



UvA-DARE (Digital Academic Repository)

Facing the Unknown Unknowns of Data Analysis

Wagenmakers, E.-J.; Sarafoglou, A.; Aczel, B.

DOI

[10.1177/09637214231168565](https://doi.org/10.1177/09637214231168565)

Publication date

2023

Document Version

Final published version

Published in

Current Directions in Psychological Science

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Wagenmakers, E.-J., Sarafoglou, A., & Aczel, B. (2023). Facing the Unknown Unknowns of Data Analysis. *Current Directions in Psychological Science*, 32(5), 362-368. <https://doi.org/10.1177/09637214231168565>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Facing the Unknown Unknowns of Data Analysis

Eric-Jan Wagenmakers¹, Alexandra Sarafoglou¹,
and Balazs Aczel²

¹Department of Psychology, University of Amsterdam, and ²Institute of Psychology, ELTE Eötvös Loránd University

Current Directions in Psychological Science
2023, Vol. 32(5) 362–368
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09637214231168565
www.psychologicalscience.org/CDPS



Abstract

Empirical claims are inevitably associated with uncertainty, and a major goal of data analysis is therefore to quantify that uncertainty. Recent work has revealed that most uncertainty may lie not in what is usually reported (e.g., *p* value, confidence interval, or Bayes factor) but in what is left unreported (e.g., how the experiment was designed, whether the conclusion is robust under plausible alternative analysis protocols, and how credible the authors believe their hypothesis to be). This suggests that the rigorous evaluation of an empirical claim involves an assessment of the entire empirical cycle and that scientific progress benefits from radical transparency in planning, data management, inference, and reporting. We summarize recent methodological developments in this area and conclude that the focus on a single statistical analysis is myopic. Sound statistical analysis is important, but social scientists may gain more insight by taking a broad view on uncertainty and by working to reduce the “unknown unknowns” that still plague reporting practice.

Keywords

planning, data management, statistical inference, modesty, transparency, open communication

Consider the following fictitious, worst-case sequence of events in an empirical research project: A small group of researchers sets out to study a hypothesis that appears consistent with a theory that the group had promoted in earlier work. An experimental design is selected and an initial sample size is chosen based on what seems to have worked well in the past. Once the data are collected, a single data analyst applies a single statistical technique in the hope to “reject the null hypothesis,” that is, to find a *p* value lower than .05. If the analyst fails, additional statistical techniques are applied, the sample size is increased, other aspects of the data are explored, or—when none of these methods yield the desired outcome—the result is not reported. If the analyst is successful, the results are submitted for publication, but the data and the analysis code are not shared or shared only “upon request.” Although the researchers do not explicitly quantify their confidence in the main claim, the contents of the article title, abstract, and conclusion leave little doubt that the result is expected to replicate and generalize. The only replications that are ever reported are by the same group of researchers who published the original

claim—remarkably, these replications are almost always successful.

If this worst-case scenario (described by one of the reviewers as “stone-aged”) is even remotely realistic, it is clear that the inevitable uncertainty surrounding an empirical claim cannot be gleaned from the usual statistical report that centers on the *p* value and the confidence interval. First, the researchers are studying their pet hypothesis, and the analysis outcome has an impact on their careers. This creates a conflict of interest that threatens the unprejudiced spirit in which the hypothesis ought to be evaluated; this is analogous to taxpayers auditing their own returns. Psychologists should be especially aware of the myriad of subtle biases that affect behavior even in the absence of any conscious effort to deceive. Second, even if researchers were perfectly unbiased, they usually present the outcome of a single analysis to test a particular hypothesis; this does not address

Corresponding Author:

Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam
Email: EJ.Wagenmakers@gmail.com

the extent to which the outcome is fragile, that is, apparent only under a narrow set of analysis choices.

In recent years, it has become increasingly evident that there is more to data analysis and the assessment of uncertainty than the mere inspection of a confidence interval and that a truthful assessment of uncertainty requires a high degree of transparency throughout the entire research cycle. Consequently, new methods have been proposed to enhance transparency and methodological rigor, covering all stages of the empirical investigation, that is, planning an experimental study, data management, statistical inference, reporting, and publishing. In this article we provide a select overview of some of the most exciting changes to the status quo that have recently been proposed for each of these stages. Our focus is on experimental psychology, but we believe that methodological reform is also relevant for other subfields of psychology (e.g., Tackett et al., 2017 [clinical psychology]; Kapiszewski & Karcher, 2021 [qualitative psychology]; and Crüwell et al., 2019 [mathematical psychology]). We adhere to academic tradition and highlight our own contributions, if only because we know these relatively well.

