

EDITORIAL

Bad science and how to avoid it, a movement analysis perspective: Study design, Statistics and Publication Ethics

Andy Kerr, Robin Prescott, Tim Theologis

Introduction (Tim Theologis)

Stepping down from the Editor-in-Chief post for *Gait and Posture* after 10 years gave me the opportunity to write a last editorial as an expression of gratitude to all of those who made these years a fantastic experience. I thought that a modest article offering general advice on how to avoid pitfalls when writing papers would be appropriate. I asked for the contribution of two excellent colleagues, who participated in an ESMAC seminar on the same subject in 2013. I hope that the result is informative and helpful. I am very grateful to Dr Andy Kerr and to Professor Robin Prescott for their work on this article.

I would like to take this opportunity to express my deep gratitude to all Readers, Reviewers and Authors of *Gait and Posture* for their contributions to the success of this journal. I am fully aware of the fact that Reviewers dedicate valuable time within a busy schedule in order to provide a service to *Gait and Posture* and maintain the good quality of published articles. I am also aware that Authors choose *Gait and Posture* among a large number of potential target journals and would like to thank them for their confidence in our journal.

I am very grateful to the Editorial Board for the time and effort they dedicate in ensuring that quality is maintained and that the journal continues to reflect the interests and future directions of the Societies it represents. For the last 10 years I have relied heavily on the hard work, dedication and commitment of the Associate Editors' team. Each one of them provides leadership in a specific field of expertise. As a team, I believe we ensured that *Gait and Posture* continued to publish research reflecting the multi-disciplinary character of our journal and remained relevant to both clinicians and scientists. I could not have done it without them and I am very grateful for their support. Finally, I am grateful to our Publishers, Elsevier, for their support and their excellent work, which contributes greatly to the success of the journal.

I leave the journal entirely confident that, under its new leadership, its future is bright. With the continuing support of Authors, Reviewers and the Societies, *Gait and Posture* can only go from strength to strength.

Bad Science and Study Design (Andy Kerr)

Let's be positive, researchers and research groups, do not intend to conduct bad science, at least not at the outset. Nevertheless, we are all aware of alarming and risible examples of bad science. When this work passes through the peer review system to publication the effect not only damages progress in those areas but also erodes trust from the public who, after all, fund our work. In our area of movement analysis we are not immune. Like most communities, bad science should only be encountered by our band of anonymous reviewers and journal editors, but the review process is not perfect, some questionable papers slip will, and do, slip through.

Many factors contribute to bad science being given a full natural life, from conception to dissemination; ignorance, statistical naivety, money (*both a lack and possibly too much*) and technology (*having the kit and no wit*). At the ESMAC conference in Glasgow (2013) we organised a seminar aimed at shining a light on these factors, in the hope that we can all learn by them, much like the Raspberry awards in Hollywood try to do for the film making industry.

As every conference presenter knows an audience needs a simple take home message. To round off our seminar we generated a list of top tips from all the presenters which we presented to the audience with the ambition of reducing this list to a top 5 through debate and discussion. We began with 11 contributions: 1) "Understand any assumptions in the methods you are using"; 2) "Be clear in your presentation"; 3) "Don't try to hide 'bad news' about your research"; 4) "Have clear objectives and pre-specified hypotheses"; 5) "Put a lot of effort into designing your study"; 6) "Make a genuine a priori hypothesis for any research that you do"; 7) "Don't report too many results in the results section!" 8) "Write statements in your Discussion that make a tingle go up and down your spine and that you would be proud to read in 20 years' time"; 9) "have a clear question"; 10) "demonstrate the need to answer the question"; and 11) describe clearly the method

The short period of lively discussion revised this list down to a top 5, ranked in order of popularity.

1. Have a clear question.
2. Understand any assumptions you are using and describe the methods clearly.
3. Put a lot of effort into designing your study.
4. Don't try to hide 'bad news' about your research.
5. Demonstrate the need to answer the question.

We consider this a valuable take home message.

Using Statistics to come to the wrong conclusion (Robin J Prescott)

One definition of statistical methods is those that are used to elucidate data that are affected by a multitude of causes. This aim of elucidation is often unappreciated and an alternative

aim of maximising the number of ‘statistically significant’ p-values often appears to be an objective of some authors. Thus we might see “patients walked significantly slower in the slow speed trials (mean 0.81 m/s, SD=0.11) than in the comfortable speed trials (1.03 m/s, SD=0.14; $t(9)=8.46$, $p<0.001$)”. The results are very clearly presented but there is no sensible Null hypothesis to test. The hypothesis that speed is the same in slow speed trials and comfortable speed trials is untenable irrespective of the data.

