PERSPECTIVE

Petitions in scientific argumentation: dissecting the request to retire statistical significance

Tom E. Hardwicke¹ & John P. A. Ioannidis^{1,2,3}

¹Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Charité –
Universitätsmedizin Berlin

²Meta-Research Innovation Center at Stanford (METRICS), Stanford University

³Departments of Medicine, of Health Research and Policy, of Biomedical Data Science, and of
Statistics, Stanford University

Author Note

Correspondence concerning this article should be addressed to John P. A. Ioannidis, MD, DSc, Stanford University, 1265 Welch Rd, Medical School Office Building, Room X306, Stanford, CA 94305 (jioannid@stanford.edu).

KEY MESSAGES

- Petitions have a long history of being used for political, social, ethical, and injustice issues, however, it is unclear how/whether they should be implemented in scientific argumentation.
- Recently, an extremely influential commentary published in Nature (Amrhein et al., 2019) calling for the abandonment of “statistical significance” was signed by 854 scientists.
- We surveyed signatories and observed substantial heterogeneity in respondents’ perceptions of the petition process, motivations for signing, and views on aspects of abandoning statistical significance.
- The top-cited signatories were strongly concentrated in a few scientific fields.
- In a random sample of 100 signatories, 62 published at least one paper in 2018 using statistical inference and most of them had used the phrase “statistical significance”.
- When scientists sign petitions, they may have variable views on important aspects and it is useful to understand this diversity.

Keywords: meta-research, petitions, p values, statistical significance, statistics

Petitions have a long history of being used for political, social, ethical, and injustice issues. The accumulation of large numbers of signatories suggests that a given cause has wide popular support and deserves attention from those who can act to address the concerns of the petitioners.^{1,2} It is unclear whether this method of raising attention and instigating change can or should also be implemented in the context of scientific argumentation. We use the term "petition" not in the strict sense of addressing a cause to a court or magistrates but in the broad sense of collecting multiple signatures to try to influence an agenda of decision making or practice.

Scientists have signed petitions on political, social, ethical, and injustice matters where science is relevant, such as climate change, priorities for research, alluded ethical breaches, or calls to retract flawed papers³. Additionally, scientists have organized petitions in order to influence aspects of the scientific ecosystem in which they operate. For example, over 80 academic signatories with editorial positions in diverse scientific journals petitioned for journals to offer the *Registered Reports* article format which they claimed will reduce bias and improve the quality of scientific outputs⁴.

Recently, more than two thousand researchers signed a petition to boycott a new journal - *Nature Machine Intelligence* – on the grounds that it was not open access⁵ and a petition supporting funder-mandated open access has gained over 1,900 signatories.⁶ Conversely, a petition opposing an initiative to mandate open access for scientific publications resulting from European public grants (“Plan S”) – has gained over 1,700 signatories⁷. The rationale in these cases is similar to the rationale behind the collection of signatories from the general population:

by weight of numbers the organizers are hoping to draw attention to their cause and convince key decision makers to act in their favour.

In the cases described above, the argument is of the form: N people support this view; therefore, this view deserves attention. However, this strategy may not be appropriate in the context of scientific argumentation in which one is trying to convince others that the signed viewpoint is “correct”. In this case the argument is of the form: N people support this view; therefore, this view is correct. This approach is arguably a logical fallacy (argumentum ad populum) because it conflates *justification* of a belief with the *acceptance* of a belief by a given group of people⁸. Science is a communal effort and it is essential to integrate the contributions of all scientists; however, the validity of arguments and data, should take precedence over number of voters. A single scientist may be correct in a debate against thousands, if his/her evidence or argument is stronger.

Recently, *Nature* invited a commentary⁹ to accompany the release of an issue of *The American Statistician* with over 40 papers on the theme of statistical inference in the 21st century (‘Moving to a World Beyond “ $p < 0.05$ ”’).¹⁰ In the commentary, Amrhein and colleagues explicitly and strongly urged the abandonment of the notion of “statistical significance” across science. The use and abuse of null hypothesis significance testing as a mechanism of scientific inference has been a hotly debated topic over many decades.¹¹ Very different views have been expressed and consensus is distinctly lacking among experts (e.g., see 21 heterogeneous commentaries accompanying the American Statistical Association’s 2016 Statement on p-Values¹²). Lack of consensus is also immediately manifest in everyday statistical practice and scientific publications, where statistical methods and inference tools are variably used, misused, and debated.^{13,14}

Amrhein inquired with *Nature* how to handle potentially multiple authors, as a very large number of people had expressed interest in these concepts. Given the difficulty of having many authors in a short commentary, *Nature* instead proposed to limit the number of the authors of the commentary to only three, and have additional scientists be signatories. A campaign followed that aimed at the collection of signatures in what was called a “petition” on the widely popular blog site of Andrew Gelman.¹⁵ Ultimately, 854 scientists signed the petition and the list of their names was published along with commentary.⁹ We should clarify that the authors of the *Nature* commentary expressly stated when we sent them a pre-print of this Perspective that they do not see their paper as a petition.

During the process of signature collection, invited scientists were given a short description of the effort (e.g., “Our comment is titled Retire Statistical Significance-...and it focuses on the absurdities generated by so-called ‘proofs of the null’”) and asked to sign in support. Invitations were distributed via social media (Twitter) and e-mail communications. Typically, the full comment was not sent alongside the invitation but the inviting authors said they “would be delighted to send you a near final draft of the comment (subject to sub-editing by the journal).” Signatories also received a link with the final version of the commentary just prior to its publication.

The commentary received enormous attention upon its publication resulting in the highest Altmetric score ever to-date¹⁶. Clearly the social dynamics of having so many co-signatories contributed to this wide dissemination. Concurrently, strong opposing views were also expressed, finding major problems with the proposal and concluding that abandoning statistical significance would do more harm than good across science.^{17,18,19,20}

Given that this is a very unusual, but also interesting opportunity, we decided to perform a survey of the signatories to understand how and why they signed the endorsement, and their scientific perspectives on various aspects of abandoning statistical significance.

