Available online at www.sciencedirect.com
ScienceDirect
Clinical Biochemistry 39 (2006) 1109–1111
CLINICAL BIOCHEMISTRY

Editorial

Quality of the scientific literature: All that glitters is not gold

One of the pleasures of our professional endeavors is reading exciting scientific papers. The scientific literature is a printed archive of ideas, experiments and reviews prepared by a myriad of scientists and includes great, good and not so good pieces. Here, I will try to make a point that not everything published represents valid work and the conclusions of a published paper may not necessarily be correct, or they could even be misleading.

In this essay, I do not intend to complain about the scientific literature. I will only attempt to identify or reiterate on some of its well-known problems. My students challenge me with the notion that although they know that there is good and bad literature, the most prestigious scientific journals (which can now be ranked by various "factors" such as the impact factor [1] and have very rigorous peer review systems) must publish papers that should be believable and represent major advances in the field. I respond with the lyrics of a famous Greek song: "the family does not make a man, but the man makes the family." The actual value of a scientific paper is its content, nothing else. One could argue that a paper should be of the highest quality if it is published by a great scientist or in a high-impact and highly prestigious journal, if it comes from a highly prestigious institution and if it passed a very rigorous scientific review. None of this is necessarily true. It is well-known that even the brightest of minds, including Einstein, published blunders. One of the most versatile minds of the 20th century, Linus Pauling, a double Nobel Prize winner, thought that he found a cure for cancer in the mid-70s, through administration of mega doses of vitamin C, a proposal that was finally shown to be wrong [2,3]. A few years ago, we became aware of a series of publications in *Nature* and *Science* from physicist H. Schon, which were proven to be fabricated [4]. More recently, two highly promising papers by Woo Suk Hwang on stem cells were found to be fraudulent [5,6]. Now, we have the Qiu Xiao-Quing case [7] and the Taira case [8]. Astonishingly, a Norwegian researcher recently admitted that a high-profile paper he published in *Lancet* was fiction [9]. One can identify a long list of papers that have been published by either prominent scientists or in the highest impact journals, which proved in the end to be invalid or fraudulent. Careful monitoring of retractions gives the impression that this activity is increasing, especially for papers in the highest impact journals.

In the following paragraphs, I will briefly summarize some limitations and threats to the current scientific literature:

What is published is not tested

With the current system, which is unlikely to change, papers published are not reproduced beforehand. This leaves the door open for publication of data that may not be reproducible, have been fabricated or they are false due to bad experimental design, biases or wrong interpretations. Recently, after the Hwang scandal, *Nature* and other journals wondered whether some seemingly influential and highly promising papers should be duplicated before publication [10]. This is clearly not a practical option because it requires money, time and skill which may not be readily available. Thus, the practice of publishing data without prior duplication will continue in the future, along with the associated dangers.

Data fabrication

In my view, data fabrication represents one of the least threatening challenges to the scientific literature. Those who fabricate data will eventually be caught and severely punished. It is like attempting to rob a bank. Very few try it, although "that's where they money is," and even fewer escape arrest and punishment. Despite the recent indications that data fabrication is on the rise, the alternative explanation is that it was always there but the tools for catching the offenders were not as sophisticated as they are now. Ironically, a scientist was recently caught fabricating his curriculum vitae! [11].

Experimental design

Especially in medical sciences, appropriate experimental design is paramount, in order for a study to produce meaningful and reproducible data. Proper experimental design is not easy [12]. It requires expertise and understanding of all the factors that may lead to false conclusions and these should be controlled as much as possible. Epidemiologists and Biostatisticians are specifically trained to assist others in designing experiments in which one parameter is compared between groups, while all other parameters (confounders) are non-contributory. I do not know how many published papers are out there with false

doi:10.1016/j.clinbiochem.2006.08.015

conclusions due to design biases, but I would guess that there may be way too many.