Are statistical errors common? There have been multiple surveys over the past 40 years to suggest that the published medical literature contains many errors, and that the wrong conclusions can be drawn. Errors in the manuscripts submitted to journals will be even more common and a recent survey of 100 papers submitted to Injury and subject to statistical review led to rejection or changes in 90. Some aspect of the analysis needed to be changed in 47 and errors were found in 45. Suggestions for an improved presentation were made in 73.

The problems that arise cover a wide spectrum and the following topics are only illustrative. Multiple testing can easily lead to errors in interpretation of results. In totally uninformative datasets, from its definition, 5% of significance tests can be expected to show statistical significance at the 5% level. Thus when many tests are conducted, some of the ‘positive’ findings are likely to be false positives. In a recently published paper, one of the tables showed tests applied to 12 variables. Eleven were shown as ‘NS’ (bad practice in itself) while one gave $p=0.017$. The corresponding text highlighted this ‘significant’ finding which, in reality, is highly likely to be due to chance. Interpretation of findings after many tests must always be cautious, which is why a single primary outcome measure is encouraged in clinical trials.

Misleading conclusions can also be drawn after the application of data driven hypotheses. There is a logical fallacy in using data to derive a hypothesis and then using the same data to test the hypothesis. This type of problem occurs in a range of forms. A simple one is finding a predictive cut-point in a variable such as age or some clinical variable from the data and then showing its effectiveness on the same data. This extends into classification trees and clinical decision rules. These methods can be very valuable but it is essential that they are evaluated on an independent dataset or the evaluation will be severely biased.

It seems that many of the standard statistical tests are applied without an understanding of their assumptions. An assumption that is often overlooked is that the observations should be independent of each other. A common reason for this to be violated is when we have bilateral data. Ignoring the pairing and treating every observation separately is invalid and leads to overly optimistic p-values and confidence intervals. Using the average of left and right is

valid, as is a repeated measures analysis of variance, but will be inefficient if there are any missing observations. Mixed or multi-level models allow an optimal analysis of such data.

Sub-group analysis is another minefield for the unwary. As well as examining overall differences between groups (often treatment groups), it is tempting to repeat the analysis for sub-groups, say for men and women. The fallacy is to test each group separately when the question that should be posed first is whether there is evidence that men and women differ from each other in the effect of treatment. In statistical terminology this is testing for a treatment by gender interaction. Only if this is significant or there is a priori information to suggest such an interaction, should sub-groups be analysed separately.

Last, but certainly not least, it is important to appreciate the reasons why tests may be statistically significant or non-significant and interpret our findings accordingly. We may achieve $p < 0.05$, because there is good evidence of a genuine departure from the Null hypothesis: e.g. there may be a genuine treatment difference. On the other hand it could be due to chance, there might be bias in our comparisons, or we might have applied an inappropriate test. Conversely, a non-significant difference might occur because the Null hypothesis is true or it might be due to chance, inappropriate analysis or, very commonly, because the study is too small to have a good chance of detecting an important difference.

Statistical methods should be applied with thought given to their initial use and to the interpretation of the findings. Many statisticians would argue that statistical thought has most benefit at the stage of designing a study, so that important differences or associations have a high probability of being shown as statistically significant. Statistics should never be thought of as black box technique to add respectability to a publication. The priority should be to conduct a well-designed study and present the results as fully and clearly as space permits, using statistical methods to enhance understanding.

Publication Ethics (Tim Theologis)

Publication of scientific research is governed by legal and ethical rules. The aim of those rules is to ascertain that the published material is the genuine product of *bona fide* research based on ethical standards and selected for publication through a fair peer review process. The rules that govern publication ethics are complex as the boundaries of ethical acceptability are often unclear. This section touches upon some of the most commonly encountered issues related to publication ethics but does not comprehensively cover this enormous and complicated subject. Additional information and help for authors, reviewers and editors can be found on the website of the Committee of Publication Ethics (COPE) at <http://publicationethics.org/>.

Publication bias is a bias with regard to what is likely to be published, among what is available to be published. Publication bias occurs when the publication of research results depends on their nature and direction rather than the quality and relevance of the research. A common form of publication bias concerns publications presenting “negative” results, which confirm the null hypothesis or, in other words, they do not prove the hypothesis on which the research question is based. It is true that journals often show a preference in publishing papers with positive findings. Researchers may also decide not to submit papers which confirm the null hypothesis because they perceive a lack of interest on the readers’ behalf or because they have lost interest and motivation. Therefore, negative research findings often remain unpublished. This leads to significant bias when systematic reviews and meta-analyses are conducted. The positive findings of multiple publications are not counter-balanced by the negative findings of unpublished work.

