Biomarkers, 2012; 17(3): 193–200
© 2012 Informa UK, Ltd.
ISSN 1354-750X print/ISSN 1366-5804 online
DOI: 10.3109/1354750X.2012.656287

informa
healthcare

COMMENTARY

Commentary: statistics for biomarkers

David P. Lovell

Division of Biomedical Sciences, St. George's, University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom

Abstract
This short commentary discusses *Biomarkers'* requirements for the reporting of statistical analyses in submitted papers. It is expected that submitters will follow the general instructions of the journal, the more detailed guidance given by the International Committee of Medical Journal Editors, the specific guidelines developed by the EQUATOR network, and those of various specialist groups. *Biomarkers* expects that the study design and subsequent statistical analyses are clearly reported and that the data reported can be made available for independent assessment. The journal recognizes that there is continuing debate about different approaches to statistical science. *Biomarkers* appreciates that the field continues to develop rapidly and encourages the use of new methodologies.

Keywords: Computational biology, proteomics, genetic polymorphisms

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

Introduction

Biomarkers is a successful and well-respected peer-reviewed journal covering a wide range of research fields. It has become the journal of choice for researchers publishing in the area of biomarkers research. Getting a paper published in the journal is competitive and is resulting in an increasingly high standard of accepted publications. The Editorial Board's objective is to maintain and improve on this high quality.

Biomarkers publishes research on *in silico*, *in vitro*, *in vivo*, and human studies. Despite this diverse field, a common feature in their acceptance for publication is evidence of the quality of the design and analysis of the studies reported. This is the crux of evidence-based research with the methods and results sections forming the core of a paper. The peer-review process aims to ensure that the conduct of studies and the reporting of results have been correctly carried out. *Biomarkers* expects high standards but recognizes that it needs to be vigilant as scientific research continues to be affected by errors in the conduct and reporting of research and by fraudulent research. There have been reports of the high incidence of statistical errors, poor statistical practice, and limitations in the designs used in papers published by peer-review journals. *Biomarkers* recognizes that the strength of the peer-review system is its claim to ensure the integrity of the evidence base necessary for scientific progress and by checking "whether the research stands a chance of being 'truthful' " (Vandenbroucke 2009).

Biomarkers, therefore, starts from the position that it is of paramount importance that studies published in a peer-reviewed journal should have been correctly designed, carried out, and reported, and that the results are provided in such a way that the experimental and statistical methods could be repeated. This is also important for both economic and ethical considerations. Transparency in terms of the availability of and access to the original raw data is a key component for the critical assessment of evidence-based research.

Monitoring this process is not easy. The increasing number of journals, the increased output of researchers, and the pressure to publish in high impact journals put a strain on the refereeing system. Referees face increasing workloads, the need to give priority to their own work, and a lack of employer support for the activity.

This is particularly so when it comes to the assessment of the statistical content of the papers. There are probably not enough statisticians familiar with the scientific fields in biomarker research to help with the refereeing process and those who are familiar face considerable demand

Address for Correspondence: David P. Lovell, Division of Biomedical Sciences, St. George's, University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom. Tel: 020-8725-5363. Fax 020-8726-2993. E-mail: dlovell@sgul.ac.uk

(Received 22 November 2011; revised 05 January 2012; accepted 07 January 2012)

RIGHTSLINK

on their time. They may find the requests for (free and unfunded) advice on statistical issues neither feasible nor helpful for their careers.

In practice, these precious statistical skills are probably not used most effectively. A paper developing specific statistical methodology or involving novel methods or where the interpretation of the statistical analyses is complex clearly needs a referee with appropriate statistical expertize. However, problems often arise in papers which are not referred to a statistician because they are thought to have low or no statistical content but, in fact, have serious experimental design or statistical analysis issues which affect their quality.

Therefore, statisticians may not be best used in specific statistical refereeing because it is not particularly efficient to have statistical expertize being brought into issues at the refereeing stage. There are appreciable costs involved in reviewing fairly. The effort put in to a single paper (which is subsequently rejected) provides no "value added" in terms of improving other papers. An important consideration is how to improve the statistical expertize of the non-statistician so they can get to a position where they are capable of knowing when to call in more expert assistance.

Statistical guidelines

One approach has been the development of guidelines to authors for the reporting of studies, including the results of statistical analyses. *Biomarkers*, at present, does not have statistical guidelines. It does, though, have instructions to authors, which provide sensible general requirements.