The survey instrument (available here: <https://osf.io/zkwgf/>) consisted of 10 multiple choice questions presented on a single webpage. All questions were optional. The survey invitation e-mail (available here: <https://osf.io/b9sdv/>) was dispatched from the e-mail address of the first author (T.E.H.) using the survey software *SoSci Survey* (<https://www.soscisurvey.de/>). Contact e-mail addresses were provided by signatories alongside their signatures⁹.

We aimed to publish an editorial perspective along with a survey of this sort where fellow expert scientists are asked voluntarily and anonymously to express their scientific opinions on technical statistical, not private, matters. Nevertheless we did seek independent approval and in consultation with the leadership of QUEST where METRIC-B is hosted in Berlin determined that no IRB approval is needed for this survey. QUEST itself was not involved in any way in the survey or the perspective.

The invitation was sent on 4th April 2019 and the survey was closed two weeks later on 18th April. No reminders were sent. For each signatory, the invitation e-mail contained a unique and anonymous link to the survey which allowed them to complete the survey only once. This prevented any non-signatories from completing the survey, unless a signatory shared their link with somebody else and had not already completed the survey via that link themselves. The link could be revisited if responses remained incomplete. Due to the “Anonymous” privacy mode of the SoSci Survey software (see <https://perma.cc/XP69-HCPK>), it was impossible for anybody to link participants’ responses to their identities, or to see who had or had not participated. One

scientist changed their mind and wanted to have their responses removed, but this was not possible.

Of all signatories, we received 248 survey responses (after removing one where no questions were answered): a 29% response rate. Responses are presented in Table 1.

TABLE 1. Responses to survey questions. Verbatim free text responses for “Other please specify” options are available in Table S1.

Question and response options	<i>n</i>	Percentage [95% CI]
Q1. When Amrhein et al. invited signatories they initially provided a summary statement about the Comment and offered to share a draft version of the Comment upon request. How did you decide to sign?		
A. I requested and read the draft version of the Comment before signing and had no concerns/comments to share with the authors.	172	69% [64% to 75%]
B. I requested and read the draft version of the Comment, made queries/comments to the authors and waited for satisfactory replies before signing.	6	2% [0% to 8%]
C. I requested and read the draft version of the Comment before signing and made queries/comments to the authors but did not feel they were so essential as to have to wait for their responses before I signed.	13	5% [0% to 11%]
D. I did not read the draft version of the Comment and I signed based on the summary statement since this is what my support was intended for.	43	17% [12% to 23%]
E. Other, please specify.	13	5% [0% to 11%]
<i>Not answered.</i>	1	0% [0% to 6%]

Q2. Amrhein et al. also sent a link to the finalized version of the Comment to all people who had signed the petition. Did you read the finalized version of the Comment before its publication in Nature?

A. Yes.	194	78% [73% to 83%]
B. No, but I read it after its publication.	35	14% [9% to 19%]
C. No, but I may read it in the future.	12	5% [0% to 10%]
D. No, and I do not plan to read it.	4	2% [0% to 6%]
E. Other, please specify.	2	1% [0% to 6%]
<i>Not answered.</i>	1	0% [0% to 5%]

Q3. To what extent do you agree with the final published Comment?

A. I agree 100% with every single word and every claim made in it.	38	15% [9% to 22%]
B. I agree with the large majority of what is stated there*.	96	39% [32% to 45%]
C. I agree with the majority of what is stated there.	97	39% [33% to 46%]
D. I agree with a minority of what is stated there, but still felt that it was worth supporting.	3	1% [0% to 8%]
E. I have not read it.	5	2% [0% to 9%]
F. Other, please specify.	8	3% [0% to 10%]
<i>Not answered.</i>	1	0% [0% to 7%]

Q4. Are you concerned that abandoning (retiring) statistical significance may lead to people interpreting results subjectively and with biased preconceptions?

A. I am very concerned about this possibility.	8	3% [0% to 10%]
B. I am somewhat concerned about this possibility.	53	21% [15% to 28%]
C. I am a little concerned about this possibility.	90	36% [30% to 43%]
D. I am not at all concerned about this possibility.	90	36% [30% to 43%]
<i>Not answered.</i>	7	3% [0% to 10%]

Q5. Are you concerned that abandoning (retiring) statistical significance may make things worse overall across the scientific literature?

A. I am very concerned about this possibility.	3	1% [0% to 8%]
B. I am somewhat concerned about this possibility.	19	8% [2% to 14%]
C. I am a little concerned about this possibility.	75	30% [24% to 37%]
D. I am not at all concerned about this possibility.	145	58% [52% to 65%]
<i>Not answered.</i>	6	2% [0% to 9%]

Q6. What do you think about replacing 'confidence interval' with 'compatibility interval'?

A. I strongly prefer 'compatibility interval' over 'confidence interval'.	29	12% [6% to 19%]
B. I modestly prefer 'compatibility interval' over 'confidence interval'.	68	27% [21% to 34%]
C. I have no preference for either term.	107	43% [37% to 50%]
D. I modestly prefer 'confidence interval' over 'compatibility interval'.	32	13% [7% to 20%]
E. I strongly prefer 'confidence interval' over 'compatibility interval'.	7	3% [0% to 10%]

<i>Not answered.</i>	5	2% [0% to 9%]
----------------------	---	---------------

Q7. Reliance on statistical significance may lead to unwarranted statements of 'no difference', but also unwarranted claims of associations that don't really exist. Which problem is more serious/common?

A. Unwarranted statements of 'no difference' is a more serious/common problem than unwarranted claims of associations that don't exist.	27	11% [5% to 17%]
---	----	-----------------

B. Unwarranted statements of 'no difference' and unwarranted claims of associations that don't exist are equally serious/common problems.	163	66% [60% to 72%]
---	-----	------------------

C. Unwarranted statements of 'no difference' is a less serious/common problem than unwarranted claims of associations that don't exist.	46	19% [13% to 24%]
---	----	------------------

<i>Not answered.</i>	12	5% [0% to 11%]
----------------------	----	----------------

Q8. When you signed the petition, what benefits did you feel this would bring to the argument to retire (abandon) statistical significance? (Please tick all that apply)

A. I felt it would help to draw attention to the argument.	207	83% [79% to 88%]
--	-----	------------------

B. I felt that it would make the argument more convincing.	77	31% [25% to 37%]
--	----	------------------

C. I did not consider the potential benefits to the argument of signing the petition.	16	6% [4% to 10%]
---	----	----------------

D. Other, please specify.	39	16% [12% to 20%]
---------------------------	----	------------------

Q9. In a recent pre-registered randomized trial comparing two septic shock treatments with formal power calculations for determining the necessary sample size to evaluate the primary outcome of mortality, the mortality was 43 per cent for the older treatment and 35 per cent for

the newer one and the p-value was 0.06. What do you think about the statement 'the new treatment clearly reduced mortality'?