Current Directions in Sample Size Planning, Restricting Analytic Freedom, and Data Management

The standard procedure in planning an experimental study concerns a statistical power analysis, which seeks to determine the sample size required to have a reasonable chance to reject the null hypothesis when it is false (e.g., when power is 80%, the Type II error rate—the probability of erroneously retaining the null hypothesis—equals 20%). A Bayesian version of power analysis is known as Bayes factor design analysis (e.g., Stefan et al., 2019), and it can be used to assess (a) the distribution of evidence for any fixed sample size and (b) the distribution of sample sizes for a fixed set of evidence thresholds in a sequential design.

The determination of sample size remains opaque, however, if readers are not informed about the reasoning behind the power calculations. The *SampleSizePlanner* application (Kovacs et al., 2022) has been developed for frequentist or Bayesian sample size determination, and it also shows what steps of the determination process need justification.

There is increasing recognition that planning an empirical project involves much more than sample size determination, and these considerations are relevant for experimental and nonexperimental studies alike. Specifically, advance measures can be taken to mitigate the deleterious impact of hindsight bias and confirmation bias. Essentially, these measures constitute a Ulysses contract: Anticipating the possibility of bias,

the researcher immunizes themselves by voluntarily restricting their freedom of analysis. For instance, the researcher may choose not to analyze the original data immediately but to first develop an analysis plan on an altered version of the data (e.g., the original data but with shuffled labels). After the analysis plan is complete, the data are “unblinded” and the plan is executed on the original data as specified (e.g., Dutilh et al., 2021; Sarafoglou et al., 2023). Similarly, researchers may adopt preregistration to outline their hypothesis, method, and analysis of the study before the data are known (e.g., Hardwicke & Wagenmakers, 2021). Registered Reports, a special case of preregistration, is a relatively new publication format where preregistration plans are peer reviewed and journals commit themselves to publishing the results independent of whether or not these turn out to be statistically significant (see Figure 1 in Chambers & Tzavella, 2022). Preregistration could serve a secondary purpose: It may force researchers to consider the methodological and practical details of their project more deeply (Sarafoglou et al., 2022).

Proper planning can prevent mistakes in the data management workflow as well. A recent survey found that for about 20% of psychology researchers, data management mistakes in the past 5 years had resulted in major adverse consequences (e.g., project failure, serious time and/or money loss, damaged professional reputation)—these mistakes were partly due to poor project preparation (Kovacs, Hoekstra, & Aczel, 2021). To prevent these mistakes, researchers should therefore carefully consider how data are collected, processed, documented, and stored.

Part of the planning phase also means formulating the research question precisely and incorporating prior literature, theory, or expert opinions into the study design (Borsboom et al., 2021; Kekecs et al., 2020). With regard to the development of an analysis plan, this may mean that such knowledge informs the specification of the model equation, the implementation of theory-based predictions as competing statistical models, and the development of appropriate parameter priors within the Bayesian framework.

A final innovation in planning is the advent of team science or “many labs” initiatives: By combining forces, a consortium of different labs is able to address an empirical question of interest conclusively while at the same time examining the extent to which the finding holds across different cultures and geographies (e.g., Forscher et al., 2022).

Current Directions in Statistical Inference

The promise of empirical science is to address substantive questions by taking a well-reasoned path from theory to design and from observation to conclusion.

The step from observation to conclusion is quantified by statistical inference, usually culminating in a p value or a confidence interval. This practice has come under increasing scrutiny. Some methodologists have argued that the threshold for new discoveries should be lowered from .05 to .005 (Benjamin et al., 2018), some have argued that hypothesis testing should be replaced by parameter estimation (e.g., Cumming, 2014), and others have argued that p values ought to be supplanted or supplemented with Bayes factors (e.g., Wagenmakers et al., 2016, 2018).

This discussion tacitly assumes that there exists only a single plausible analysis path (e.g., choice of statistical model, choice of preprocessing pipeline) or—if it is admitted that multiple plausible paths may exist—that all paths lead to the same conclusion. Several large-scale projects across different fields explored the analytic space either by inviting independent teams to analyze the same data set or by having the authors conduct multiverse analyses on the data themselves (Hoogeveen et al., 2022; Wagenmakers et al., 2022). These projects demonstrate that—at least in the social and behavioral sciences—the theoretical framework does not constrain the researchers to a limited analytic path. Indeed, only rarely did two teams use the exact same analysis. More worryingly, the different paths introduce substantial variability in the conclusions; sometimes even opposing conclusions were drawn from the same data (e.g., Botvinik-Nezer et al., 2020).