Relevance of in vitro data

A large amount of medical research is being conducted with cell lines isolated from either humans or animals. While these systems are highly valuable in some cases, and they have led to significant discoveries, numerous papers are also reporting data that do not necessarily represent the actual physiological or pathobiological situation of a human being. Cell lines change with time and they frequently have abnormal chromosome numbers, contain translocations, mutations, etc. These parameters may introduce a bias which could make the results irrelevant to human physiology and pathobiology.

Sometimes, it is astonishing on how much information is generated on molecules that appear to be "in fashion". A classical example is the putative tumor suppressor gene, p53, for which a Medline search identified, until now, more than 20,000 published papers. The myriad of putative interacting proteins with p53 and the large number of pathways that apparently this "tumor suppressor" is participating in, beg the question as to which of these interactions are physiological and which are artifactual. So much has been attributed to this molecule that it is highly unlikely that all these functions are biologically relevant. Despite intensive efforts for over 20 years and a huge number of publications, many of them in top scientific journals, no practical application, diagnostic or therapeutic, has as yet emerged for this protein.

Animal models

While these models have been proven invaluable in some situations, and they play a major role in understanding biology, test new therapies, establish toxicity, etc., they can also provide misleading results related to human diseases. For example, models for prostate cancer rely on animals that never naturally develop prostate cancer, whose prostates have been transfected with oncogenic viruses, oncoproteins, etc. Large data sets have been generated with such animals which do not necessarily have any relevance to the human disease.

Data over-interpretation

Scientists tend to be highly biased towards their hypotheses and they can go a long way in proving them correct. Sometimes, due to the small differences between experiments, overzealous statistical analysis (so called "data massage") and elimination of "outliers" can achieve statistical significance that is either artifactual or practically meaningless.

Bioinformatics biases

The emergence of powerful bioinformatics tools (so-called "artificial intelligence") has tempted many investigators to use statistical packages without really understanding the tests performed and their limitations. On the other hand, "skilled" biostatisticians can identify specific cutoff points in clinical studies which will give them the hard sought "statistical significance" ($p < 0.05$). These identified differences could be so fragile, that once tested in other settings, they are destined to collapse. Careful re-analysis of original data sets can reveal such shortcomings, an example being the highly touted "Ovacheck" for early ovarian cancer diagnosis. For details, see the cited literature [13–16].

Reviewers' biases

We all know that the current peer review process is not perfect and it is only used for the lack of a better system. However, we should accept the fact that the review process does not always work well. For example, papers from prominent investigators are not rejected as frequently as equivalent papers of less known investigators. Authors' connections are highly valuable in achieving acceptance of high-profile papers in high-impact journals and authors' enemies can ruin them. Additionally, reviewers may have vested interest in the papers that they review. Not only they can gain important information before a paper is printed, but if they are highly cited, they may also be reluctant to reject them because they will miss citations in their "Bibliometric" records.

Fortunately, after years of debating if the current peer review system is the only way to assess scientific papers, some fresh ideas, supported by the Internet, may gain widespread acceptance. For example, an author could post a paper on an on-line journal and prospective readers could act as "reviewers" by posting their comments along with the paper. This and other ideas are now being tried by some journals.

Outlook

Scientific publishing is a business controlled by powerful organizations and follows the usual rules of supply and demand. A high-profile journal can afford to reject a large percentage of submissions, while smaller journals thrive by accepting papers that have already been rejected by higher impact journals. Scientists rely on their publications to build their careers and compete for scarce resources and, for that matter, they strive towards publishing as much as possible ("publish or perish"). Academic institutions and granting agencies rely heavily on the published record of an individual to make decisions. Scientific publications also follow the rules of evolution; those papers that are likely to make a difference (e.g. introduction of a new and effective therapy, discovery of a new diagnostic procedure, etc.) will, sooner or later, find their way to practical applications. There is no apparent harm in letting low-impact and probably "worthless" [17] scientific literature to sit on bookshelves of libraries or on hard-drives of computers. But, one item of interest is to how to treat high-profile papers which attract media attention as soon as they are published, and they appear to describe major advances. Such "breakthroughs" are very useful to the authors since they increase their prestige and the fame of their organizations, and they can even drive-up stock prices of commercial organizations, sometimes with suspicion for fraud.