A. I strongly agree. The treatment clearly reduced mortality. Claiming that there is no difference between the two treatments is appalling science.	29	12% [6% to 18%]
B. I mostly agree.	55	22% [16% to 28%]
C. I tend to agree.	68	27% [21% to 34%]
D. I tend to disagree.	45	18% [12% to 24%]
E. I mostly disagree.	25	10% [4% to 16%]
F. I strongly disagree. Claiming that there is a difference between the two treatments is appalling science.	8	3% [0% to 10%]
<i>Not answered</i>	18	7% [1% to 14%]

Q10. For which fields do you currently find that the idea to abandon (retire) statistical significance is appropriate?

A. It is appropriate for all fields of science.	93	38% [31% to 44%]
B. It is appropriate for my field and perhaps also other fields, but not necessarily all fields of science.	123	50% [43% to 56%]
C. It is mostly not appropriate for my field, but it is mostly appropriate for many other fields of science.	5	2% [0% to 9%]
D. At this point, I think it is mostly not appropriate.	6	2% [0% to 9%]
E. Other, please specify.	16	6% [0% to 13%]
<i>Not answered</i>	5	2% [0% to 9%]

*Note that, contrary to the pre-registered protocol, one of the response options for question 3 (“I agree with the large majority of what is stated there.”) was initially missing from the live survey due to researcher error (T.E.H.). The omission was noticed within 60 minutes of the survey launch and corrected. At the time of correction, 25 complete surveys had been submitted, but an unknown number of surveys had been started.

CI: confidence interval

As per our original protocol, we explored whether the responses to questions 8 (perceived benefits of signing the petition) and 10 (appropriateness for all science, own field, and other fields) were associated with the responses to each of the other 8 questions of the survey. No strong associations were observed (see Supplementary Information).

Overall, although most respondents reported that they agreed with the ‘majority’, ‘large majority’, or ‘every single word’ of the commentary (Question 3), responses to certain survey questions revealed some diversity of viewpoints that were not otherwise apparent. For example, the commentary focused on when reliance on ‘statistical significance’ might lead to unwarranted claims of ‘no difference’ but put less emphasis on the issue of unwarranted claims about associations that are actually spurious.²⁰ Most respondents (66%) felt that the two issues were equally serious/common and 19% felt that unwarranted claims of ‘no difference’ were less serious/common than unwarranted claims about associations that are spurious (Question 7). Similarly, Question 9 probed a statement by an extremely influential statistician (also a signatory), David Spiegelhalter. In a Financial Times interview,²¹ Spiegelhalter, a towering giant in Bayesian statistics, described as “appalling science” the conclusion of no difference in a large randomized trial of sepsis treatment²² that was published in JAMA. The trial involved formal power calculations and mortality was 43% and 35% respectively in the two arms with $P=0.06$. About one third of signatories agreed with the interpretation in JAMA. If one has high skepticism and a low prior for effectiveness of sepsis treatments (many have been tested, but almost all have failed, despite frequent earlier promises) than the conclusion of no difference would be appropriate.

It is unclear if the endorsement of the 854 signatories should be restricted to the request to abandon statistical significance or applied to the Nature commentary more broadly. Surprisingly, 6% of respondents reported that they had never read the Nature commentary and 14% only read it after publication (Question 2). 17% reported that they had signed on the basis of a summary statement and that this is what their support was intended for (Question 1). Notably, responses to the question about replacing the term “confidence intervals” with “compatibility intervals” (Question 6), suggested that many respondents (43%) were ambivalent on this issue and some (16%) even expressed the opposite preference. This definitional issue is somewhat tangential to the main goal of abandoning statistical significance and these responses reinforce the idea that, at least for some signatories, their endorsement should not necessarily be interpreted as applying to all aspects of the commentary.

Most respondents stated that they signed the petition in order to draw attention to the statistical significance issue and almost a third felt it would make the argument more convincing (Question 8). Arguably the latter goal is a logical fallacy (*argumentum ad populum*)⁸. Many respondents also used the opportunity to provide free text ‘other’ responses to this question, some of which conveyed frustration about the decades long effort to address problems related to statistical significance (see Table S1). The sentiment amongst some appeared to be that scientific argumentation alone had not been sufficiently effective, and that in order to address what they viewed as a cultural/political problem, a cultural/political solution (i.e., a petition) was justifiable.

As statistical methods are hotly debated and attract the interest of many scientists, several papers proposing ways to improve the current status include large numbers of authors. The Nature commentary increased 10-fold the number of names involved compared with previous

papers, but did not ask the signatories to be involved at the level required for authorship. It is unknown to which depth formal authorship criteria can be met for all authors involved in extensively multi-authored papers. Conversely, it is also unknown whether disengaging scientists from the intellectual commitment and scrutiny required to reach authorship credit and asking them to be simply signatories is beneficial or harmful.

One perspective on scientific argumentation states that it is irrelevant *who* has made an argument; it is the substance of the argument that is of primary importance²³. However, when a petition is used to support a scientific argument, the expertise, experience, and perspective of the signatories seems highly relevant to how that petition should be interpreted. We selected a random sample of 100 individuals from the 854 signatories to examine their stance towards statistical significance in their own published work. For 7 of these signatories, we were unable to locate them in Web of Science or find a publicly accessible online profile. Among the other 93, CVs could be located for 13 individuals, but only 2 of them had been definitely updated after the publication of the Nature commentary, so we cannot tell if/whether the signatories would eventually include the commentary in their CVs. As shown in Table 2, based on Web of Science slightly less than two thirds of the sampled signatories had published at least one article in 2018 reporting quantitative data. When they did publish something, inferential statistics had almost always been used in their most recent 2018 quantitative article, with the large majority including the phrase “statistical significance” (or similar) - the practice which the commentary suggests should be abandoned. We cannot exclude the possibility that these scientists were won over to the cause of abandoning statistical significance very recently, e.g. in 2019, or that they were pressured by reviewers, editors, or funders to use terminology that they did not wish to use.