> You should describe your selection of the observational or experimental participants, identify the methods, apparatus and procedures in sufficient detail to allow others to reproduce the results, and describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.

The International Committee of Medical Journal Editors (ICMJE 2008) provides more detailed guidelines for reporting statistics in its "Uniform Requirements for Manuscripts Submitted to Biomedical Journals." These provide a good starting point for a statistical section in a paper. These state as follows:

> IV. A. 6. c. Statistic
>
> Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as *p* values, which fail to convey important information about effect size.
>
> References for the design of the study and statistical methods should be to standard works when possible (with pages stated). Define statistical terms, abbreviations, and most symbols. Specify the computer software used.

This statement is a good and succinct synopsis of the basic requirements required for a statistical section. *Biomarkers* does not, at present, want to take a prescriptive approach to how statistical analyses are carried out. For instance, it will avoid being prescriptive in controversial areas, such as the choice of Frequentist or Bayesian methods, the use or not of multiple comparison methods, or the presentation of data with confidence intervals rather than probability (*p*) values. It will, however, expect the authors to be aware of such issues and present their data accordingly.

Many specific fields of research have developed their own guidelines for reporting of studies, including for the statistical analyses. Many are part of the Enhancing the Quality and Transparency of Health Research (EQUATOR) network (http://www.equator-network.org/). These guidelines have built upon the success of the CONSORT statements (Moher et al. 2001) on the reporting of clinical trials and are often jointly published by a series of journals (Vandenbroucke 2009) and have the aim "to improve the quality of scientific publications by promoting transparent and accurate reporting of health research." Table 1 provides a nonexhaustive list of some of the guidelines available. In many specific technical areas, such as the – omics technical reviews of appropriate statistical methods have been produced; the referee can reasonably expect that an author is aware of them, applies the methods, and cites the publications where appropriate.

Statistical input to the paper

It is expected that the contribution of one or more statisticians to a paper with respect to the design and analysis should be recognized and any statistical help acknowledged. If there is a sizable statistical component to the design and/or analysis, then it is expected that one or more statisticians would be the co-authors and accept all the responsibilities associated with authorship. If no statistician is explicitly named, one of the authors should be identified as having sufficient expertize to take responsibility for the statistical aspect of the paper and to be the point of contact for any issues related to the data and/or the statistical analysis. This will enable faster and better communication on statistical issues and help improve the quality of the reporting of the work.

If an acknowledgment is made to the advice or help of a statistician, he/she should sign a confirmation that they are content with the way their advice has been interpreted. The statistician should be aware that they are being acknowledged and have the right to withhold their name if they are not content with how their advice has been used.

Experimental design and statistical analysis

Although much emphasis is on the statistical analysis of data, it should be recognized that analysis of data is only part of the statistical input into a study and that the study

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

RIGHTSLINK

Table 1. Guidelines for reporting statistical methods.

There are many guidelines for reporting statistical methods. Here is a non-exhaustive list. See also the Cochrane Organization http://www.cochrane.org/about-us/evidence-based-health-care/webliography/books/reporting

Altman, D.G., Gore, S.M., Gardner, M.J. & Pocock, S.J. (1983) Statistical guidelines for contributors to medical journals. Br Med J *286* 1489–1493.

American Physiological Society Guidelines for reporting statistics in journals published by the American Physiological Society. http://physiolgenomics.physiology.org/cgi/content/full/18/3/249

Bailar, J.C. III & Mosteller, F. (1988) Guidelines for statistical reporting in articles for medical journals. Ann Intern Med *108* 266–273.

Chinn, S, (2001) Statistics for the European Respiratory Journal. Eur. Respir. J. *18* 393-401.

Clayton, M.K. (2007) How should we achieve high-quality reporting of statistics in scientific journals? A commentary on 'Guidelines for reporting statistics in journals published by the American Physiological Society: The sequel.' Adv Physiol Educ. *31* 302–304.

Curran-Everett, D. & Benos, D.J. (2004) Guidelines for reporting statistics in journals published by the American Physiological Society. J Appl Physiol *97* 457–459.

Ludbrook, J. (2005) Comments on journal guidelines for reporting statistics. Clin Exp Pharmacol Physiol *32* 324–326.

