

Bedside to bench: a look at experimental research with a clinical trial checklist

Karin R. Sipido^{1*}, Barbara Casadei², Paul Holvoet¹, Stefan Janssens¹, Aernout Lutun¹, and Maurilio Sampaolesi³

¹Department of Cardiovascular Sciences, KU Leuven, Herestraat 49, Leuven B-3000, Belgium; ²Department of Cardiovascular Medicine, University of Oxford, Oxford, UK; and ³Department of Regeneration and Development, KU Leuven, Leuven, Belgium

1. Introduction

Editors across all disciplines and approaches ask the same questions to reviewers and judge the merit of a study against a similar checklist: is the study well designed and does it provide reliable findings? Have we learnt something new? Are the results relevant to the community of scientists and/or medical professionals? In short, does the study make an important contribution to existing knowledge?

As clinical practice has become increasingly evidence-based, the number of clinical trials has increased significantly (Presently, >6500 trials with 'heart' as keyword are registered and are open or recruiting; www.clinicaltrials.gov.) and stringent criteria have been put into place to guide the design, execution, and analysis of clinical investigations. These are the standards of good clinical practice ("GCP is an international ethical and scientific quality standard for designing, recording, and reporting trials that involve the participation of human subjects. Compliance with this standard provides public assurance that the rights, safety, and wellbeing of trial subjects are protected, and that clinical-trial data are credible"; http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000072.jsp). Training courses are available to educate clinicians in GCP and the management of patient-based studies. In the past 10 years, the impact factor of journals that focus on clinical research has steadily increased, due, to a large extent, to the recognition that large, well-designed clinical trials are able to produce robust findings that are immediately relevant to patient management and underpin practice guidelines. (The AHA and ACC published more than 70 guidelines in their journals since 2005; <http://my.americanheart.org/professional/StatementsGuidelines>, and the ESC similarly published more than 30; <http://www.escardio.org/guidelines-surveys/esc-guidelines/Pages/GuidelinesList.aspx>.) Not surprisingly, this results in widespread citation of the work, reflecting its impact on the community.^{1,2}

The care in the design, conduct, and analysis of a clinical trial, and the subsequent further scrutiny in meta-analysis, outcome research, and registries are motivated by the high values at stake, namely the direct impact on patient outcome and quality of life. The values at stake may

not seem that high for experimental studies but ultimately they are. Misleading results can lead other researchers to pursue unrewarding avenues of research, and wrong results may take many years to redress. Among reasons for retraction, doubt about data is a major factor³ and lack of reproducibility a major concern.⁴ In particular, given the aspiration that findings in animal models will ultimately inform our understanding and management of disease states in humans, it is important to carefully consider the current standards in bench research.

2. Experimental design

2.1 A robust number of observations to ensure data are reliable

When designing a clinical trial, a hypothesis to be tested is formulated, defining the type of measurements and data to be obtained. The design includes a careful evaluation of the number of patients to be recruited to detect a meaningful difference between interventions (i.e. the calculation of the statistical power of the study). This takes into consideration the reproducibility of the measurements as well as the expected rate of the outcome of interest in the control group. Once a protocol has been established and initiated, it will be applied to all patients unless independent data monitoring reveals a highly significant (larger than expected) difference between groups, futility (no hope of detecting a difference between interventions in the number of patients recruited in the study) or safety issues, all of which would warrant termination of the study for ethical reasons. In experimental studies, the hypothesis to be tested can remain rather general and often research is partly hypothesis-driven, partly explorative. In either case, however, a formal pilot study allowing accurate power calculations is rarely performed. The number of animals that are eventually examined can vary (e.g. from 3 to >10) and it is rarely adequately justified. For studies involving large animals, numbers tend to be even lower, and during the review process 'ethical reasons' are sometimes advocated to justify the low number of observations. Yet, animal sacrifice to obtain a set of inconclusive, non-reproducible

The opinions expressed in this article are not necessarily those of the Editors of *Cardiovascular Research* or of the European Society of Cardiology.

* Corresponding author. Tel: +32 16 330815; fax: +32 16 345844, Email karin.sipido@med.kuleuven.be

© The Author 2013. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

data resulting in controversy is, by definition, not ethically justifiable, likely leading to even more animals being used.

2.2 Randomization and blinded observations

In clinical studies, genetic and environmental differences in the population underpin the need for patient randomization or, at the very least, careful matching between groups. Neither practice is rigorously followed in experimental research. For instance, genetically modified mice should be compared with their wild-type littermates, a simple rule that is not always observed. Sex differences should ideally be addressed by carrying out studies of larger size including similar numbers of male and female animals. A rigorous plan for randomization is rarely in place; likewise, appropriate sham interventions are not always provided, all of which are common reasons for rejection when a manuscript is submitted. Such wasted data could have been prevented by due consideration to the study design *ab initio*.

Avoidance of bias by 'blinding' the investigators in charge of recording outcome data of the treatment allocation or mouse genotype should be mandatory, yet it is often not applied to experimental studies. This is of great concern as animal welfare regulations may prevent the use of survival rates resulting in the collection of surrogate endpoints that are often operator-dependent and prone to bias (e.g. assessment of left ventricular function by echocardiography or quantification of atherosclerotic plaque size in histological sections of the aortic root).