TABLE 2. Characteristics of paper-level analysis for random sample of 100 signatories.

Characteristic	<i>n</i>
Published at least one article in 2018 reporting quantitative data	62
Most recent 2018 quantitative article...	
...reports inferential statistics	55
...uses the term “(statistically) significant” or “(statistical) significance”	43
...reports at least one p-value	35
...reports only p-values	20
...reports p-values and confidence intervals	14
...reports p-values and false discovery rate	1
...reports using only confidence intervals	17
...reports using only Bayesian techniques	3

For each of the 854 signatories we used Scopus to identify whether they were among the ‘top-cited’ 100,000 scientists across science based on a composite citation indicator²⁴ that combines total citations, H index, co-authorship adjusted Hm index, citations to papers as single author, as single or first author, and as single, first, or last author. Data have been extracted on these 6 indicators for citations received in 1995-2017 and the composite calculated on all author records in collaboration with Scopus/Elsevier.³³ 68 signatories (8%) were among the ‘top-cited’ 100,000 scientists across science based on that composite citation indicator. As shown in Table 3, Medicine, Psychology/Cognitive Sciences, Public Health/Health Services, and Economics/Business were the most common fields for 53/68 top-cited signatories. Looking at more granular fields, for 44/68 top-cited signatories 4 of the 176 fields of science (Epidemiology,

Social Psychology, Oncology/Carcinogenesis, Public Health) were the most common or second most common field of their publications.

TABLE 3. Characteristics of 68 signatories who were among the ‘top-cited’ 100,000 scientists.

Characteristics	Median (25th-75th percentile)
Ranking (with self-citations)	28379 (7881-51096)
Ranking (without self-citations)	28370 (7827-49053)
Most common higher-level Science-Metrix category (out of 22 categories)	<i>n</i>
Medicine	25
Psychology & Cognitive Sciences	13
Public Health & Health Services	10
Economics & Business	5
Biology	4
Social Sciences	4
3 Others	< 4
Most common Science-Metrix subject categories (out of 176 categories)	<i>n</i>
Epidemiology	11
Social Psychology	8
Oncology & Carcinogenesis	5
Public Health	5

	17
Business & Management	4
Ecology	4
21 Others	< 4
<hr/>	
Second most common Science-Metrix subject categories (out of 176 categories)	<i>n</i>
<hr/>	
Epidemiology	7
General & Internal Medicine	7
Social Psychology	6
Evolutionary Biology	4
Experimental Psychology	4
Obstetrics & Reproductive Medicine	4
Statistics & Probability	4
25 Others	< 4
<hr/>	

Interestingly, the heavily represented fields (epidemiology, public health, medicine, social psychology) among the top-cited signatories are mostly fields where serious concerns and debates about low reproducibility have been raised.²⁵⁻³⁰ This gives some insight into certain characteristics of the signatories, but knowing the scientific field of all signatories would have been informative, particularly in light of responses to Question 10 – 50% of respondents said they felt that abandoning statistical significance was appropriate for their field and were more cautious about whether it should also apply to other fields.

For fields where there are prevalent concerns about low reproducibility, abandoning the concept of statistical significance would make claims of “irreproducibility” difficult if not impossible to make, although Amrhein and Greenland disagree with this perspective.^{31,32} In our opinion this approach may give bias a free pass.¹⁹ There is tension therefore on whether abandoning statistical significance would improve statistical reasoning or inappropriately silence worrisome claims of irreproducibility in fields that seem to have the greatest problems of irreproducibility. The former consequence would be most welcome, while the latter could be a disaster.

Some important limitations and caveats should be considered. Firstly, it was not our intention to make broad inferences about the views of all scientists; the 854 signatories are likely a highly biased sample of the wider scientific community. They may not be representative even of all scientists who would otherwise support many of the positions in the Nature Comment, but for various reasons did not have an opportunity or did not wish to sign. Thirdly, the response rate was reasonable (29%) for this type of single communication survey, but it is likely that responders are different to non-responders. It is not clear in what direction this would bias the results. For example, it is unknown if signatories who were very much in favor of the positions of the Nature commentary would be more likely to complete the survey (to testify their fervent enthusiasm) or less likely to complete it (feeling that the survey is causing unnecessary questioning).. Thirdly, this is a single case study of signature collection, and thus inferences may not extend necessarily to other petitions that may have higher or lower agreement between signatories in the positions expressed. Petitions that recruit signatories may differ also on whether single, simple positions are expressed, or many statements are incorporated that may attract different degrees of enthusiasm. Nevertheless, we feel that given the scientific

importance of the matter addressed in this specific situation and the great attention it received, any insights obtained from it would be useful. Finally, these data may bear on the utility of the petition, but do not directly pertain to the merits of the argument to retire statistical significance, which has been debated elsewhere.¹⁷⁻²⁰

In conclusion, we are grateful to all colleagues who shared their insights about their participation in the Nature petition and their perspectives about the hot topic of retiring statistical significance. If and/or how petitions should be used in scientific argumentation remains an open question, however we hope that the insights obtained in this exercise might provide the foundations for an informed debate. We have invited and received extensive insightful comments from the authors of the Nature commentary and we have revised our Perspective paper accordingly. The original version and revised version are available at <https://doi.org/10.31222/osf.io/73xm5>. We hope the current paper will stimulate further debate.

Open practices statement

The study protocol was pre-registered on April 4, 2019 and is available at <https://osf.io/hbkj3/>. The protocol was hidden under embargo for the duration of the survey to avoid contamination of participant responses. The introductory text of this report is largely based on the text of the pre-registered document. All procedural deviations from the pre-registered protocol are explicitly acknowledged in this article. All data exclusions and measures conducted during this study are reported in this article. Data, materials, and analysis scripts related to this study are publicly available at <https://osf.io/8aenp/>. To facilitate reproducibility the analysis script is available in a Code Ocean container (<https://doi.org/10.24433/CO.3912558.v1>) which re-creates the software environment in which the original analyses were performed.