International Committee of Medical Journal Editors (1988) Uniform requirements for manuscripts submitted to biomedical journals. Ann Intern Med *108* 258–265.

International Committee of Medical Journal Editors (1997) Uniform requirements for manuscripts submitted to biomedical journals. Ann Intern Med *126* 36–47.

Specific Guidelines

ARRIVE Animal Research: Reporting In Vivo Experiments

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M. & Altman, D.G. (2010) The ARRIVE guidelines Animal Research: Reporting In Vivo Experiments. PLoS Biol 8(6): e1000412. doi:10.1371/journal.pbio.1000412
http://www.nc3rs.org.uk/downloaddoc.asp?id=1206&page=1357&skin=0

CONSORT

Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F. et al. (2001) The Revised CONSORT statement for reporting randomized trials: explanation and elaboration. Annals of Internal Medicine *134* 663–694.

Moher, D., Schulz, K.F., Altman, D.G. for the CONSORT Group (2003). The CONSORT statement: revised recommendations for improving the quality of reports of parallel- group randomised trials. Clin Oral Investig *7* 2–7.

Schulz, K.F., Altman, D.G., Moher, D, CONSORT Group (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS; 7(3):e1000251.

Altman, D.G. (2005) Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. BMJ *330* 1056-1057.

Moher, D., Hopewell, S., Schulz, K.F, Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M. & Altman, D.G. (2010) CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials

BMJ *340* c869. http://www.bmj.com/content/340/bmj.c869.full

EQUATOR (Enhancing the QUAlity and Transparency of health Research)

Altman, D.G., Simera, I., Hoey, J., Moher, D. & Schulz, K. (2008) EQUATOR: reporting guidelines for health research. Lancet *371* 1149–1150.

Simera, I., Altman, D.G., Moher, D., Schulz, K.F. & Hoey, J. (2008) Guidelines for reporting health research: the EQUATOR network's survey of guideline authors. PLoS Med *5*:e139.

GRIPS (Genetic Risk Prediction Studies)

Janssens, A.C.J.W., Ioannidis, J.P.A., van Duijn, C.M., Little, J & Khoury. M.J. for the GRIPS Group (2011) Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. European Journal of Human Genetics *19* 833–836.

Guidelines for Meta-Analyses

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.W., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B. for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. (2000) Meta-analysis of Observational Studies in Epidemiology A Proposal for Reporting JAMA *283* 2008– 2012.

QUOROM (Meta Analysis)

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D. & Stroup, D.F., for the QUOROM Group (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet *354:*1896–1900.

REMARK (REporting recommendations for tumour MARKer prognostic studies)

McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M. & Clark, G.M. (2005) Reporting recommendations for tumor MARKer prognostic studies (REMARK). Nat Clin PractUrol *2* 416–422.

SQUIRE (Standards for Quality Improvement Reporting Excellence)

Davidoff, F., Batalden, P., Stevens, D., Ogrincm G. & Mooney S., SQUIRE Development Group.(2008) Publication guidelines for improvement studies in health care: evolution of the SQUIRE Project. Ann Intern Med *149*:670e6.

STARD (Standards for Reporting of Diagnostic Accuracy)

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M. et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Clin Chem *49* 1–6.
(Also published in other journals BMJ (2003) *326* 41–44.)
http://www.consort-statement.org/stardstatement.htm

(*Continued*)

© 2012 Informa UK, Ltd.

RIGHTSLINK

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

Table 1. *(Continued)*.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M. Moher, D., Rennie, D., de Vet, H.C.W. & Lijmer, J.G. (2003) The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. Clinical Chemistry *49* 7–18.

STREGA (STrenghtening the REporting of Genetic Association Studies)

Little, J., Higgins, J.P. Ioannidis, J.P. et al (2009) STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6: e22.

Little, J., Higgins, J.P.T, Ioannidis, J.P.A., Moher, D., Gagnonm F., Von Elmm E., et al (2009). STrengthening the REporting of Genetic Association studies (STREGA) – an extension of the STROBE Statement. J Clin Epidemiol *62* 597e608.

