2011

Balanced decisions and unbiased reporting in bioscience and medicine

Merrill F. Elias
University of Maine, mfelias@maine.edu

Adam Davey
Temple University

Amanda L. Goodell
University of Maine

Balanced decisions and unbiased reporting in bioscience and medicine

inaugural editorial

Merrill F. Elias[1,2], Adam Davey[3], Amanda Goodell[2] (1. Graduate School of Biomedical Sciences, University of Maine, USA, 2. Department of Psychology, University of Maine, USA, 3. Department of Public Health, Temple University, USA.)

Receive date: September 21st, 2011

Merrill F. Elias (PhD, MPH) is an editor of *Journal of Bioscience and Medicine* (JBM). His training is in research psychology and cardiovascular epidemiology. His work focuses on cardiovascular risk factors in relation to cognitive functioning and dementia. Risk factors he has studied include hypertension, plasma homocysteine, ApoE-e4 genotype, diabetes mellitus, left ventricular mass, obesity, central adiposity, atrial fibrillation, angina pectoris, nutrition, and depressed mood. His current interest is in relating pulse wave velocity (a measure of arterial stiffness) to cognition and relations between mild renal dysfunction and cognition. His studies, and those of my his, are done in a life-span developmental context and involve cross-sectional differences and longitudinal designs. Much of the data for his work, and the work of his students, comes from their own Maine-Syracuse Longitudinal Study (begun in 1975), the Framingham Heart Study and Framingham Offspring Study. His teaching interests are experimental design, biostatistics and epidemiological approaches to the study of cognitive functioning.

Contact information: University of Maine, 5742 Little Hall, Orono, ME 04469-5742; Phone: (207) 581-2097, Fax: (207) 581-6128; Email: mfeliasumaine@aol.com

In this editorial we comment on editorial policies of the *Journal of Bioscience and Medicine* (JBM) in relation to advancing the knowledge base in bioscience and medicine. These include a rapid review process, a free open publication format, an emphasis on "building block" research as well as novel studies, no limitation in word count, open discussion and active distribution of papers to a relevant audience of scientific colleagues; and most importantly, publication of both positive and negative results papers. We argue that given the role probability in science and self-correcting nature of scientific inquiry, the scientific community must have access to the full range of evidence for and against specific hypotheses and associations; from this perspective, negative results in papers meeting quality standards should be treated no differently than positive results. Moreover, we suggest specific guidelines for presenting negative results, reviewing negative results, and most importantly, deciding when to submit negative results to peer review. Among these guidelines is the GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) approach to evaluating the quality of the research and presentation of data in formats reasonable for interpretation of negative and positive finding. More generally we argue that one of the most significant reasons that negative results are not published is the reluctance of investigators to publish negative results. We suggest that positive decisions to submit negative results data may be more likely for journals such as JBM that publish both positive or negative results in a balanced fashion rather than focusing on either negative and positive results papers.

The *Journal of Bioscience and Medicine* (JBM) will publish its inaugural issue in winter of 2011. There are a number of editorial policies that make this journal an attractive publication vehicle for researchers in the biological, epidemiological, bio-behavioral, and medical sciences. We review each of these in the sections that follow and comment on their scientific merits. Finally, we provide a detailed commentary on the issues involved in reviewing null results papers and in making a decision to publish null results.

Quick Review Process.

Quick Review Process.

The goal is a manuscript turnaround time of four weeks from submission. This goal is within the time envelope developed by the major journals in bioscience and medicine. The "proof of the pudding is in the tasting" and thus the editors' success in achieving this goal will be very important to the success of the journal. A number of documents on the ethics and considerations in review of scientific papers are available and useful, and include issues such as confidentiality, conflict of interest, objectivity, expertise, and timeliness [1].

No Publication Costs.

No Publication Costs.

This is a very important element in the success of the journal and balanced science at a time when funding for research is decreasing and grant funding for publication fees is exhibiting a parallel decrease. It is important that science not be represented solely by those investigators who have resources to pay for publication fees. Moreover, there are few researchers, at least in the older generation of scientists, who do not recall the stigma of the "pay journal." Clearly that stigma may not be deserved, as even high impact journals now charge for extra pages using a very conservative figure for standard journal length.

There are experimental data indicating that articles made free-to-the reader via the open access journals were cited more often [2, 3]. It is likely that this trend will continue and increase demand for open access journals for which there is no publication charge.

Novelty is Not a Requirement for an Acceptable Paper.

Novelty is Not a Requirement for an Acceptable Paper.