The results from recent “multianalyst” projects invite the bleak view of statistical inference displayed in Figure 1: Focusing exclusively on the finer details of the statistical modeling may be likened to rearranging the deck chairs on the *Titanic*. The primary component of uncertainty may reside not in what is reported (the “known unknowns”—the width of a confidence interval, the strength of a Bayes factor) but in what is *not* reported, either because it was unflattering or because it was never explored in the first place (Donald Rumsfeld’s “unknown unknowns”).

In order to obtain a more realistic impression of the uncertainty of an empirical claim, it is therefore important to assess the extent to which the conclusion is robust to plausible changes in the analytic path. Such changes can involve different preprocessing pipelines (e.g., Steegen et al., 2016), different dependent and independent variables, different models, and of course, also different outcome metrics (e.g., intervals, p values, and Bayes factors). Guidelines for conducting and reporting many-analysts studies are provided by Aczel, Szaszi, Nilsson, et al. (2021). In Wagenmakers et al. (2021), we outline seven general recommendations for more transparency in statistical inference that will result in a fairer acknowledgement of uncertainty: (a)

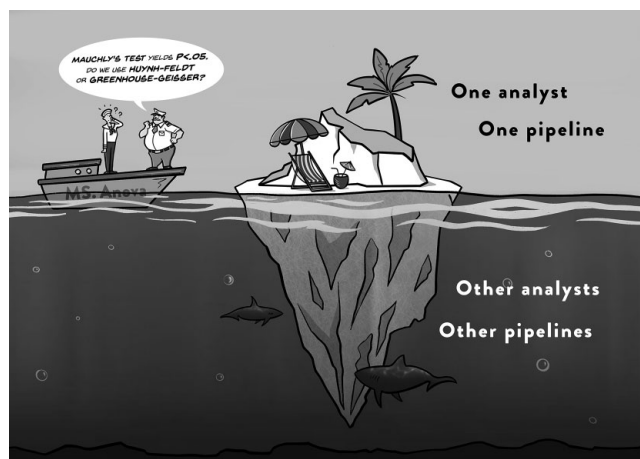


Fig. 1. The traditional data analysis framework wherein a single data analyst reports the outcome of a single statistical analysis procedure. The first problem with this framework is the risk of bias, that is, the danger that the analyst has (perhaps unwittingly) cherry-picked a procedure that yields a relatively flattering outcome. The second problem is overconfidence, that is, the fact that by ignoring alternative plausible analysis procedures, much of the uncertainty that accompanies an empirical claim remains hidden beneath the surface. See also Wagenmakers et al. (2022). Figure available under a CC-BY license from BayesianSpectacles.org. Design by Viktor Beekman; concept by Eric-Jan Wagenmakers.

visualizing data, (b) quantifying inferential uncertainty, (c) assessing data preprocessing choices, (d) reporting multiple models, (e) involving multiple analysts, (f) interpreting results modestly, and (g) sharing data and code. In that article, we provide explanations and guidance for each of these recommendations.

Current Directions in Transparency, Communication, and Publishing

The credibility of an empirical claim hinges on the transparency with which the results are reported. The practical challenge, however, is to know what and how to report. Some scientific subfields have already developed their own standards for reporting (e.g., START, PRISMA, and CONSORT; see <https://www.equator-network.org>). To assist researchers in the behavioral and social sciences, we recently developed a comprehensive transparency checklist to improve and document the transparency of their research (Aczel et al., 2020). The transparency checklist and accompanying online application allow researchers to obtain an overview of the transparency-related aspects of their work (e.g., pre-registration; data, code, and materials availability) and communicate these aspects to the journal at the time of submission.¹

Another challenge in the reporting stage is to avoid overconfidence when statistical outcomes are translated