Another strategy to avoid bias and an effective way of reaching an adequate sample size in a shorter time is the involvement of several research centres in the same study. In experimental research, multicentre studies have become more frequent, but their aim is usually to obtain different sets of data from a number of research laboratories with complementary expertise.⁵ Although this results in a more comprehensive testing of a particular hypothesis, it does not address whether key outcomes are reproducible across laboratories. The incentives for collaborative studies that run parallel experiments and pool data have not been high until now. Yet, good experimental practice is unlikely to be widely adopted unless failure to adhere to it would seriously prejudice the outcome of funding applications or the publication of experimental results. In many ways, for no good reasons, experimental practice in basic research lags behind that for clinical investigations. The extent to which these issues have precluded effective translation of findings from animal investigations into the clinical arena may be considerable.⁶

3. Data processing

3.1 Statistics

Random sampling of publications in journals that are predominantly reporting cardiovascular experimental work shows that studies on animals, tissue samples, and cells mostly employ parametric statistics but rarely test for normality, despite the use of small data samples. Together with due consideration to study design and sample size, a formal plan for data analysis in hypothesis-driven experiments would substantially enhance the quality of the evidence and reproducibility of the findings. The *Journal of Physiology* launched a series of articles on statistics a few years ago, with valuable guidelines, including the presentation of results.⁷ The popular bar graph is too often used inappropriately and hides information on the distribution of data of interest. Several data analysis programmes, such as GraphPad Prism (GraphPad Software;

<http://www.graphpad.com>), will also guide researchers in the use of basic statistics, whereas more complex analyses can be performed using SPSS (IBM; <http://www-01.ibm.com/software/analytics/spss/products/statistics/>) and the freeware R (<http://www.r-project.org/>). In an era of multidisciplinary research, consulting a statistician at the time of study design and data analysis should be a given; yet, it is rarely done.

3.2 Access to raw data and meta-analysis

Clinical trial registration and access to raw data by an independent steering committee ensure adequate monitoring of clinical studies. Sharing of individual patient data for meta-analysis and exploratory subgroup analyses further increases the information and impact that can be accrued from clinical investigations. In bench research, data deposition of large-scale studies, typically gene and protein data, has become mandatory for a number of journals. However, the utility of such data to other investigators remains unclear, as the data are not always well annotated and the number of databases is large (see, e.g. http://mybio.wikia.com/wiki/Microarray_databases). Yet, as recently shown, a consortium effort to organize data sets in a rigorous manner can greatly enhance their value.⁸ It is conceivable that sets of experimental data can be used and included in a meta-analysis,⁹ and efforts have been launched to obtain more standardized data acquisition and annotation.¹⁰ At present, calls for provision of raw data to reviewers and for institutional data repositories are more driven by fears of data manipulation than by the potential for a scientific added value.

4. The basic research checklist

Among the three criteria of 'correct, novel, relevant', it is novelty that is often stressed in basic research. Prominent journals with a focus on basic science feature exciting new observations, most often based on novel technology, tools, or unique new animal (mouse) models. Technological advances, such as the -omics approaches and advanced imaging, have been a major driver to the wealth of new insights into cellular and molecular mechanisms. In the checklist, truly novel data are nearly always considered relevant, as they open uncharted territory and attract scientists and media attention. Yet, there has been concern that the pressure for novelty might lead to less stringent methodological review and disappointing reproducibility.⁴ Balancing the 'pressure to publish' with the need for publishing evidence that is sufficiently robust to be considered for translation 'from bench to bedside' is a target that requires active involvement of all stakeholders; the investigators, the journals, and the funding bodies, in particular. Adhering to recommendations and standards, such as the ARRIVE guidelines,¹¹ can improve the quality and reproducibility of the data. In the present issue, new instructions to authors illustrate the commitment of this journal to these important issues. Further translation from 'bedside to bench' of the GCP standards can enhance the information and knowledge to be gained from experimental research, will help addressing the 3 R's of animal studies, and will increase the overall impact and contributions of basic research.

Conflict of interest: none declared.

References

1. Luscher TF, Gersh B, Hendricks G, Landmesser U, Ruschitzka F, Wijns W. The best of the European Heart Journal: look back with pride. *Eur Heart J* 2012;**33**:1161–1171.
2. Sipido KR. What does the future hold for Cardiovascular Research? *Cardiovasc Res* 2013; **97**:1–3.
3. Grieneisen ML, Zhang M. A comprehensive survey of retracted articles from the scholarly literature. *PLoS One* 2012;**7**:e44118.
4. Announcement: reducing our irreproducibility. *Nature* 2013;**496**:398.

5. Unsöld B, Kaul A, Sbroglio M, Schubert C, Regitz-Zagrosek V, Brancaccio M et al. Melusin protects from cardiac rupture and improves functional remodelling after myocardial infarction. *Cardiovasc Res* 2014;**101**:97–107.
6. Howells DW, Sena ES, Macleod MR. Bringing rigour to translational medicine. *Nat Rev Neurol* 2013. doi: 10.1038/nrneurol.2013.232.
7. Drummond GB, Vowler SL. Show the data, don't conceal them. *J Physiol* 2011;**589**: 1861–1863.
8. Zong NC, Li H, Li H, Lam MP, Jimenez RC, Kim CS et al. Integration of cardiac proteome biology and medicine by a specialized knowledgebase. *Circ Res* 2013;**113**:1043–1053.
9. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL et al. Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods* 2013;**221C**:92–102.
10. Quinn TA, Granite S, Allesie MA, Antzelevitch C, Bollensdorff C, Bub G et al. Minimum Information about a Cardiac Electrophysiology Experiment (MICEE): standardised reporting for model reproducibility, interoperability, and data sharing. *Prog Biophys Mol Biol* 2011;**107**:4–10.
11. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010;**8**: e1000412.