Clarifying the Meaning of Statistical Collaboration

Musings on why I think increasing statistical collaboration is a good thing

Aaron Caldwell

2020-08-23

Updated on: 2020-08-24

I equate trying to change the reporting practices of statistics with trying to change the direction of an ocean liner with a kayak. Good luck with that. When I mentioned this analogy to a colleague, he said, “What we need are more kayaks. A lot more kayaks.” I understand it is difficult to change entrenched practices. I get that change is slow. But that does not mean we should not try. – Douglas Curran-Everett (2020)

This week a manuscript I co-authored was finally published by the British Journal of Sports Medicine. The point of this opinion piece was to encourage more exercise & sport scientists engage with the statistics literature and, when necessary, seek out collaborators with statistical expertise. As part of our “Call to Action” we highlighted our fields problems with statistics which include mistakes made in published literature and our fields proclivity towards creating novel statistical methods. I hope this manuscript is a rallying call to everyone in the field with in an interest in statistics, but it appears we already have some detractors. In this blog post, I want to address some of what I consider to be “legit criticisms” of our paper and reject some misinterpretations of our opinions.

## Some Background

I think it may help the reader understand the perspectives contained within the paper if they understood more about how and why this paper was written. First, the lead author, Kristin Sainani, organized the writing effort and invited authors to join the paper based on mutual frustrations with the current state of sport and exercise science. This is a fairly diverse group that includes senior academics, actual statisticians, people working in industry/government, and even a graduate student. As one might expect, we all have different perspectives, and often disagree on what constitutes appropriate statistical practice. In fact, this paper took quite some time to finish because of disagreements among the authors on what policies we should recommend, what things we should criticize, and even the tone of our message. In fact, there was a (very) short period of time that I almost had myself removed as co-author. It would be a mistake to assume we are acting as monolithic group vying for control of the scientific literature. We eventually did come to a consensus and we all agreed to the final version that was published this past week. Kristin did a wonderful job organizing this effort (which was probably akin to [hearding cats](https://www.youtube.com/watch?v=m_MaJDK3VNE)), and I am proud to be included as a co-author.

Since there was large amount of disagreement among authors, it is understandable that our colleagues would also have criticisms of our manuscript. Disagreement with our opinion is fine and I personally welcome it. If we are right, it will help sharpen our future arguments for this cause, and, if we are wrong, it would hopefully change our minds. However, what I will not tolerate is bad faith arguments that are intended to distract from the thesis of our paper.

# Legitimate Criticisms

## Our analysis in Figure 2 could be interpreted many ways and has a serious limitation

In the paper, we state only 13.3% of articles we surveyed (k = 299) employed some type of statistical methods expert. From this information, we state there is a shortage of “statisticians” to collaborate on these studies and imply that increasing that proportion would be beneficial to sport and exercise science. However, there are limits to what we can actually say with this data. All the data can really tell us is that the majority of sport and exercise scientists do not collaborate with statisticians in way that merits authorship. This excludes situations where statistcians were consulted but not included on the manuscript, and excludes those with formal statistics training embedded with departments that do sport and exercise science research (more on that below). As we stated in the supplement this analysis is also limited by the criteria we used to count statistical collaborators. On a personal note, many of my own papers would *not* met the criteria we outlined. Now, I do not consider myself a statistician (yet), but a lot of my papers in graduate school did include a statistician (Big thanks to Ronna Turner and Sean Mulvenon). However, in most cases we simply listed our college (College of Education and Health Sciences) of which the Department Educational Statistics and Research Methods was located within. Nonetheless, Kristin Sainani and David Borg checked a subset of 30 articles and found only 1 case where a staff statistician, who was embedded within a non-stats department, was included on a manuscript.

Moreover, I think most of my co-authors would agree that some quantitative studies do not require much statistical expertise in order to be analyzed properly. I am reminded of the time a colleague, who is well trained in statistics, told me “I don’t think there is a single research question or experimental design I am interested in that would require more than a t-test”. While I think such a situation is rare, it does illustrate an important point: sometimes a statistician is not necessary. I am reminded of some of the basic (i.e. animal/cell model) physiologists I have worked with in the past where their experiments are so precise and well-controlled that statistics are rarely necessary. I like to call these “light switch studies” because if the effect the physiologists were studying is real it would be as clear as turning on a light switch in a windowless room. I think the problem is more of knowing *when* to consult a statistician rather than always having a statistician review your work. It is difficult to have the humility to admit when you require an additional assistance. Those with some statistical training, I include myself in this group, should be wary of the [“Beginners Bubble Effect”](https://psycnet.apa.org/record/2017-49272-001), or the tendency to be overconfident in our knowledge/abilities once we have gained only a beginners-level knowledge (Sanchez and Dunning 2018).

Overall, our data cannot prove that sport scientists lack statistics training or that all researchers absolutely need to have a statistician on their papers/projects for their results to be valid.