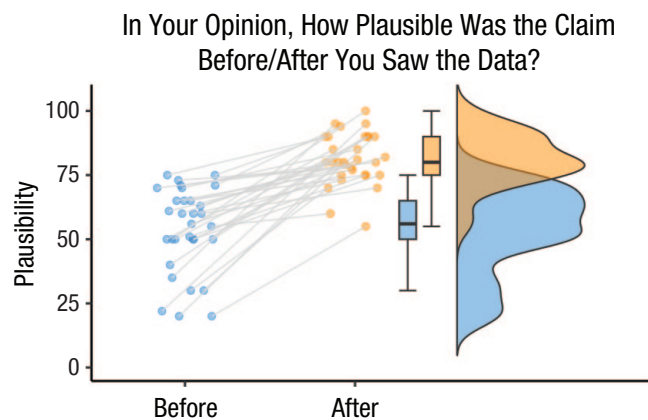


Fig. 2. Results from van Doorn and colleagues (2021). The gray lines connect the responses for each respondent. See text for details. Figure from preprint available at <https://psyarxiv.com/pc4ad>.

to substantive conclusions (e.g., Hoekstra & Vazire, 2021; Van Doorn et al., 2021). For instance, van Doorn et al. (2021) surveyed authors who had published a strong claim in *Nature Human Behaviour* (*NHB*). As shown in Figure 2, all 31 respondents indicated that the data reported in the *NHB* article made their claim more plausible than it was before. However, the size of the increase is relatively modest, and the final confidence was considerably lower than the article title suggests. The prior plausibility assessment centers around 50 (median = 56), and the posterior plausibility assessment centers around 75 (median = 80). It is not unlikely that the implicit demands of the traditional academic publishing system frustrate a modest interpretation of the results, as authors are generally expected to present compelling data that warrant strong claims.

A transparent report also contains the contributions of the authors. As the number of authors on journal articles increases, it becomes more and more difficult to know who did what. Tenzing (Holcombe et al., 2020; Kovacs, Holcombe, et al., 2021) is a tool that aims to assist researchers to document the roles each author played in the project. This online application makes it easy to create a list of authors, to report each author's contribution, and to provide funding information.²

The reform of scientific publishing concerns not just the authors' reporting practices but also the whole publication system. Current reform initiatives concern open review (in which the reviews are publicly accessible; Ross-Hellauer, 2017), double-blind review (in which the reviewers do not know the identity of the authors; Snodgrass, 2006), or noncommercial peer review that is independent from journals.³ In addition, there is increasing realization that reviewing is mostly done by researchers without any compensation or recognition.

Aczel, Szaszi, and Holcombe (2021) estimated that in 2020 alone, this service totaled more than 100 million hours, equivalent to 15,000 years. The fact that much published research is subsequently hidden behind a publisher's paywall suggests that the research community and all stakeholders of science should invest more effort into developing alternative models for the peer-review process and the entire publication system.

It is clear that science can progress even in the absence of publishers and academic journals. For instance, in data science and engineering, considerable prestige is associated with conference proceedings; in physics, preprints on ArXiv communicate key results well before they are ultimately published in a journal. It remains an open question how exactly such alternatives can be made normative and rewarding for researchers in other fields. However, we speculate that in order to change the status quo, two elements are critical: (a) The alternative to the current publishing system needs to be noncommercial, that is, open access for the reader and with near-zero article processing fees for the authors (e.g., <https://scipost.org/>); (b) funders need to demand that researchers publish only in these noncommercial outlets. In other words, funders need to prohibit publication in commercial outlets. Unless funders prohibit commercial publication, individual researchers will too easily succumb to the lure of the prestige associated with a high-impact journal, propagating a practice that in other walks of life would be considered racketeering.

Concluding Comments

Recent methodological work in psychological science has revealed that there is more to statistical uncertainty than meets the eye. Specifically, the uncertainty as routinely reported through confidence intervals, *p* values, and Bayes factors represents only the tip of an iceberg of uncertainty. In order to obtain a more realistic impression of the uncertainty that accompanies an empirical claim, it is necessary to know much more than the final outcome of a single statistical analysis. As Figure 3 illustrates, an ideal empirical report is freely available and features (a) assurance against cherry-picking (e.g., through preregistration or other Ulysses contracts), (b) publicly shared and properly anonymized data and code, (c) an assessment of inferential robustness (e.g., through a multiverse analysis or a multianalyst effort), (d) a modest interpretation of the outcome, and (e) a transparent description of the research.