Funding statement

The authors received no specific funding for this work. The Meta-Research Innovation Center at Stanford (METRICS) is supported by a grant from the Laura and John Arnold Foundation. John Ioannidis is a Berlin Institute of Health visiting Einstein fellow and the Meta-Research Innovation Center Berlin (METRIC-B) is supported by a grant from the Einstein Foundation and Stiftung Charité.

Competing interests statement

John P.A. Ioannidis is editor-in-chief of the European Journal of Clinical Investigation and the opinions expressed in this perspective are those of the authors and not of any institution. John Ioannidis has previously taken a stance on issues of what statistical inference methods should be used and how and his positions can be seen in the references that he has authored or co-authored which are cited in this Perspective. His expressed comments of the Nature commentary content and process can also be seen in the highly popular blog of Andrew Gelman. Tom E. Hardwicke declares no competing interests.

Acknowledgements

We are grateful to all the scientists who responded to our questions and who have contributed constructively on the ongoing debate about statistical significance. Many thanks to the authors of the Nature commentary for their insightful comments and criticisms on an earlier version of this Perspective.

References

1. Miller H. Petitioning and demonstrating. In D. Brown, G. Pentland, & R. Crowcroft (Eds.), *The Oxford Handbook of Modern British Political History, 1800-2000*. Oxford University Press, 2018.
2. Norris P. *Democratic phoenix: Reinventing political activism*. Cambridge University Press, 2002.
3. Ripple WJ, Wolf C, Newsome TM., Galetti M, Alamgir M, Crist E, ... & 15,364 scientist signatories from 184 countries. World scientists' warning to humanity: a second notice. *BioScience* 2017;67:1026-1028.
4. Chambers C et al. Trust in science would be improved by study pre-registration. *The Guardian*. Retrieved from: <https://perma.cc/CT9A-T3RD> on March 21, 2019.
5. McCook A. Thousands boycott new Nature journal about machine learning. *Retraction Watch*. Retrieved from: <https://perma.cc/Q4SK-3K4M> on March 21, 2019.
6. Open Letter in Support of Funder Open Publishing Mandates (n.d.). Retrieved from <https://perma.cc/G3PS-48Q6> on March 21, 2019.
7. Plan S Open Letter. Retrieved from <https://perma.cc/PE47-HKFE> on March 21, 2019.
8. Walton DN. Why Is the 'ad Populum' a Fallacy? *Philosophy & Rhetoric* 1980;13:264-278.
9. Amrhein V, Greenland S, McShane B. Retire statistical significance, *Nature* 2019;567:305-307.
10. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond " $p < 0.05$ ". *The American Statistician* 2019;73:1-19.
11. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods* 2000;5:241-301.
12. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: context, process, and purpose. *The American Statistician* 2016;70:129-133.
13. Gigerenzer G. Mindless statistics. *Journal of Socioeconomics* 2004;33:567-606.
14. Strasak AM, Zaman Q, Marinell G, Pfeiffer KP, Ulmer H. The use of statistics in medical research. *The American Statistician*, 2007;61:47-55.
15. Gelman A. "Abandon / Retire Statistical Significance": Your chance to sign a petition! [Blog post]. Retrieved from <https://perma.cc/2L3Q-HEH2> on April 4, 2019.

16. Altmetric. Scientists rise up against statistical significance - Overview of attention for news story in Nature. Retrieved from <https://perma.cc/2865-B7KA> on June 20, 2019.
17. Haaf JM, Ly A, Wagenmakers EJ. Retire significance, but still test hypotheses. *Nature* 2019;567:461.
18. Johnson VE. Raise the bar rather than retire significance. *Nature* 2019;567:461.
19. Ioannidis JPA. Retiring statistical significance would give bias a free pass. *Nature*, 2019;567:461.
20. Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: Do not abandon significance. *JAMA* 2019;321:2067–2068.
21. Ahuja A. Scientists strike back against statistical tyranny. *The Financial Times*. Retrieved from: <https://www.ft.com/content/36f9374c-5075-11e9-8f44-fe4a86c48b33> on June 26, 2019.
22. Hernández G, Ospina-Tascón GA, Damiani LP, et al. Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock: The ANDROMEDA-SHOCK randomized clinical trial. *JAMA* 2019;321:654–664.
23. Merton RK. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, 1973.
24. Ioannidis JPA, Klavans R, Boyack, KW. Multiple citation indicators and their composite across scientific disciplines. *PLOS Biology*, 2016;14:e1002501.
25. Ioannidis JPA. Why most published research findings are false. *PLoS Medicine* 2005;2:e124.
26. Pashler J, Wagenmakers E-J. Editors' introduction to the special section on replicability in psychological science. *Perspectives on Psychological Science* 2012;7:528-530.
27. Feinstein A. R. (1988). Scientific standards in epidemiologic studies of the menace of daily life. *Science*, 1988;242:1257–63.
28. Le Fanu J. *The rise and fall of modern medicine*. New York: Little Brown, 1999.
29. Skrabanek P. *False premises, false promises*. Whithorn: Tarragon Press, 2000.
30. Taubes G. Epidemiology faces its limits. *Science* 1995;269:164–69.
31. Amrhein V, Korner-Nievergelt F, Roth T. The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research. *PeerJ* 2017: e3544.

32. Amrhein V, Trafimow D, Greenland S. Inferential statistics as descriptive statistics: there is no replication crisis if we don't expect replication. *The American Statistician* 2019;73(sup1):262-270.
33. Ioannidis JPA, Baas J, Klavans R, Boyack KW. A standardized citation metrics author database annotated for scientific field. *PLOS Biology*, in press.

Supplementary information

for

Petitions in scientific argumentation:

dissecting the request to retire statistical significance

Tom E. Hardwicke & John P. A. Ioannidis

Contents

1. Table S1. Verbatim free text responses for the “Other, please specify” option available for questions 1, 2, 3, 8, and 10.
2. Supplementary methods: Associations between survey responses
3. Table S2. Association between question 8 and other questions 1-7, 9, and 10.
4. Table S3. Association between question 10 and other questions 1-7 and 9.

Table S1. Verbatim free text responses for the “Other, please specify” option available for questions 1, 2, 3, 8, and 10.

Q1. When Amrhein et al. invited signatories they initially provided a summary statement about the Comment and offered to share a draft version of the Comment upon request. How did you decide to sign?