STROBE (STrengthening the Reporting of OBservational studies in Epidemiology)

Vandenbroucke, J..P., von Elm, E., Altman, D. G., Gøtzsche P. C., Mulrow C. D., Pocock S. J., Poole C., Schlesselman J. J. & Egger M(2007) STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 16:e297.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. & Vandenbroucke, J.P. for the STROBE Initiative.(2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. Ann Intern Med *147* 573-577.
http://www.strobe-statement.org/index.php?id=strobe-publications

von Elm, E., Altman, D.G., Egger, M., Pocock, S. J., Gotzsche, P.C. & Vandenbroucke J. P. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS.Med. *4*: e296.

STROBE-ME (STROBE-Molecular Epidemiology)

Gallo, V., Egger, M., McCormack, V., Farmer, P.B., Ioannidis, J.P.A., Kirsch-Volders, M., Matullo, G., Phillips, D.H., Schoket, B., Stromberg, U., Vermeulen, R., Wild, C., Porta, M. & Vineis, P. (2011) STrengthening the Reporting of OBservational studies in Epidemiology – Molecular Epidemiology (STROBE-ME): An extension of the STROBE statement. PLoS Medicine (in press).

QUADAS

Whiting, P.F., Weswood, M.E., Rutjes. A.W., Reitsma. J.B., Bossuyt. P.N. & Kleijnen, J. (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol *3* 25.

Guidelines in specific research areas

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. Nat Genet 29 365-371.
Preparation of data submission (MIAME checklist)
http://www.mged.org/Workgroups/MIAME/miame_checklist.html

Lindon, J.C., Nicholson, J.K., Holmes, H.., Keun, H.C., Craig, A., Pearce, J.T.M., Bruce, S.J., Hardy, N., Sansone S.A. et al. (2005) (The Standard Metabolic Reporting Structures working group**).** Summary recommendations for standardization and reporting of metabolic analyses. Nat Biotechnol. *23*: 833-838.
MIAME: Minimum Information About a Microarray Experiment http://www.mged.org/Workgroups/MIAME/miame.html

MIBBI: Minimum Information for Biological and Biomedical Investigations
http://mibbi.org/index.php/Main_Page

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

or experimental design is as, or even more, important. The design can be considered the strategic aspect, while the methods of analysis are the tactical part. Good study design leads directly to an appropriate statistical analysis; poor or no design can result in no feasible statistical analysis being possible. This links with the common advice to "consult a statistician at the experimental design rather that statistical analysis." A quote of R.A. Fisher is apt: "To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of" (Fisher 1938).

Biomarkers, therefore, expects to see evidence of the planning that went into a study and to see statistical analyses, which make full use of the design. Examples would be details of the statistical analysis plan (SAP), consideration of the primary endpoint, and whether the primary aim of the study was hypothesis testing or hypothesis generation. A failure to declare in the "methods" section that blinding and randomization were carried out would be interpreted as implying that this had not been done. Details, such as the type of randomization, e.g., block or stratified and methods used for blinding, should be given when relevant.

Biomarkers expects to see a justification of the sample sizes used and, where relevant, the power calculations, which were carried out as part of the development of the SAP for both experimental and observational studies. Such sample size determination is likely to be carried out at the design stage, because this is increasingly expected to be transparent and subject to scrutiny by a statistician on a grant, regulatory, or ethics board. A wide range of software, web-based resources, books, and formulae are now available, which can give sample sizes for a given power or power for a given sample size for simple designs. Examples of books on sample size methods include Machin et al. (1997) and Chow et al. (2003). Reference to the methodology used will increase the transparency. Studies with small sample sizes, which may consequently have low power, are likely to be of particular concern to referees. However, experimental designs, such as those related to the factorial design, are very efficient for investigating multiple experimental factors using small sample sizes (Festing 2003).

Referees should also ensure that statistical analyses are carried out on the appropriate experimental unit. The experiment unit is the entity that is randomly allocated

RIGHTSLINK

to the treatment group. Biological replicates may be the experimental unit but technical replicates are generally not. Analyzing the cell as an experimental unit, rather than the animal, results in pseudo-replication and, by inappropriately increasing the sample sizes used in a statistical test, can lead to overestimation of the statistical significance of results.