## One of our examples is not like the others (and maybe unfair to include)

One of the subsections of our opinion was titled “Case Studies: Inventing New Statistics” and, among many examples, we highlighted the case of a preprint titled “On the use and abuse of principal component analysis in biomechanics.” from Dr. Dan Cleather, PhD. Like the other examples, we only gave a brief explanation of the problematic information contained with Dr. Cleather’s preprint. This is unfortantate because, as of now, there are no other public criticisms of his preprint. Therefore, you really have to trust our very brief statement that the information contained within Dr. Cleather’s work is inaccurate or at the very least misleading. However, I am told (via personal correspondance) that Dr. Cleather was privately made aware, on multiple occasions, the statistical issues with his preprint. Nonetheless, it was unfair to include his work as an example since it really hasn’t been cited (therefore isn’t proliferating) and it had not been formally criticized in a public forum (i.e., no one has shown via math and simualtions the faults in his proposed approach). On the other hand, Dr. Cleather doesn’t really outline the mathematics of his position nor does he demonstrate via simulation the superiority of his proposed approach. Therefore, I find his paper lacking the information needed to state that any use of PCA constitutes “use and abuse”. I am told that preprint should be uploaded soon that directly discusses the problem with Dr. Cleather’s position on PCA.

I should also note that I was the SportRxiv moderator that accepted Dr. Cleather’s work onto our server last November. The [insinuation](https://twitter.com/dr_jump_uk/status/1296800260082339840?s=20) that I, or anyone else, would block or otherwise supress his preprints on statistical methods are unfounded considering I’ve accepted his work in the past. We have criteria at SportRxiv, which are [outlined on our website](https://www.sportrxiv.org/submission-guidelines/), for accepting manuscript, and personal disagreements on technical matters should *never* influence our acceptance policies.

## Statistics is not a necessary condition for good science inquiry

On Twitter, Jamie Burr brought up a [good point](https://twitter.com/Dr_Burr/status/1297436782938722305?s=20) that, “statistics are but one tool to help derive confidence in the conclusions to experiments”. For the most part, I agree with this sentiment. I think one omission in our manuscript is that we do not differentiate between quantitative and qualitative research. I think, and I believe most co-authors would agree, that many scientists do not need statistics in order to make scientific discoveries or make scientific advances. There is great value in qualitative, descriptive, or any other variety of work that does not rely upon statistical inference. I want to be clear here and state that such research is not lesser than or deserve less respect than research that is quantitative in nature. In fact, I believe that many exercise and sport scientist may be better off if they ignored statistics, or at least inferential statistics, and simply spent more time describing the phenomena they are studying. Some have even argued that many scientists may be better off sticking to descriptive statistics (Amrhein, Trafimow, and Greenland 2019) to avoid the pitfalls of inference.

# Debunking Bad Criticisms

## TBD

# Miscellaneous notes

**More kayaks**

I started this blog post with a quote from one my role models, Dr. Douglas Curran-Everett PhD. He, like myself, has PhD in Physiology but later shifted his focus to statistics (he is an accredited statistician of the American Statistical Association). He has been instrumental in my education on statistics and his series in *Advances in Physiology Education* called “Explorations in Statistics” were very helpful in my early understanding of statistics. If you are interested in statistics, I highly suggest you read this series. Dr. Curran-Everett exemplifies what I think we need more of in our field: people who are well-trained in the subject matter *and* statistics. I believe that, in addition to increased statistical collaboration, we need more [“stats mavens”](https://thehardestscience.com/2015/02/16/top-10-signs-you-are-a-statistics-maven/) within the field getting in their metaphorical kayaks and pulling us in the right direction.

Sometimes there is going to be disagreement among these stats mavens. Some of my closest collaborators, Andrew Vigotsky and Matt Tenan, disagree on many statistics related topics. That is fine and our debates and disucssion have been extremely useful in my own education. The point is we expect a high level of discourse on this subject matter, and expect each other to be able to justify our opinions through simulations and formal mathematics. The latter portion (simulations and math) is what I believe is missing from the current conversation. If we are to have discussions about statistics in the sport and exercise science literature it must be done on the merits of the proposed statistical techniques, and we should expect those outlining an opinion to come with verifiable evidence (e.g., simulations). If we can have this level of discourse I have no doubt our statistical methods will improve over time.

**We aren’t alone**

As we mention in the opening paragraph of the manuscript, other fields have similar issues with poor statistical practice. A very good example is psychology, and more specifically social psychology (although I am told other sub-fields have similar issues). I would like to highlight the peculiar case of the p-rep ([probability of replication](https://en.wikipedia.org/wiki/P-rep)) that would have been in our “Inventing New Statistics” section if it had occurred in sport and exercise science. Overall, the p-rep was simply a transformation of the p-value obtained from a frequentist statistical test. Like other “novel” statistical techniques that are not evaluated by traditional statistics standards, the p-rep led to an inappropriate interpretation of probabilities and led people to believe their results were more reliable than they were in actuality. Unfortunately, p-rep was given some praise from those in the psych community, and the Association for Psychological Science (a flagship organization in psychology) actually encouraged authors submitting to their journals to report p-rep over p-values. Not only did this lead to the misinterpretations of results, it has made the work of those doing error detection more difficult or even impossible (This is according to James Heathers on the *Everything Hertz* podcast). Luckily, this is was swiftly criticized (see Iverson, Lee, and Wagenmakers (2009) for details) and the Association for Psychological Science soon abandoned their policy on p-rep.