Novelty is clearly the hallmark of exciting new science, but it requires subjective judgment on the part of the reviewer and precludes publication of failures to replicate or applications of previous findings to new problems. After all, buildings are supported by many more bricks than cornerstones, and there is growing recognition of the important role of negative results in clinical trial reporting [4] and negative results in the general research literature [5].

No Limitation on Length.

No Limitation on Length.

This is an attractive feature at a time when many print publication journals are forced to limit word count, and utilize supplements and references to previous papers with the methodological details. While these are reasonable solutions, they force readers with significant interest in a paper to make the extra effort to track previous papers and access supplementary materials. Moreover, there are few authors who have not been frustrated by peer review advice as to reduce the word count while simultaneously to add information on results, methods, and elaboration of discussion points. Obviously the researcher who cannot practice reasonable constraint will produce a product that will not be well accepted by reviewers or readers, but many authors are likely to welcome the opportunity to match the length of their papers to the scope and the detail required to communicate their research accurately and effectively.

Open Discussion.

Open Discussion.

This is a particularly attractive feature of the journal, i.e., an option that allows reviewers to publish their comments in the journal. Here again good ethics and constraint are important as reviewers must refrain from "pulling their punches" during the review process to save material for commentary. Allowing rebuttal by authors serves to prevent this use of the review commentary and initiates a very exciting flow of scientific discourse which was alive and well in all sciences at the turn of the century, but has faded in parallel to constraints on length and increase in publication cost for print journals.

Distribution of Papers by JBM.

The journal will actively distribute published articles to potentially interested readers. Most scientists realize that public relations considerations are part of the scientific process, but few have the time to circulate papers to all potentially interested colleagues. Most will welcome this feature and, as scientists, we welcome receipt of papers relevant to our work.

Negative Results.

Negative Results are Acceptable for Publication.

This is one of the most important features of science and thus one of the major positive features of JBM. However, it is important to have knowledge of the important technical issues with regard to the publication of null findings. Thus, we will expand our comments on the merits of this negative results publication policy by providing a review of the important reasons why publication of negative results is so critical to good sciences.

Why Are Negative Results Important?

As most all of us have been taught in methods courses: "Most broadly, findings from good research are the subject matter of science, regardless of whether the results support particular hypotheses" [6, p. 742]. In a short essay on this topic, Probst [6], citing Dickersin [7], points out that the systematic failure of journals to publish articles with null results leads to a disproportionate representation in the literature on positive results (most often defined as statistically significant results). She [6] makes the important point that statistical methods for meta-analyses have been developed to compensate for the under-representation of null findings [8] because this problem is so widely recognized. In his book "Hyping Health Risks," epidemiologist G.C. Kabat [9] makes excellent points about the roles of societal beliefs and preferences and biases internal and external to medicine, including political agendas, which encourage the selective reporting of negative and positive results.

Giraud-Carrier and Dunham [10] make the important point that failure experiences need to be accessible to fellow researchers who may otherwise spend needless time on a project or a hypothesis, or who could build from those results with a new methodological approach. Probst [6] argues that ethical considerations in human research require that the research findings, negative or positive, be communicated and that this is most important in clinical trials when an intervention is not successful. In the context of educational policy this author [6] points out that an apparently sound but ineffective intervention, if not contradicted via publication of negative results, might be employed for generations of students.

Magnitude of effect is an important consideration even when the null hypothesis is rejected. Despite clinically significant results, investigators are obligated to comment on the clinical and/or epidemiological significance of the findings and to be very clear where effect sizes or practical importance of the results are minimal. The confusion between statistical significance and significance in terms of effect size continues in the literature despite arguments against strict adherence to p-value rules as indicators of importance of a research finding [11].

Journal policy, reviewer and editor bias, indirect pressures related to funding, and self-censorship are some reasons offered for this trend to not publish null results [6], although a recent report from the CONSORT group [12] suggests that rejection by a journal was much less commonly mentioned by researchers as a reason for failing to publish negative findings. Thus, selective submission may be as important in creating bias toward positive findings as selective recommendations through peer review or selective acceptance by

editors. As a solution to bias toward positive results the *Journal of Cerebral Blood Flow and Metabolism* has introduced a negative result section [13], and there are journals that are dedicated to negative results, including the *Journal of Negative Results in Biomedicine* [14], the *Journal of Negative Results* [15], and the *Journal of Interesting Negative Results* [16]. While this is a very positive trend in publication of scientific information, investigators may be much more willing to submit negative result findings to a journal that accepts papers of both varieties as the imbalanced presentation of the negative results journal may suggest that negative results papers are of intrinsically less value and not worthy of mainstream publication. In this context we note that the *Journal of Interesting Negative Results*, although devoted to important negative results papers, was started in April 2008 and appears to have published only 1 paper [16]. The *Journal of Negative Results in Biomedicine* has a better record as it has published approximately 96 articles, including commentaries on articles, between 2002 and 2011 [14].