6 provided no details.

1. "I signed the statement based on the summary statement, before I got the draft version of the Comment. I read the Comment in full afterwards, but still believe the support is equally valid when reading just the summary statement."
2. "I did not read the draft version and signed nevertheless, because I knew from previous papers of Amrhein and Greenland that I share their opinion."
3. "I requested and read the draft version of the Comment before signing. I had objections but did not communicate them, as I saw that they were echoed by others and felt confident that they would be addressed before the final draft, and they were."
4. "The above statement is incorrect: I had a full draft available along with the invitation and I read the draft before making up my mind."
5. "I requested and read the draft version of the Comment before signing. Had a few queries/comments to the authors but did not feel they were so essential as to have to contact them before I signed."
6. "I requested and read the draft version of the Comment before signing and agreed with the general sentiment and thought the suggestions were feasible and likely to have impact. I did not feel, as a non-coauthor, that my (minor) questions/concerns were relevant."
7. "I decided before reading what would be a \"deal breaker\" for me: things that were (not) in the document, that would prevent me from supporting it. Then I read the draft version of the Comment before signing and had no concerns/comments to share with the authors."

Q2. Amrhein et al. also sent a link to the finalized version of the Comment to all people who had signed the petition. Did you read the finalized version of the Comment before its publication in Nature?

1. "Skimmed to see the changes"
 2. "I read it but am unsure of the timing. Given I was behind in email, I'd give 1:2 odds for Yes:After."
-

Q3. To what extent do you agree with the final published Comment?

1 provided no details.

1. "I agree with the analysis, but still do not know how to pass medical papers past reviewers. Examples should be discussed, e.g how to convert epidemiological papers with 90 p-values"
 2. "I agree with the majority of what is stated, but what is most relevant is that I agree with the main conclusion"
 3. "I agree with the text completely, but there are caveats and extensions that I believe would be appropriate (but more for a full paper, not in a comment/letter)"
-

-
4. "In light of reading the rest of TAS19 articles, I think each has merits & demerits. In Valentin et al, the 'compatibility interval' could have been clearer"
 5. "I agree with the notion that there should be debate about this issue. As to the solutions offered in the comment, my agreement is mixed."
 6. "I've never come across any similar statement on statistical issues with which I totally agree, confirming Tukey's view that the collective noun for a group of statisticians is a \"quarrel\""
 7. "I think the first option is inappropriate and misleading. It reads as if you have some specific ideas in mind with this survey ... I think it is unfortunate that a person who expressed is concerns with the statement already (JPAI) is responsible for the survey. The other questions miss a freetext field ... Q9: the question is not whether it reduced mortality in this trial (43% and 35% is different or am I missing something here?) but what it means beyond the trial itself (the past tense in your statement is the problem)!"
-

Q8. When you signed the petition, what benefits did you feel this would bring to the argument to retire (abandon) statistical significance? (Please tick all that apply)

1. "As an editor of a biomedical journal, I plan to recommend citation of the comment in the editorial policy."
 2. "I felt that, more than just drawing attention to the argument, it would spark off a discussion. And it did."
 3. "I don't want scientists misuse the statistical significance"
 4. "I have read many dozens of similar articles; there are hundreds of them, everybody seems to agree with them and nothing ever changes. I don't sympathize with the current article's focus on false negatives, but that's ok. At the end, making binary "decisions" based on one * of one study is pretty much insanity, in my opinion. I would never do anything like that if I wanted to learn something from my data. Whenever I summon up those *s I am doing this due to cultural reasons. This is why the response needs be a cultural one as well. I think that statistical significance is already being abandoned and am modestly hopeful about the future. I just want it to happen more quickly, so I will have fewer articles of mine to look back on and be embarrassed about. To all the tenured professors who can't imagine what to do when the stock prices of those *s crash: maybe ask your PhD students who are pushing predictive models to GitHub all day what they are doing. You might learn something."
 5. "I felt it would help young researchers that want to make better science, and interpret all aspects of a study (not just a p-value or "statistical significance") to argue with their supervisors that are often mostly concerned about the value of the p-value. Young researchers need support presented in a way similar to this paper with a list signatories, because purely scientific arguments (on how to use statistics) seem not to work."
 6. "The petition may help scientists to look at their data in a much more thoughtful way instead of using a single potentially biased cut-off. A 'strong' statement and position paper I think will help and improve the scientific research. The first 2016 ASA statement had to be remarked."
 7. "I do not think adding more voices to the argument strengthens it and that logical arguments made by the author stand on their own. The main point in signing was to draw attention to the problem"
 8. "While a petition may sound as an awkward instrument in a scientific context I think it is relevant to show that many people share the same concerns."
-

-
9. "I did not consider it to be a petition. I signed to show I strongly supported the arguments made in the commentary"
 10. "I felt it would demonstrate that the "norms of scientific research" do not require statistical significance testing"
 11. "In my humble opinion the petition would provide more evidence on avoid binary decisions based on p-value only."
 12. "Sometimes political pressure is needed to make changes. It is not always only about science, it is often about money and political power! Everything else is naive."
 13. "I mainly wanted to encourage the authors and tell them "I am also here, you are not alone, I agree with you""
 14. "I get so annoyed when people dismiss a results because P is not less than 0.05. Focus on the confidence limits."
 15. "I wanted to publicly commit to the principles in the statement"
 16. "The Nature comment is an excellent article and if more people read it, that will increase understanding of the issues."
 17. "I just agreed"
 18. "It was a means to express my professional opinion."
 19. "It would help my teaching practice at home, with my students. I come from [name of country], and we have a gap to bridge."
 20. "Topics as such need visibilization. The battle might be lost for regular NHST or basic frequentists users, but it is important to sensibelize the discussion for current and future students. Developing a critical stance towards NHST is an important motivation to go beyond mindless point-and-click, threshold-searching analyses."
 21. "I think that the ML community (and causal inference experts) should help to structure the alternative tools for validation of analysis (instead of p-value)"
 22. "I thought that there were inherent fallacies in the exercise of NHST & p values. I am still scoping out whether p values can be interpreted in a useful way. Not fully sure yet."
 23. "I also hoped it would make people aware that science based on significance test only is impossible. A scientific claim requires taking into account all relevant evidence, not only a decision to reject an imaginary statistical model based on a single index (p-value)."
 24. "I wanted to show support for the authors and encourage them to continue pushing for change."
 25. "I felt it would show the range of researchers (statisticians, epidemiologists, clinicians, etc) who are concerned about this issue."
 26. "It acts as a strong counterweight to ongoing pressure to normalize erroneous dichotomous reasoning (already a huge problem in my field, consistent with the main argument)"
 27. "I thought: I agree with the argument, why not sign it?"
 28. "The sad reality is that, among applied researchers, the majority view of statistics is that it is a ritual to be performed in the Results section in order to confirm the preconceptions that went into the design of the experiment. When the data aren't cooperative, they are very commonly tortured into submission by extensive undocumented revision. Getting people to employ hypothesis testing in a thoughtful, targeted way will first require that the scientific community
-