Adequately powered studies with appropriate assessment of the probability for a false positive result can overcome this difficulty as they usually convince peer reviewers of the genuine result. Registration of research study protocols and data collection with the relevant clinical or academic authority would also prevent studies leading to negative results from remaining unfinished or un-reported.

Another form of publication bias has been termed “*HARKing*” or Hypothesising After the Results are Known [1]. Good quality research should include an *a priori* hypothesis or research question. Collecting data and undertaking multiple comparisons (as it is often the case with gait analysis data) without a pre-defined hypothesis can lead to false positive results and an often complicated and improbable interpretation of those in the discussion.

Redundant publication or duplication is the publication of work based on previously published datasets and findings by the same team of researchers. In cases of major redundancy, the content of the manuscript is similar to the already published material. Attempts to hide the redundancy by changing the title or the authors’ order and by failing to refer to the previous publication usually indicate intentional duplication. This constitutes a serious misconduct which contradicts the authorship statement required by most publishers. If the attempt to publish redundant material is substantiated, the editors are obliged to report the attempt to the authors’ institution and to reject or retract the manuscript. More often, the boundaries of redundant publication are less clear. Publishing multiple papers from subsections of the same research project, based on a single dataset obtained from the same experiment or same sample, is common practice. It is usually left to the editors to decide what constitutes duplication. If different papers address a different audience and raise different points in the discussion, present an extended follow-up or are written in a different language, they are usually considered acceptable. The editors, however, have to be satisfied that the manuscript under consideration constitutes original work as a stand-alone document.

Plagiarism is the unattributed use of large portions of text or data presented as if they were produced by the plagiarist. This also represents serious misconduct contradicting the authors’ statement submitted to the publishers. If the editors are convinced of the plagiarism they are obliged to report it to the authors’ institution and to reject or retract the manuscript. The same

principles apply to **Data Fabrication**. An increasing number of scientific journals request submission of the raw data by the authors in order to prevent this misconduct.

Authorship disputes are unfortunately common in scientific publication. Defining clear roles and ownerships at the start of a research project should prevent such disputes. Most journals require written confirmation of all authors in order to change the authorship of a paper. They also require written explanation of the reasons for the change. Editors and reviewers should always be aware of Gift authorship (the inclusion of an author who did not contribute to the work) or Ghost authorship (the exclusion of an individual who contributed significantly to the project). The boundaries of what is ethical in this practice are often ill-defined and reaching a judgement in some cases of authorship dispute can be challenging.

The World Association of Medical Editors (WAME, www.wame.org) has defined as **Conflict of Interest (CI)** the “divergence between an individual’s private interests and his or her responsibilities to scientific and publishing activities such that a reasonable observer might wonder if the individual’s behavior or judgment was motivated by considerations of his or her competing interests”. The matter of CI concerns authors, reviewers and editors and is an important ethical issue to consider in scientific publication. Financial ties in the form of funds, patents, stocks, gifts or services relating to a specific research project have to be declared and considered. Financial CI should also be considered when there are ties to the industry or commercial bodies and include payment for research, ownership of stock and stock options, honoraria for advice, public speaking or consultation, service on advisory boards, service on scientific/medical education companies, receipt of patents or patents pending or holding a post funded by the industry. However, CI is not always financial. Academic CI may exist when individuals have strong beliefs in a particular research domain or belong to a competing team or institution. Personal CI may exist when family and close friends, competitors or current/previous colleagues are involved in the process of peer review and publication. Political or religious CI should also be considered. Most scientific journals expect authors to declare CI on submission of an article. Whether or not the review and publication of an article can proceed despite a CI or not is left to the discretion of the editors. Reviewers, editors and the readers should exercise their judgement on the credibility of scientific articles taking into account the declared CI. Finally, editors are expected to consider CI, not only when assessing routinely submitted work but – particularly- when dealing with their own research or work submitted from their own institution.

Good quality research and the scientific publication process should be based on solid ethical principles. Researchers, reviewers and editors should ensure that ethical questions and dilemmas are fully addressed before an article is submitted and published. This will protect the reader from misinformation and potential harm. As stated above, additional information and help for authors, reviewers and editors can be found on the website of the Committee of Publication Ethics (COPE) at <http://publicationethics.org/> . Two interesting as well as entertaining articles referenced below provide further guidance on conducting scientific research and on writing an original article [2,3].

REFERENCES

1. Kerr N.L. (1998). "HARKing: Hypothesizing After the Results are Known". *Personality and Social Psychology Review* **2**(3): 196–217
2. Linton JD (2012). Editorial: How to get your papers rejected (or not). *Technovation*; 32: 6-8
3. Isaacs D, Fitzgerald D (1999). Seven alternatives to evidence based medicine. *British Medical Journal*; 319: 1618