Statistical significance

The finding of statistical significance in a test of a null hypothesis is often considered to be the primary objective of a study. However, this is a complex and controversial topic. The widely used Null Hypothesis Significance Testing Procedure (NHSTP) is a hybrid of methods developed independently by Ronald Fisher in the 1920s and by Jerzy Neyman and Egon Pearson in the 1930s. This approach is the basis of the method described in many standard textbooks and the statistical tests found in much software. However, the compromise between the Neyman–Pearson and Fisher approaches results in serious theoretical objections to its use by many statisticians, although it has its supporters who stress its pragmatic use (Frick 1996; Berger 2003). Alternative approaches, such as the use of estimation and confidence intervals, have been proposed (Altman et al. 2000) and adopted as journal policy by a number of journals. Others have argued that the known limitations of the current hypothesis testing approach can be overcome by switching to an alternative Bayesian view of statistical methodology (Lee 2011).

In practice, however, the NHSTP is currently so embedded in biomedical research that it will continue to be used for some time. However, uncritical use of hypothesis testing and the reporting of results as either statistically significant ($p < 0.05$) or, preferably, with the exact p values is not acceptable. Statistical significance alone is not a justification for publication. Ioannidis, for instance, identified the inappropriate use of p values as one explanation for "why most published research findings are wrong" (the title of one of his papers) (Ioannidis 2005). It is increasingly recognized that there is a distinction between the concepts of the statistical significance and biological importance (EFSA 2011). An over emphasis on statistical significance as evidence for biological importance partly explains the evidence suggesting that the publication policies of journals is creating a publication bias detectable in meta-analyses by, for instance, funnel plots.

It is, therefore, important to note that *Biomarkers* policy is that well-designed studies, which produce negative results, are viewed favourably for publication. This policy also meets the ICMJE (2008) obligation to publish negative studies.

Reproducible research

Gentleman and Lang have developed the concept of "reproducible research," where anything in a scientific paper should be reproducible by the reader (Gentleman and Lang 2004). By reproducible research, Gentleman and Lang meant research papers with accompanying software tools, such as the open-source R statistical system, that would allow the reader to directly reproduce the results and employ the methods that are presented in the research paper. Baggerly and Coombes (2009) have championed this approach as an important aid in their use of forensic bioinformatics to identify problems and misinterpretations in their reanalyses of large-scale biological data.

The expectation is that reanalysis of the data should be feasible. However, at present, very little "raw data" is available for independent review. This means that many results and the conclusions drawn from them must be taken on trust. However, the nature of scientific publishing is changing. While printed space remains at a premium and restricted, restraints on the length or details of information which can be included in a publication can be relaxed or even removed through the use of electronic versions where fuller versions, detailed analyses, and the full original and supplementary data can be provided.

Studies should, therefore, be more easily repeatable, results checkable, alternative analyses and or sensitivity analyses by multiple authors and multiple groups increasingly feasible. Fuller descriptions of the methods and more detailed tables etc. can be provided as supplementary information. Examples might include printouts which would allow the diagnostic statistics derived from modelling to be examined and for other opportunities to check the analyses, assumptions underlying them, sensitivity analyses etc.

Although there may be practical problems, the technology exists to make the raw data and the actual statistical analyses available in an electronic format either at the journal's website or at the author's own web page. Other journals now place an obligation on the authors "to make sufficient data publicly available for an experiment to be reproduced" (Anonymous 2007).

Availability of data and presentation of results

The presentation of the results of analyses of data forms the core of any paper and the basis for evidence-based practice. The opportunity to assess the quality of data in a study is crucial for the assessment of the quality of research work, and the central tenet, here, is transparency.

The potential complications of selective reporting of results, multiple endpoints, multiple testing, inappropriate combination of data in meta-analyses require that it is mandatory that a full description of the statistical methods, including specific tests options, are full reported.

It is important that the full statistical analysis could be made easily available. The condensed version provided in a paper can be difficult to follow. A referee may be able to save time by looking at the original analysis rather than an edited version. This should not be a practical problem as many statistical packages now have a facility to maintain a complete record or log of all the analyses carried out and any manipulations of the data. These records should

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

© 2012 Informa UK, Ltd.

RIGHTSLINK

also be kept and deposited as part of the requirement for reproducible research. This log allows the re-creation of the statistical analyses carried out. This has long been a feature of work carried out for regulatory purpose in the pharmaceutical industry under good laboratory practice (GLP) and good clinical practice (GCP) and is now receiving attention within the academic community. The logs and analyses should form part, together with the raw data, of the supplementary information which is deposited as outlined by Green (2003).

