

## Correspondence

### ***The American Statistician* Special Issue on Statistical Inference<sup>1</sup>**

Caroline A. Watt      J. E. Kennedy

University of Edinburgh

To the Editor,

*The American Statistician* journal recently devoted a special issue to the topic of statistical inference. The special issue included a summary editorial (Wasserstein, Schirm, & Lazar, 2019) and 43 articles (401 pages) about proposed statistical methods that could modify, supplement, or replace the hypothesis testing methods that have dominated past research. The assumption for the articles was that the common past statistical practices are not acceptable. The articles were diverse, from a wide range of academic and applied disciplines with varying degrees of generality and mathematics.

The purposes of this letter are to point out some limitations of this special issue that are relevant for parapsychological and psychological research, and to clarify why the arguments in the special issue do not alter the requirements and expectations for study registrations submitted to the KPU Study Registry (2012), which we manage.

The article by Tong (2019) in the special issue aligns very closely with our views and with the requirements for the KPU Study Registry. Tong believes that exploratory research must be distinguished from confirmatory research and that the failure to make this distinction has been a major cause of past statistical problems. Efforts to address the statistical problems will not be successful until this distinction becomes a fundamental component of statistical thinking.

The statistical methods for exploratory and confirmatory research are different. Tong points out that  $p$ -values and hypothesis tests are applicable for preregistered confirmatory research, but not for exploratory research. He also notes that “most scientific research is (and should be) highly exploratory” (p. 252) and that attempts to adapt statistical practices to give strong inferences from exploratory research will not be successful. We agree with these points. Clearly defined confirmatory research is essential for strong statistical scientific evidence.

Unfortunately, the majority of the articles in the special issue do not address the distinction be-

---

<sup>1</sup> Send correspondence to: Caroline A. Watt, Ph. D., School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Scotland, [Caroline.Watt@ed.ac.uk](mailto:Caroline.Watt@ed.ac.uk)

tween exploration and confirmation. Many articles appear to have a “one size fits all” approach that implicitly emphasizes exploration and promotes a lack of distinction between exploration and confirmation. We expect that in time these proposals will be recognized as continuing the statistical weaknesses and abuses that have been common in the past.

In the editorial for the special issue Wasserstein, Schirm, and Lazar (2019) recommend elimination of the term “statistically significant” and associated dichotomous classification of statistical results—however, they note late in the editorial that  $p$ -value thresholds and decision criteria appear to be useful or needed in some situations of applied research and for confirmatory research. Our impression is that the editorial writers implicitly focused on the exploratory stage of research, with confirmation and application basically treated as afterthoughts. Given that formal confirmatory research historically has been rare outside of regulated clinical trials, it may be unsurprising that statistical discussions focus on the exploratory stage of research. However, as noted above, we believe the best solution to the past statistical abuses is to place greater emphasis on formal confirmatory methods, rather than discounting those methods in favor of methods suitable for exploratory research.

Emerging evidence indicates that the efforts to abandon hypothesis tests are not an effective approach for statistical analysis of confirmatory research. Fricker, Burke, Han, and Woodall (2019) evaluated 31 articles published in a journal that had banned the use of inferential statistics. They noted that the statistical methods allowed researchers to make claims for effects that were actually consistent with random fluctuations:

We found multiple instances of authors overstating conclusions beyond what the data would support if statistical significance had been considered. Readers would be largely unable to recognize this because the necessary information to do so was not readily available. (p. 374)

They also expressed the suspicion that in the cases when the conclusions were consistent with inferential statistical analyses, the researchers may have based the conclusions on hypothesis tests and then excluded the hypothesis tests from the final report.

Our expectation is that statistical errors and abuses will not be reduced until virtually all writings on statistical methods distinguish between exploratory and confirmatory analyses. The standards for evidence should be strengthened for confirmatory research and relaxed for exploratory research. The description of any newly proposed statistical method should clarify how the method applies to each of these stages of research. Unfortunately, the special issue gives the impression that this critical distinction has not yet been widely incorporated into statistical thinking outside of regulated clinical trials.

These principles are implemented on the KPU Study Registry with the requirement that all planned analyses be identified as exploratory or confirmatory. For confirmatory analyses, the planned statistical analysis must be fully pre-specified, including the numerical criteria that the experimenters will consider as acceptable evidence supporting the hypothesis of interest. For exploratory analyses, the planned statistical analysis need not be fully pre-specified or pre-specified at all. In fact, preregistration of exploratory research is considered optional, with the stipulation that any research that is not preregistered is presumed to be entirely exploratory.

We also believe that the methods for confirmatory research need to be strengthened to emphasize falsifiable research. The special issue does not discuss falsifiable research methods, and unfortunately these methods are not widely known among psychological researchers. Parapsychological and psychological studies are rarely designed to be falsifiable. We have described elsewhere the methods for falsifiable confirmatory research (Kennedy & Watt, 2018). The steps include specification of a minimum effect of interest, selecting a sample size that provides power substantially higher than the common .8, and study preregistration with specific numerical criteria for evidence that the hypothesis of interest is false as well as true. Examples of studies that do have falsifiable hypothesis tests include the prospective meta-analysis by Watt (2017) and the Transparent Psi Project (Kekecs & Aczel, 2018).

By avoiding statistical practices that do not distinguish between exploration and confirmation and by implementing falsifiable confirmatory research methods, parapsychologists can be leaders in the ongoing evolution of research methodology.

## References

- Fricker, R.D., Burke, K., Han, X., & Woodall, W.H. (2019). Assessing the statistical analyses used in *Basic and Applied Social Psychology* after their *p*-value ban. *The American Statistician*, 73, sup1, 374-384, doi:10.1080/00031305.2018.1537892
- Kekecs, Z. & Aczel, B. (2018). Transparent Psi Project - Research Plan. Retrieved from [osf.io/d7sva/](https://osf.io/d7sva/).
- Kennedy, J.E., & Watt, C.A. (2018). How to plan falsifiable confirmatory research. Retrieved from [jeksite.org/psi/falsifiable\\_research.pdf](https://jeksite.org/psi/falsifiable_research.pdf)
- KPU Study Registry, (2012). Retrieved from [koestlerunit.wordpress.com/study-registry/](https://koestlerunit.wordpress.com/study-registry/)
- Tong, C. (2019). Statistical inference enables bad science; Statistical thinking enables good science. *The American Statistician*, 73, sup1, 246-261, doi:10.1080/00031305.2018.1518264
- Wasserstein, R.L., Schirm, A.L., & Lazar, N.A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73, sup1, 1-19, doi:10.1080/00031305.2019.1583913
- Watt, C. (2017). A prospective meta-analysis of pre-registered ganzfeld ESP studies. Retrieved from [www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU\\_Registry\\_1024.pdf](http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_Registry_1024.pdf)