their research career; he estimates ≥75% of attendees raised their hand (including himself). [↑](#footnote-ref-3)
4. Readers can see for themselves how easy it is to *p*-hack a truly null effect to significance using the simulator at: http://shinyapps.org/apps/p-hacker/ [↑](#footnote-ref-4)
5. Note, *post hoc* power analyses based on observed effects (i.e., the kind that SPSS provide for you) are *not* informative for making sample size decisions because they are determined by your *p*-values. [↑](#footnote-ref-5)
6. Researchers interested in methods of “data peeking” that do not inflate Type-I error rates should consult the sequential analysis technique described in Lakens’s (2014) paper. [↑](#footnote-ref-6)
7. Confidence intervals are often misunderstood (see Belia, Fidler, Williams, & Cumming, 2005) as providing an interval estimate that the researcher can be certain, with a certain level of confidence (e.g., 95%), the true population value lies between. A visualization by Magnusson, <http://rpsychologist.com/d3/CI/>, helps to render the correct interpretation of confidence intervals clearer. [↑](#footnote-ref-7)
8. The APA describes relying on confidence intervals as, “…in general, the best reporting strategy,” and effect size measures as, “…almost always necessary to include” (p. 34). [↑](#footnote-ref-8)
9. Methodologists are beginning to notice that, somewhat ironically, meta-analyses are susceptible to the very same issues of replicability as individual studies (Lakens, Hilgard, & Staaks, 2016), alongside replicability-limiting issues that are unique to meta-analyses, which continue to go unsolved (e.g., accurately correcting for publication bias; see Carter, Schönbrodt, Hilgard, & Gervais, 2017). [↑](#footnote-ref-9)