Sufficient information should, therefore, be available to be able to repeat an experiment. Ideally, the original (or "raw") data, statistical inputs, result outputs, and program logs should be available initially for referees and, ultimately, the wider scientific community (through non-heroic means) so that the full analysis carried out could be reviewed, rerun and, if necessary, reanalyzed by an independent statistician by different methods using different statistical software.

Biomarkers will, therefore, expect "raw" data to be available for inspection and reanalysis. In future, all "raw" data should be available for scrutiny in an electronic format (such as in supplementary tables or as data held at websites) in a format that guarantees availability for 10 years. Any relevant annotations (meta-data) associated with the data should also be available and similarly archived.

The ICMJE (2008) includes in its guidelines the following text (*IV. A. 7.*) about the "results" section:

> Present your results in logical sequence in the text, tables, and illustrations, giving the main or most important findings first. Do not repeat all the data in the tables or illustrations in the text; emphasize or summarize only the most important observations. Extra or supplementary materials and technical detail can be placed in an appendix where they will be accessible but will not interrupt the flow of the text, or they can be published solely in the electronic version of the journal.
>
> When data are summarized in the Results section, give numeric results not only as derivatives (for example, percentages) but also as the absolute numbers from which the derivatives were calculated, and specify the statistical methods used to analyze them. Restrict tables and figures to those needed to explain the argument of the paper and to assess supporting data. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables. Avoid nontechnical uses of technical terms in statistics, such as "random" (which implies a randomizing device), "normal," "significant," "correlations," and "sample."
>
> Where scientifically appropriate, analyses of the data by such variables as age and sex should be included.
>
> More information is found in IV. A. 10.

New statistical approaches and description of methods

Statistical methodology is a fast developing field with new methods and techniques being developed and used (Rafferty et al. 2002). Statistics has changed from the calculator-based algorithms familiar in many standard textbooks to computer-intensive methods. The breadth and complexity of statistical methods means that non-statisticians can now understand appreciably fewer papers than they could have 30 years ago (Switzer and Horton 2007) and few statisticians can now be experts in all areas.

Biomarkers encourages the use of innovative approaches. Alongside the more familiar, even traditional, chi-square, *t*-test, correlations, linear regression, and analyses of variance methods, there are now a range of newer, less familiar and more complex approaches. In recent issues of *Biomarkers* the statistical methods used have included Kaplan–Meier survival curves, Cox's proportional hazards multivariate analysis, generalized linear model (GLM), receiver-operator characteristic (ROC) curves, linear mixed effects modelling, autoregressive covariance structures, Bonferroni corrections, random forest methods, logistic regression, principal components analysis, and orthogonal partial least squares. This is not an exhaustive list, and to the statistical community these represent a set of tried and tested methods with a clear set of uses, potential benefits, and limitations. Most can be accessed through proprietary software using clearly defined procedures with options available for modifications of the analyses. The user should ensure that the specific procedures and options used are clearly defined. A brief overview of some of the newer methods should also be included so that the general features of the method are available.

Biomarkers recognizes that there are different "schools" of statistical thinking and methodology broadly characterized as Frequentist, Bayesian, Likelihood and Information-Theoretic with different statistical philosophies underlying statistical analyses (Anderson 2008). *Biomarkers* appreciates that such methods need encouragement but also critical evaluation. In some cases, the methods are derived from disciplines, such as bioinformatics, computer sciences or machine learning where terminologies may be unfamiliar and overlap with those used in classical statistical methods.

It is important to appreciate that there is no such thing as the correct analysis but there are many incorrect ones. A modern journal should be able to adjust to this. These more complex and less familiar methods are now being used and referees should expect to see them being applied in fields relevant to *Biomarkers*. For instance, new methods, such as Bayesian approaches, are welcome but sufficient information are to be provided to describe and explain the method so that a statistician familiar with the methods can usefully review the approach and the general reader can understand the general approach being used.

The commonly used statistical tests can be specifically described by reference to a page in widely available statistical texts books, such as, for instance, Altman (1991).

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

RIGHTSLINK

When more advanced methods are carried out using a popular statistical software package, such as SAS, SPSS, Stata, Statistica, Minitab, JMP etc., the software version, the procedures, and any specific options applied should be reported in the "statistics" section. An explicit description of more advanced procedures or unusual statistical methods should be provided in the "statistics" section to help the reader appreciate what the method does and why it has been used. Any less widely available software package should be described and reference made to how it could be obtained and the specific analyses carried out using it. The limitations of using Excel for statistical analyses should be appreciated (McCullough and Heiser 2008).