A proper assessment of uncertainty therefore demands transparency across much of the empirical cycle, and this implies a radical departure from standard practice. Psychology has been at the forefront of these

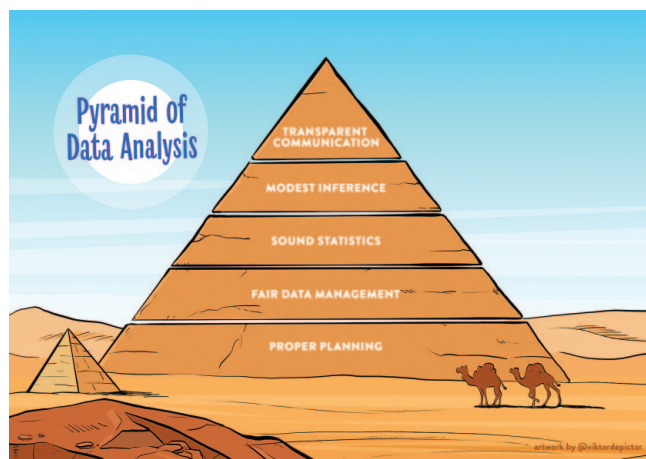


Fig. 3. A pyramid of data analysis. The figure highlights five stages of the empirical cycle (i.e., planning, data management, statistics, inference, communication) in which researchers can improve the assessment of uncertainty. The validity of higher stages relies on the validity of the underlying stages. Figure available under a CC-BY license from BayesianSpectacles.org. Design by Viktor Beekman; concept by Balazs Aczel, Alexandra Sarafoglou, and Eric-Jan Wagenmakers.

methodological developments, but for some other disciplines (e.g., medicine, neuroscience, biology, economics), it is mostly still business as usual. This is bound to change. Finally, these insights are also relevant to statistics proper. Especially for applied statisticians, it is important to realize they have a role to play from the conception of a project to its final report.

Recommended Reading

- Hoekstra, R., & Vazire, S. (2021). (See References). Proposes that researchers emphasize the statistical uncertainty of their results, stress the limitations of their studies, and provide alternative interpretations of their findings.
- Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *Royal Society Open Science*, 8(4), 201925. Summarizes common sources of uncertainty of analysis strategies and describes how different disciplines deal with them.
- MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, 526(7572), 187–189. Outlines the method of analysis blinding and shows that this Ulysses contract benefits even researchers who work in a theory-strong discipline, such as physics.
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, 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(1), 719–748. Discusses the importance of replication

studies and summarizes the recent advances on this topic in psychological science.

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, 526(7572), 189–191. <https://doi.org/10.1038/526189a>. One of the first articles to promote a multianalyst approach in psychological science.

Transparency

Action Editor: Teresa A. Treat

Editor: Robert L. Goldstone

Author Contributions

All authors were involved in every aspect of the creation of the manuscript.

Declaration of Conflicting Interests


The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported by a European Research Council grant to E.-J. Wagenmakers (283876).

ORCID iDs

Eric-Jan Wagenmakers  <https://orcid.org/0000-0003-1596-1034>

Alexandra Sarafoglou  <https://orcid.org/0000-0003-0031-685X>

Notes

1. See <https://martonbalazskovacs.shinyapps.io/SampleSizePlanner>.
2. See <https://tenzing.club>.
3. See, for example, <https://peercommunityin.org>.

References

- Aczel, B., Szaszi, B., & Holcombe, A. O. (2021). A billion-dollar donation: Estimating the cost of researchers' time spent on peer review. *Research Integrity and Peer Review*, 6, Article 14. <https://doi.org/10.1186/s41073-021-00118-2>
- Aczel, B., Szaszi, B., Nilsson, G., van den Akker, O. R., Albers, C. J., van Assen, M. A., Bastiaansen, J. A., Benjamin, D., Boehm, U., Botvinik-Nezer, R., Bringmann, L. F., Busch, N. A., Caruyer, E., Cataldo, A. M., Cowan, N., Delios, A., van Dongen, N. N. N., Donkin, C., van Doorn, J. B., . . . Wagenmakers, E.-J. (2021). Consensus-based guidance for conducting and reporting multi-analyst studies. *eLife*, 10, Article e72185. <https://doi.org/10.7554/eLife.72185>
- Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., . . . Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, 4(1), 4–6. <https://doi.org/10.1038/s41562-019-0772-6>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., Boeck, P. D., Dienes, Z.,