It is difficult to take issue with the fact that negative results are important and it is important that JBM is accepting papers with negative results. Just as for positive results, one must define which negative results papers are acceptable and worthy of publication. In the sections below, we argue that the decision should be based on the quality of the methods, not the result obtained.

***Negative Results*.**

Evaluating a Negative Result Paper.

There has been a long debate on null hypothesis significance testing (NHST) and effect size testing (EST) for generating a cumulative knowledge base in the behavioral sciences, and it has now run its course several times. Issues include the importance of considering both parameter estimates and confidence intervals [17], issues necessary for training scholars [18] and the importance of model building and model testing [19] and effect size testing [20]. What is clear in this process is that evidence differs in the quality of both inference and recommendations.

Journals which explicitly encourage or permit submission of negative findings potentially eliminate or reduce the potential bias of selective submission. Some journals further distinguish between studies which present negative results and those which are methodologically sound but under-powered (e.g., pilot studies designed to bracket effect sizes rather than to test group difference treatment effects). Hurdles that remain, then, pertain to evaluating the scientific rigor of a manuscript based upon the quality and integrity of a particular study. In recent years, epidemiologists have worked to develop consensus on how to eval-

uate research studies.

One large-scale example of this is the GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) approach [21-25]. An envisioned 20-part article series in the *Journal of Clinical Epidemiology* [26], the first nine installments of which are already in print or in press, considers guidance for researchers to evaluate the quality of evidence and strength of recommendations of research in health care. Although the overall framework is oriented toward research synthesis in the form of systematic reviews, many of the guidelines it provides are also useful for evaluating the quality of evidence within a specific study, and offer an explicit and structured approach to summarizing evidence. Reviewers of negative findings should consider the confidence that effect estimates are correct alongside confidence that effect estimates are adequate to support clinical decisions and recommendations.

Central to consideration of studies presenting negative findings is sensitivity to the precision of estimates, such as evidenced by 95% confidence intervals. Depending on goals and perspective, reviewers may wish to consider the accuracy in parameter estimation (AIPE) [27] and/or optimal information size (OIS) [28] associated with a particular study. Reviewers should be sensitive not just to evidence for or against specific hypotheses, but rather what new information is gained about the likely range within which a specific effect is likely to reside. In testing statistical hypotheses, what is obtained is the probability of the data given that the null hypothesis is true. What researchers really want to know, however, is the probability that the null hypothesis is true given the data. This latter quantity, the inverse probability, is only obtainable through Bayes' theorem [29, 30], a set of statistical rules requiring assumptions about the prior probability or distribution of probabilities. However, one need not master Bayes' theorem in order to make common-sense decisions about the quality of the methods in a negative results study or to present the data in the best possible way for clear interpretation of their importance.

Summary

Summary

Many researchers contend that a study is never completed until it has been published, and this is a valid argument. It is likely that an investigator or investigative team may be the most important single factor behind the lack of publication of null or negative findings, rather than journal editors or reviewers. We argue that research should be evaluated on its intrinsic merits, i.e., the quality of the methods, procedures, and analyses used, rather than on the nature of its results.

The use of probability in science and the self-correcting nature of scientific inquiry depends on a research community that has access to the full range of evidence for and against specific hypotheses and associations; from this perspective negative results should be treated no differently than positive results [31]. In evaluating research, particularly with negative or null findings, approaches beyond null hypothesis significance testing are essential. Confidence intervals provide very important information about the likely range within which an effect or difference is likely to be found, and there are many examples of effective presentation of data using this vehicle [23, 32, 33]. From this perspective, studies with larger sample sizes more successfully bracket these effects than studies with smaller sample sizes. In many contexts, an emphasis on accuracy in parameter estimation can be a useful extension to confidence intervals.

Many different frameworks can be applied to evaluate the soundness and rigor of scientific research, regardless of their findings. In this paper, we provide links to the emerging GRADE framework as one example. The GRADE framework is broad and systematic in its scope, and written with a clinical research audience in mind. Within the GRADE framework, imprecision is considered as just one of several components in evaluating the quality of the evidence a study provides. As more journals, editors, and reviewers provide positive experiences for researchers interested in publishing null or negative findings, perhaps this barrier to generating a more cumulative scientific knowledge base will diminish. One necessary first step in this important objective is a journal policy that encourages submission of both positive and negative results studies. Quality review with rapid turnaround, no publication charges, no limitations on length, and distribution of papers to a potential interesting audience are additional features that will make the *Journal of Bioscience and Medicine* an important vehicle for achieving this goal.