Modelling

There is likely to be a continuation of the trend away from formal hypothesis testing approaches to methods based upon estimation (such as equivalence, noninferiority, and superiority tests) and to statistical modelling.

Modelling underlies many traditional statistical methods. Many of the traditional parametric methods, such as the analysis of variance (ANOVA) and multiple regression methods, are part of a wider modelling approach, the General Linear Model (GLM), which, in turn, is part of an even wider modelling framework (Dobson and Barnett 2008). These more sophisticated experimental designs and analyses move from a traditional hypothesis testing approach into a modelling methodology where estimates are derived for various model components and attempts made to identify the best fitting and hopefully the most predictive model. In fields, such as the – omics, informatic approaches based upon computing algorithms have joined with other statistical methods, such as multivariate methods, to develop new methods of analysis, such as data mining. A large number of these statistical/bioinformatics approaches are aimed at developing statistical (and other) models which try to explain a data set and then to make predictions about data from new items. A clear distinction is to be made between the success of a model at explaining or describing a data set and its ability to predict new items.

Model development requires appreciable skill to be able to identify the set of variables to include in the model, to avoid over-fitting and to validate and test the model on new data to avoid the dangers of over-prediction. General guidelines for model fitting have been developed by Harrell (2001). In particular, any "black box" approach should be described in sufficient detail so that the methodology is transparent, because some of these approaches while being successfully applied have been criticized for being difficult to interpret.

These considerations are also important in distinguishing between studies aimed at hypothesis testing as opposed to hypothesis generation. This distinction links into the debate on the concept of "discovery biology" and whether biological studies should be discovery (Golub 2010) or hypothesis driven (Weinberg 2010).

Visualization of data

Visualization of data is increasing with methods for the presentation of data and information developing rapidly using the increased computing power available. More sophisticated diagrammatic and graphical ways of presenting data visually, including the use of colour, are feasible and are more cost-effective electronically than in hardcopy versions of publication. However, while graphical presentation is useful, it should not supersede formal statistical methods. Charts and figures to be included in the text version of the paper should be kept simple and as self-explanatory as possible. The possibility of more detailed and sophisticated presentation of data in electronic form is recognized.

Graphs using "perspective" should generally be avoided so that representing inter-sample variability using 3D displays and histograms needs care (van Belle 2008). Excel figures and graphs, while a convenient and accessible resource, are often not considered suitable for scientific publications.

Completeness of the data set

Transparency of the selection of data for analysis, results for reporting (publication bias), and the handling of missing or omitted data is important. Information should be provided on what the full data set was and what was actually used in the analysis. Data missing because of some systematic cause as opposed to random effects can introduce biases leading to incorrect conclusions. Analyses that use only the data from the subjects remaining at the end of the study (a *per protocol* analysis in a clinical trial context) can overestimate the size of effects. Alternatively, analyzing all the subjects in the study even though some have dropped out or missing can lead to an underestimation of effects. Such *Intention to Treat* (ITT) analyses, which are usually the default option for the analysis of clinical trials, usually require estimates to be made on any missing data. In all cases, the nature of the missing data and any implication on the analysis should be clearly outlined. An example of the presentation of the inclusion or exclusion of units is the type of flow diagram which depicts the passage of participants through each stage of an RCT illustrated in the Consolidated Standards of Reporting Trials (CONCORD) statement (Moher et al. 2001).

Recommendations and conclusions

The recommendations in this commentary are based upon what would generally be considered good statistical practice and relate to concepts, such as the reproducibility of research and the need to provide robust data and results. *Biomarkers*, therefore, expects that authors are aware of statistical guidelines in their field of research and prepare their submissions taking these points into account.

It is expected that a statistician would be one of the authors of a paper or, alternatively, that one author is

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

© 2012 Informa UK, Ltd.

RIGHTSLINK

identified as responsible for statistical aspects of the paper. Anyone acknowledged for their help with the statistical analysis should have confirmed that they agree to being identified.

Evidence should be available, if required, of a SAP for the studies, and all relevant data and analyses should be accessible, initially, to the referees and then to the wider scientific community. This is consistent with the policy of other journals and is in the spirit of the developing concept of "reproducible research."