In the past decade psychology has taken many steps to improve their practices. This field is also helped by the fact that psychology has sub-disciplines that *exclusively* focus on quantitative methods (e.g., mathematical psychology). There are journals now, like Meta-Psychology, that just focus on the “Science of Science” in an effort to improve research practices. The fruits of this labor are clear: a robust literature on applied statistical methods for psychologists (see the recent work of [Daniel Lakens](https://journals.sagepub.com/doi/full/10.1177/2515245918770963), [Eric Wagenmakers](https://link.springer.com/article/10.3758/s13423-017-1323-7?wt_mc=Other.Other.8.CON1172.PSBR%20VSI%20Art05%20&%20utm_medium=other%20&%20utm_source=other%20&%20utm_content=2062018%20&%20utm_campaign=8_ago1936_psbr%20vsi%20art05sch), or [Lisa DeBruine](https://psyarxiv.com/xp5cy/)). While this field is far from perfect, and there is a lot of room for improvement, I think there are many actions that our field should emulate.

# Conclusion

If I were to boil our manuscript down into a pithy statement it would be this:

As sport and exercise scientists we should have the humility to recognize our own limitations, realize as a field we have made a lot of statistical mistakes, and not be afraid to lean on statisticians for help when needed.

I explicitly want to state that I do not want some version of “statistics police” telling us what we can and cannot publish. Instead, I think my goal can be summed up by the late Doug Altman [“We need less research, *better research*, and research done for the right reasons”](https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000127/full), and, in order to accomplish the middle goal, at least some of us have to spend more time and attention on how we use statistics.

Lastly, I want to close on some encouraging words. I believe in the power of our field to advance scientific knowledge and improve our understanding of human performance. Despite our collective mistakes, I believe sport and exercise science has made a positive impact and will continue to do so. We should not “throw out” the old, established knowledge because we find statistical mistakes. Instead I also believe many of those who do not understand statistics, with the appropriate training, can develop a statistical expertise. It was only 7 years ago that I was introduced to the basic concepts of statistics, and I too found myself frustrated by this material (and the statisticians who taught it!) Now I find myself empowered by what I have learned, and I am writing [tutorial](https://www.tandfonline.com/doi/full/10.1080/23328940.2019.1624131) [articles](https://osf.io/preprints/sportrxiv/tfx95/) to help my peers better understand statistics. Yet, I do not possess any natural skill in mathematics or statistics, despite my love of it. In fact, I was told to give up on learning math during pre-calc by my high school teacher because I, “just really struggle with this material”. I believe there are many more people who, like myself, can enjoy learning and utilizing quantitative methods. Therefore, I encourage my peers to learn more about statistics, read the statistics literature, and engage in conversations about best statistical practice within our field. In order to accomplish the stated goals of our manuscript, we need more, not less, analysts/statisticians coming from our field.

So, join us in our kayaks and maybe we can pull this ocean liner in the right direction.

## P.S. Rejected titles for this blog post

So, blog post titles tend to have a funny or witty twist. I thought, with the seriousness of the claims being made about our manuscript, that it would be in bad taste for these to headline this blog post. But, I also got a good chuckle from coming up with a few of these so I included some of the titles that didn’t make the cut below. Again, all of these were made in jest and in no way are meant to mean spirited. If you followed some of the Twitter arguments I think you may get a chuckle as well.

**Explicit explanations of my opinions on statistics**

**Lies, Damned Lies, and Lies about my opinions on Statistics**

**Statistics Police: Implicit Intent Unit**

**Statistics Police: Collaborations Victims Unit**

**On the Moral Outrage from Suggesting that Collaboration is a Good Idea**

# References

Amrhein, Valentin, David Trafimow, and Sander Greenland. 2019. “Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis If We Don’t Expect Replication.” *The American Statistician* 73 (sup1): 262–70. <https://doi.org/10.1080/00031305.2018.1543137>.

Curran-Everett, Douglas. 2020. “Evolution in Statistics: P Values, Statistical Significance, Kayaks, and Walking Trees.” *Advances in Physiology Education* 44 (2): 221–24. <https://doi.org/10.1152/advan.00054.2020>.

Iverson, Geoffrey J., Michael D. Lee, and Eric-Jan Wagenmakers. 2009. “P Rep Misestimates the Probability of Replication.” *Psychonomic Bulletin & Review* 16 (2): 424–29. <https://doi.org/10.3758/pbr.16.2.424>.

Sanchez, Carmen, and David Dunning. 2018. “Overconfidence Among Beginners: Is a Little Learning a Dangerous Thing?” *Journal of Personality and Social Psychology* 114 (1): 10–28. <https://doi.org/10.1037/pspa0000102>.