REFERENCES

1. Rockewell, S. Ethics of peer review: A guide for manuscript reviewers. The Interntional Congress on Biomedical Review and Global Communications on Advisory Board, Congress on Biomedical Peer Review. http://ori.hhs.gov/education/products/yale/. Accessed September 18, 2011.
2. Craig ID, Plume AM, McVeigh ME, Pringle J, Amin M. Do open access articles have greater citation impact?: A critical review of the literature. *J Informetrics.* 2007; 1:239-248.
3. Davis PM, Lewenstein BV, Simon DH, Booth JG, Connolly MJ. Open access publishing, article downloads, and citations: Randomised controlled trial. *British Med J.* 2008; 337:a568.
4. Cals JW, van Amelsvoort LG, Kotz D, Spigt MG. CONSORT 2010 Statement-unfinished update? *J Clin Epidemiol.* 2011; 64:579-582.

5. Gupta N, Stopfer M. Negative results need airing too. *Nature*. 2011; 470:39.
6. Probst JC. Prisoners' dilemma: The importance of negative results. *Fam Med*. 2006; 38:742-743.
7. Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev*. 1997; 9:15-21.
8. Pham B, Platt R, McAuley L, Klassen TP, Moher, D. Is there a "best" way to detect and minimize publication bias? An empirical evaluation. *Eval Health Prof*. 2001; 24:109-125.
9. Kabat GC. *Hyping Health Risks: Environmental Hazards in Daily Life and the Science of Epidemiology*. New York, NY: Columbia University Press, 1993.
10. Giraraud-Carrier C, Dunham MH. On the importnce of sharing negative results. *SIGKDD Explorations Newsletter*. 2010; 12:3-4.
11. Ziliak ST, McCloskey DN. *The cult of statistical significance: How the standard error costs us jobs, justice and lives*. Michigan: University of Michigan Press, 2008.
12. Stern JM, Simes RJ. Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Med J.* 1997; 315:640-645.
13. Dirnagl U, Lauritzen, M. Fighting publication bias: Introducing the negative results section. *J Cereb Blood Flow Metab*. 2010; 30:1263-1264.
14. http://www.jnrbm.com/. Accessed September 19, 2011.
15. http://www.jnr-eeb.org/index.php/jnr. Accessed September 19, 2011.
16. http://jinr.site.uottawa.ca/. Accessed September 19, 2011.
17. Harlow LL, Mulaik SA, Steiger JH. *What if There Were no Significance Tests?* Mahwah, NJ: Lawrence Erlbaum Associates, 1997.
18. Schmidt FL. Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychol Meth.* 1996; 1:115-129.
19. Rodgers JL. The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *Am Psychol.* 2010; 65:1-12.
20. Cortina JM, Landis RS. The Earth Is Not Round (p = .00). *Org Res Meth.* 2011; 14:332-349.
21. GRADE Working Group. http://www.gradeworkinggroup.org/. Accessed September 19, 2011.
22. Guyatt GH, *et al.* GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Med J.* 2008; 336:924-926.
23. Guyatt G, *et al.* GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011; 64:383-394.
24. Guyatt G, *et al.* GRADE guidelines 6. Rating the quality of evidence-imprecision. *J Clin Epidemiol*, In Press, Corrected Proof.
25. Balshem H, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011; 64:401-406.
26. http://www.elsevier.com/wps/find/journaldescription.cws_home/525472/description. Accessed September 19, 2011.
27. Maxwell SE, Kelley K, Rausch JR. Sample size planning for statistical power and accuracy in parameter estimation. *Annu Rev Psychol.* 2008; 59:537-563.
28. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: Utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials.* 1997; 18:580-593.
29. Statistics Tutorial: Bayes' Theorem (aka, Bayes' Rule). http://stattrek.com/lesson1/bayes.aspx. Accessed September 18, 2011.
30. Yudkowsky ES. An intuitive explanation of Bayes' Theorem. http://yudkowsky.net/rational/bayes. Updated June 6, 2006. Accessed September 18, 2011.

31. Young NS, Ioannidis JP, Al-Ubaydli O. Why current publication practices may distort science. *PLoS Med.* 2008; 5:e201.
32. Brown NJ, Byiers S, Carr D, Maldonado M, Warner BA. Dipeptidyl peptidase-IV inhibitor use associated with increased risk of ACE inhibitor-associated angioedema. *Hypertension*. 2009; 54:516-523.
33. Aberson C. Interpreting null results: Improving presentation and conclusions with confidence intervals. *J Artic Support Null Hypothesis.* 2002; 1:36-42.