break the spell that $p < .05$ holds over to many of its members. A comment piece written by a few people won't accomplish this (as evidenced by decades of efforts), so it seems to me that gathering a much larger crowd is needed to get across to people that the complaints about the misuse of p-values is not a fringe opinion."

29. "Your question 9 requires more than 5 options. It is an incomplete presentation of the results and prior literature."
 30. "The general public (journalists) and indeed most practicing biologists don't understand the relationships between models and uncertainty."
 31. "Need for bold changes in the field and this is merely a first step that captures attention"
 32. "Expanding on my answer, many or most researchers who report their findings in terms of significance do so because this is what they have been told to do. They learn it in their courses and from their advisors and colleagues and the reviewers and editors who decide whether to publish their work. This sends a message that there is a strong consensus that the focus on statistical significance is appropriate. But there is not such a consensus, and a high profile commentary with hundreds of signatures from experts illuminates this fact. I know that John Ioannidis has stated his concern regarding the "popularity contest" aspect of a petition. In my opinion, the current use of statistical significance is the result of a much larger implicit popularity contest."
 33. "I have supported this idea for many years and thought this was the best simple presentation yet for the rationale"
 34. "I wanted to express my support to the argument"
 35. "I was expressing my agreement with the views stated in the paper."
 36. "I don't understand this question. What does "this" refer to? I want scientists and scientific journals to re-consider stringent demands to *always* t and p the data and *always* dichotomize findings regardless of whether a particular question demands such dichotomization."
 37. "I view this as a good tool to bring the issue to the attention of readers, researchers, and reviewers who have been unaware of the problem of the dichotomization of evidence based on p-values"
 38. "I agreed with the proposal and, since I cannot do research/have enough expertise myself on this topic, I felt it was important to support this initiative"
 39. "I felt it would help to draw attention to the current problems with the way statistical significance is used"
-

Q10. For which fields do you currently find that the idea to abandon (retire) statistical significance is appropriate?

1. "It is appropriate for fields which use to test qualitative research hypothesis by rejecting a point null hypothesis. For example: psychology, sociology, political science and all the other soft sciences. It is well explained by Paul Meehl as early as 1967 (as well as by great many other authors) and I can't see why we should go over this issue over and over again."
 2. "observational studies"
 3. "In population studies, it more appropriate for observational studies than randomised experiments; I am unsure about laboratory experiments."
-

-
4. "Part of the problem we see is with the use of this in writing, in papers. The other problem is on the verbal communication, when people make presentations without proper interpretation. That is the biggest problem for me and why I chose to sign this."
 5. "Regarding question 9: I want to answer "I am not convinced by the results""
 6. "It is appropriate for all fields where only a significance test is used to make a categorical claim of some effect or relation existing (or not) in reality."
 7. "all, except for confirmatory studies with a clear hypothesis,"
 8. "It is appropriate to my field, and others which rely heavily on erroneous dichotomous reasoning. Many fields are not prone to this issue."
 9. "The question of "which fields should abandon statistical significance" misunderstands the problem. Rather, hypothesis testing of the traditional sort should be limited to confirmatory analysis in preregistered contexts, such as (but not necessarily limited to) studies that are structured as clinical trials. The example given in question #9 is a good example of where problems can begin to bleed in. "Formal power analysis," for example, in inherently subjective even when properly executed, so given the very limited information provided in question #9, any statement that uses adverbs like "clearly" tends to be bad science, even when the result is statistical significant. Furthermore, it reflects a tendency to confuse group risk with individual risk, which is something that can only be understood by using covariates to build a model of individuals, rather than using them to "control for" those differences and in so doing treat those differences as nuisance variables. Scientists generally should get out of this mindset that the appropriateness of a statistical procedure is a function of which silo a study finds itself in, and instead focus on whether the procedure is appropriate to a particular study's experimental design. Every field has a mix of exploratory and confirmatory analysis, with the latter representing the cases that have sufficiently robust theory to make narrow predictions. Hypothesis testing is fine in those cases where the null is precise, rather than being a catch-all straw man, regardless of field. It is inappropriate in cases where theory is vague and the experimenter admits that "anything could happen," again regardless of field."
 10. "Any field anchoring to one test statistic is problematic. I think this survey fundamentally does not understand what statistical inference is. The survey itself put an unnecessary focus on one small component of drawing conclusions which is what the underlying purpose of the Nature commentary was trying to refute. There are no hard rules for statistical inference and we must appreciate models used to construct our arguments, not anchor to single measurement values."
 11. "You are giving a series of multiple choice, categorical options to a set of people who just agreed with the idea that categorization tends to oversimplify."
 12. "I'm a statistician, so I can't speak to "my field" of science. I think statistical significance can be appropriate in limited situations. But far too many students and researchers are taught that statistical significance testing should be done as a matter of routine in all of their research, and they are taught to interpret significance roughly as "the effect exists" and non-significance roughly as "the effect does not exist". Also, to comment on your question #7, I think the paradigm of "true" effects existing or not existing, while sometimes helpful, is also simplistic. If we have no reason to believe that the null is true, then statistical significance doesn't convey much useful information."
 13. "It is appropriate for my field and definitely others. Anyone in any field trying to estimate the size of an effect, is wrong if they conclude the effect to be 0 solely because the result was statistically insignificant-"
-