Many statisticians criticize the emphasis given to statistical significance and *p* values and argue that reporting estimates of the size of effects and their associated confidence intervals is more appropriate. Finding statistically significant effects should not be the primary objective of statistical analyses as this does not automatically equate with the biological importance or relevance of a finding. Well-conducted studies which result in informative but negative results justify publication.

The growth in computing power means that new and innovative statistical and bioinformatic methods are being developed to analyze the high dimensional data in research areas, such as the "–omics" and next generation sequencing. *Biomarkers* welcomes approaches that use these new methodologies.

Finally, this commentary aims to help *Biomarkers* ensure that the high quality of the journal is maintained by ensuring that the key experimental evidence in papers submitted to the journal are appropriately presented, interpreted, and reviewed. The key point is that statistical analysis is not a "prescribed ritual" as part of the publication process but rather experimental design and statistical analysis are an integral part of the process of conducting biomarker research.

Declaration of interest

The author has no conflict of interest to report.

References

Altman DG, Machin D, Bryant TN & Gardner MJ. (2000) Editors: *Statistics with confidence: confidence intervals and statistical guidelines*. London: BMJ Books.

Altman DG. (1991) *Practical statistics for medical research*. London: Chapman and Hall.

Anderson DR. (2008). *Model based inference in the life sciences: a primer on evidence*. New York: Springer.

Anonymous. (2007). Editorial: Democratizing proteomics data. *Nat Biotechnol* 25:262.

Baggerly KA, Coombes KR. (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Ann App Statist* 3:1309–1334.

Berger JO. (2003). Could Fisher, Jeffreys and Neyman have agreed on testing? *Stat Sci* 18:1–31.

Chow SC, Shao J, Wang H. (2003). *Sample size calculations in clinical research*. New York, NJ: Chapman & Hall/CRC.

Dobson AJ, Barnett AC. (2008) *An introduction to generalized linear models* (third edition). London: Chapman & Hall.

EFSA Scientific Committee. (2011). Statistical significance and biological importance. *EFSA J* 9:2372.

Festing MFW. (2003). Principles: the need for better experimental design. *Trends Pharmacol Sci* 24:341–345.

Fisher RA. (1938). Comment said to have been made at Indian Statistical Congress, Sankhya, ca 1938. (Widely cited as Fisher quotation.)

Frick R W. (1996). The appropriate use of null hypothesis testing. *Psychol Meth* 1:379–390.

Gentleman R, Lang DT. (2004). Statistical analyses and reproducible research. *Bioconductor Project* working papers. Working paper 2.

Golub T. (2010). Opinion: Counterpoint: Data first. *Nature* 464: 679.

Green PJ. (2003). Diversities of gifts, but the same spirit. *The Statistician* 52: 423–438.

Harrell FE Jr. (2001). *Regression modeling strategies*. New York: Springer.

ICMJE. (2008).Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication (Updated October 2008), The International Committee of Medical Journal Editors.

Ioannidis JP. (2005). Why most published results are false. *PLoS Medicine* 2 e124.

Lee JJ. (2011). Demystify statistical significance – time to move on from the P value to Bayesian analysis. *J Natl Cancer Inst* 103:2–3.

Machin D, Campbell M., M., Fayers P., P, Pinol A. (1997). *Sample size tables for clinical studies*. 2nd Edition, Oxford: Blackwell Science Ltd

McCullough BD, Heiser DA. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Comput Stat Data Anal* 52:4570–4578.

Moher D, Schulz KF, Altman DG, for the CONSORT Group (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *JAMA* 285:1987–1991.

Rafferty AE, Tanner MA, Wells MT. (2002). Eds. *Statistics in the 21st Century*. Edited by Monographs on Statistics and Applied Probability, London: Chapman and Hall.

Switzer SS, Horton NJ. (2007). What your doctor should know about statistics (but perhaps doesn't…). *Chance* 20:17–21.

Vandenbroucke JP. (2009). STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK and CONSORT: for whom does the guideline toll? *J Clin Epidemiol* 62:594–596.

van Belle G. (2008). *Statistical Rules of Thumb*. 2nd Edition New York: John Wiley & Sons.

Weinberg R. (2010). Opinion: point: hypotheses first. *Nature* 464:678.

Biomarkers Downloaded from informahealthcare.com by Changhua Christian Hospital on 11/14/12
For personal use only.

RIGHTSLINK