- Dreber, A., Easwaran, K., Efferson, C., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 16(4), 756–766.
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., . . . Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582, Article 7810. <https://doi.org/10.1038/s41586-020-2314-9>
- Chambers, C., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour*, 6(1), Article 1. <https://doi.org/10.1038/s41562-021-01193-7>
- Crüwell, S., Stefan, A. M., & Evans, N. J. (2019). Robust standards in cognitive science. *Computational Brain & Behavior*, 2(3), 255–265.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- Dutilh, G., Sarafoglou, A., & Wagenmakers, E.-J. (2021). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese*, 198(23), 5745–5772. <https://doi.org/10.1007/s11229-019-02456-7>
- Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2022). *The benefits, barriers, and risks of big-team science*. Advance online publication. <https://doi.org/10.1177/17456916221082970>
- Hardwicke, T. E., & Wagenmakers, E.-J. (2021). *Preregistration: A pragmatic tool to increase transparency, reduce bias, and calibrate confidence in scientific research*. MetaArXiv. <https://doi.org/10.31222/osf.io/d7bcu>
- Hoekstra, R., & Vazire, S. (2021). Aspiring to greater intellectual humility in science. *Nature Human Behaviour*, 5(12), 1602–1607.
- Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Documenting contributions to scholarly articles using CRediT and tenzing. *PLOS ONE*, 15(12), Article e0244611. <https://doi.org/10.1371/journal.pone.0244611>
- Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., Altay, S., Alzahawi, S., Amir, Y., Anthony, F.-V., Kwame Appiah, O., Atkinson, Q. D., Baimel, A., Balkaya-Ince, M., Balsamo, M., Banker, S., Bartoš, F., Becerra, M., Beffara, B., . . . Wagenmakers, E.-J. (2022). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior*. Advance online publication. <https://doi.org/10.1080/2153599X.2022.2070255>
- Kapiszewski, D., & Karcher, S. (2021). Transparency in practice in qualitative research. *PS: Political Science & Politics*, 54(2), 285–291.
- Kekecs, Z., Szaszi, B., & Aczel, B. (2020). *Expert consensus procedure (ECO): Facilitating robust scientific outputs*. PsyArXiv. <https://doi.org/10.31234/osf.io/9gqru>
- Kovacs, M., Hoekstra, R., & Aczel, B. (2021). The role of human fallibility in psychological research: A survey of mistakes in data management. *Advances in Methods and Practices in Psychological Science*, 4(4), 25152459211045930.
- Kovacs, M., Holcombe, A., Aust, F., & Aczel, B. (2021). Tenzing and the importance of tool development for research efficiency. *Information Services & Use*, 41(1–2), 123–130.
- Kovacs, M., van Ravenzwaaij, D., Hoekstra, R., & Aczel, B. (2022). SampleSizePlanner: A tool to estimate and justify sample size for two-group studies. *Advances in Methods and Practices in Psychological Science*, 5(1), 25152459211054060.
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>
- Sarafoglou, A., Hoogeveen, S., & Wagenmakers, E.-J. (2023). Comparing analysis blinding with preregistration in the many-analysts religion project. *Advances in Methods and Practices in Psychological Science*, 6(1). <https://doi.org/10.1177/25152459221128319>
- Sarafoglou, A., Kovacs, M., Bakos, B. E., Wagenmakers, E.-J., & Aczel, B. (2022). A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science*, 9(7), 211997.
- Snodgrass, R. (2006). Single- versus double-blind reviewing: An analysis of the literature. *ACM SIGMOD Record*, 35(3), 8–21. <https://doi.org/10.1145/1168092.1168094>
- Stegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, 51, 1042–1058. <https://doi.org/10.3758/s13428-018-01189-8>
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, 12(5), 742–756.
- Van Doorn, J., Van den Bergh, D., Dablander, F., van Dongen, N., van Derks, K., Evans, N. J., Gronau, Q. F., Haaf, J. M., Kunisato, Y., Ly, A., Marsman, M., Sarafoglou, A., Stefan, A., & Wagenmakers, E.-J. (2021). Strong public claims may not reflect researchers' private convictions. *Significance*, 18(1), 44–45. <https://doi.org/10.31234/osf.io/pc4ad>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . Morey, R. D. (2018). Bayesian inference for psychology: Part II. Example applications with JASP.

- Psychonomic Bulletin & Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25(3), 169–176. <https://doi.org/10.1177/0963721416643289>
- Wagenmakers, E.-J., Sarafoglou, A., Aarts, S., Albers, C., Algermissen, J., Bahnik van Dongen, Š. N., Hoekstra, R., Moreau, D., van Ravenzwaaij, D., Sluga, A., Stanke, F., Tendeiro, J., & Aczel, B. (2021). Seven steps toward more transparency in statistical practice. *Nature Human Behaviour*, 5(11), 1473–1480. <https://doi.org/10.1038/s41562-021-01211-8>
- Wagenmakers, E.-J., Sarafoglou, A., & Aczel, B. (2022). One statistical analysis must not rule them all. *Nature*, 605(7910), 423–425.