-
14. "The p-value may serve a purpose when it comes to distinguishing between two treatments A and B, in a situation where there is little circumstantial evidence to support either or. However, in many situations in epidemiology a study is based on prior and circumstantial evidence, and it is not meaningful to dismiss all that based on one single study not reaching $p < 0.05$. Patterns are more important than p-values! So given the choice between conserving or retiring the p-value, I would retire it - but I would agree to preserve it for the industry's treatment approvals."
 15. "Any field that is prepared to not let interpretations run wild"
 16. "At this point I'm not sure given the alternatives, but I think progress should be made. When good alternatives are available, I think abandoning would be appropriate for all fields."
-

Supplementary methods: Associations between survey responses

In the pre-registered protocol, we stated that we would examine the correlations between responses, however this was ambiguous because the data are ordinal and nominal rather than continuous. We therefore conducted Pearson Chi-Squared Tests of Independence, and the specification of these analysis (below) was not pre-registered. We excluded all "Other" and Missing responses from these analyses (they are available verbatim in Table S1). Questions 1, 2, 8, and 10 were treated as nominal variables; whereas questions 3 to 7 and 9 were treated as ordinal variables. The Chi-Squared Tests involving ordinal variables were adjusted for trend.

Due to low expected values in some cells, p -values were computed using Monte Carlo simulations. For question 8, responses were not independent because respondents could select all options that applied to them. Therefore, we recoded the responses into independent categories: *A only* (i.e., "I felt it would help to draw attention to the argument."); *B only* (i.e., "I felt that it would make the argument more convincing."); *C only* (i.e., "I did not consider the potential benefits to the argument of signing the petition."); and *A and B*. For the purpose of these analyses, two participants who chose A, B, and C were classified as *A and B* and one participant who chose A and C was classified as A. This is because the C option said they had not considered the benefits of signing, which directly contradicts the selection of A and or B (which are both benefits of signing).

Contingency tables and the outcome of Chi-Squared Tests for all inter-question associations are displayed in Table S2. In brief, no notable associations were observed and there were no “statistically significant” associations (all $p > .05$).

Table S2. Association between question 8 and other questions 1-7, 9, and 10.

		Question 1				
Response option		A	B	C	D	
Question 8	AB	46	0	3	13	
	A	99	4	8	22	
	B	8	1	0	4	
	C	8	1	0	3	
$\chi^2(9) = 7.88, p = 0.52$						
		Question 2				
Response option		A	B	C	D	
Question 8	AB	45	12	6	0	
	A	116	17	5	4	
	B	11	2	0	0	
	C	10	2	1	0	
$\chi^2(9) = 8.58, p = 0.45$						
		Question 3				
Response option		A	B	C	D	E
Question 8	AB	13	27	21	0	2

A	14	54	64	3	2
B	5	3	5	0	0
C	3	6	2	0	1

$$\chi^2(12) = 6.40, p = 0.10$$

		Question 4			
Response option		A	B	C	D
Question 8	AB	2	11	28	23
	A	3	34	52	52
	B	2	3	4	4
	C	1	3	2	5

$$\chi^2(9) = 2.01, p = 0.57$$

		Question 5			
Response option		A	B	C	D
Question 8	AB	0	3	20	40
	A	2	14	44	81
	B	1	0	3	9
	C	0	2	3	7

$$\chi^2(9) = 1.93, p = 0.59$$

		Question 6				
Response option		A	B	C	D	E
Question 8	AB	9	24	23	8	0
	A	15	33	66	20	6

B	3	3	5	2	0
C	2	3	7	0	1

$$\chi^2(12) = 5.27, p = 0.16$$

		Question 7		
Response option		A	B	C
Question 8	AB	8	41	13
	A	13	99	25
	B	4	8	0
	C	1	8	3

$$\chi^2(6) = 7.14, p = 0.07$$

		Question 9					
Response option		A	B	C	D	E	F
Question 8	AB	9	17	17	12	4	3
	A	16	28	42	27	18	3
	B	2	5	0	4	2	0
	C	0	3	4	1	0	2

$$\chi^2(15) = 1.74, p = 0.64$$

		Question 10			
Response option		A	B	C	D
Question 8	AB	28	31	1	1
	A	52	75	3	4
	B	5	7	0	1

C	5	4	0	0
---	---	---	---	---

$$\chi^2(9) = 3.57, p = 0.93$$

Table S3. Association between question 10 and other questions 1-7 and 9.

		Question 1			
Response option		A	B	C	D
Question 10	A	59	3	7	21
	B	92	2	5	17
	C	1	0	0	2
	D	4	0	0	2

$$\chi^2(9) = 10.48, p = 0.29$$

		Question 2			
Response option		A	B	C	D
Question 10	A	69	14	8	2
	B	99	18	2	2
	C	2	1	2	0
	D	6	0	0	0

$$\chi^2(9) = 19.48, p = 0.05$$

		Question 3				
Response option		A	B	C	D	E

Question 10	A	18	33	36	0	3
	B	17	49	51	2	1
	C	1	1	2	0	1
	D	1	1	3	1	0

$$\chi^2(12) = 2.54, p = 0.46$$

	Response option	Question 4			
		A	B	C	D
Question 10	A	6	19	25	43
	B	1	30	55	36
	C	0	1	1	3
	D	1	1	1	3

$$\chi^2(9) = 1.42, p = 0.73$$

	Response option	Question 5			
		A	B	C	D
Question 10	A	1	10	20	62
	B	1	7	49	64
	C	0	1	1	3
	D	1	1	0	4

$$\chi^2(9) = 2.02, p = 0.57$$

		Question 6			
--	--	------------	--	--	--

	Response option	A	B	C	D	E
Question 10	A	11	23	40	12	4
	B	14	39	53	15	2
	C	0	3	1	1	0
	D	1	0	2	2	1

$$\chi^2(12) = 3.70, p = 0.30$$

	Response option	Question 7		
		A	B	C
Question 10	A	13	61	15
	B	14	85	23
	C	0	4	1
	D	0	4	2

$$\chi^2(6) = 2.29, p = 0.52$$

	Response option	Question 9					
		A	B	C	D	E	F
Question 10	A	17	20	25	14	6	4
	B	10	33	38	21	17	3
	C	0	0	2	2	1	0
	D	2	0	1	3	0	0

$$\chi^2(15) = 4.36, p = 0.23$$