While solutions are not easy, one useful measure would be to reassess such papers in 5–6 years to see what happened to them and examine if, and why, they did or did not deliver the promised goods. It would also be highly desirable for scientific journals to publish validation and reproducibility studies aiming to either confirm or refute the published data [18]. Currently, papers with negative data are not viewed as favorably as those publishing positive results [19,20]. Papers that are found not to be reproducible or to contain methodological flaws or other mistakes should be reported and the original authors should be encouraged or pushed to retract them in order to clean-up the database. I was quite encouraged to see [20] that there is now a growing list of journals devoted to publishing “negative” data. With one of my own experiences, the journal *Lancet*, who published a highly acclaimed paper on early ovarian cancer diagnosis by using proteomic profiling [21], denied my contribution dealing with the weaknesses of the method (which is now known not to work) and for which I proposed a retraction. Fortunately, other journals later welcomed my contribution [16,22].

Conclusion

We need to alert the younger scientists that the scientific literature is full of artifactual and/or irreproducible data and train them to exercise caution when they read the literature. Editors and publishers should encourage publication of validation studies on published papers and give space to negative data, so that invalid papers are identified and eliminated from the pool. In realistic terms, so much is at stake regarding publications (fame, promotions, grants, etc.) that over-publication, over-interpretation and fierce competition for high-profile space will undoubtedly continue. Unfortunately, we have to rely on natural selection to find out which papers, in the long-run, have made a difference and which represent loads of information that is either useless, irrelevant or misleading. Mentorship of young scientists on publication practices may help alleviate the problem.

References

[1] Ball P. Prestige is factored into journal ratings. Nature 2006;439:770–1.

[2] Cameron E, Pauling L. Supplemental ascorbate in the supportive treatment of cancer: prolongation of survival times in terminal human cancer. Proc Natl Acad Sci U S A 1976;73:3685–9.

[3] Moertel CG, Fleming TR, Creagan ET, Rubin J, O’Connell MJ, Ames MM. High-dose vitamin C versus placebo in the treatment of patients with advanced cancer who have had no prior chemotherapy. A randomized double-blind comparison. N Engl J Med 1985;312:137–41.

[4] Kennedy D. Next steps in the Schon affair. Science 2002;298:495.

[5] Fuyuno I. Hwang scandal hits Korean biotech hard. Nature 2006;439:265.

[6] Kwok LS. Reactions to the Hwang scandal. Science 2006;311:606–7.

[7] Yidong G, Marshall E. Biotechnology in China. Doubts over new antibiotic land co-authors in court. Science 2006;311:937.

[8] Fuyuno I, Cyranoski D. Doubts over biochemist’s data expose holes in Japanese fraud laws. Nature 2006;439:514.

[9] Marris E. Doctor admits Lancet study is fiction. Nature 2006;439:248–9.

[10] Marris E. Should journals police scientific fraud? Nature 2006;439:520–1.

[11] Anonymous. Bogus background. Science 2006;312:193.

[12] Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev, Cancer 2005;5:142–9.

[13] Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst 2005;97:315–9.

[14] Coombes KR, Morris JS, Hu J, Edmonson SR, Baggerly KA. Serum proteomics profiling—a young technology begins to mature. Nat Biotechnol 2005;23:291–2.

[15] Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst 2005;97:307–9.

[16] Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics 2004;3:367–78.

[17] de Carvalho LB. Pressure also leads to worthless publications. Nature 2006;439:784.

[18] Shamoo AE. Data audit would reduce unethical behaviour. Nature 2006;439:784.

[19] DeCoursey TE. It’s difficult to publish contradictory findings. Nature 2006;439:784.

[20] O’Hara B, Fowler MS, Johnson CA. Why negatives should be viewed as positives. Nature 2006;439:782.

[21] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.

[22] Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst 2004;96:353–6.

Eleftherios P. Diamandis

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, 60 Murray Street, Toronto, ON, Canada M5G 1x5

Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

E-mail address: ediamandis@mtsinai.on.ca.

Fax: +1416 586 8628.

10 July 2006