PRP Pharmacology Research & Perspectives
Open Access

ORIGINAL ARTICLE

Helping to drive the robustness of preclinical research – the assay capability tool

Katrina Gore & Phil Stanley

Research Statistics, Clinical Research, Pfizer, Cambridge, United Kingdom

Keywords
Experimental design, replication, reproducibility, robustness

Correspondence
Katrina Gore, The Portway Building, Granta Park, Great Abington, Cambridge CB21 6GS, United Kingdom. Tel: +44 1304 648044; E-mail: Katrina.Gore@Pfizer.com

Received: 11 June 2015;
Accepted: 12 June 2015

Pharma Res Per, 3(4), 2015, e00162, doi: 10.1002/prp2.162

doi: 10.1002/prp2.162

The article originated from discussions emanating from the Pharmacology 2014 meeting on London (December 2014) at which Pharmacology Research and Perspectives hosted a symposium on replication in research http://www.bps.ac.uk/SpringboardWebApp/userfiles/bps/file/Meetings/meetings%202014/Pharmacology%202014/Pharmacology%202014_Reduced.pdf.

Abstract

Numerous articles in Nature, Science, Pharmacology Research and Perspectives, and other biomedical research journals over the past decade have highlighted that research is plagued by findings that are not reliable and cannot be reproduced. Poor experiments can occur, in part, as a consequence of inadequate statistical thinking in the experimental design, conduct and analysis. As it is not feasible for statisticians to be involved in every preclinical experiment many of the same journals have published guidelines on good statistical practice. Here, we outline a tool that addresses the root causes of irreproducibility in preclinical research in the pharmaceutical industry. The Assay Capability Tool uses 13 questions to guide scientists and statisticians during the development of in vitro and in vivo assays. It promotes the absolutely essential experimental design and analysis strategies and documents the strengths, weaknesses, and precision of an assay. However, what differentiates it from other proposed solutions is the emphasis on how the resulting data will be used. An assay can be assigned a low, medium, or high rating to indicate the level of confidence that can be afforded when making important decisions using data from that assay. This provides transparency on the appropriate interpretation of the assay's results in the light of its current capability. We suggest that following a well-defined process during assay development and use such as that laid out within the Assay Capability Tool means that whatever the results, positive or negative, a researcher can have confidence to make decisions upon and publish their findings.

Abbreviations

ARRIVE, animal research: reporting in vivo experiments; BJP, British journal of pharmacology; PR&P, pharmacology research and perspectives; QC, quality control; SOP, standard operating procedure.

Introduction

It is hard to pick up a recent copy of Nature, Science, Pharmacology Research and Perspectives (PR&P), or many biomedical research journals without seeing an article on the issue of nonreproducible research. They all acknowledge that research is plagued by findings that are not reliable and cannot be reproduced. The pharmaceutical industry is not immune to these issues. Replication of published research findings is a key component of drug target identification and provides confidence to progress internal drug projects. Additionally, in-house in vitro and in vivo assays are used to generate data to assess the biological and pharmacokinetic activity, selectivity, and safety of novel compounds and make decisions which impact their progression toward nomination for clinical development and these assays need to be reliable and reproducible.

Attrition is a key concern in the pharmaceutical industry. As Hayes (2015) highlights in her article in this

themed section of PR&P, the most common reason that compounds fail in clinical trials is a lack of either efficacy or safety (toxicology and clinical safety), which then brings into question the preclinical data that were generated to support the compound's transition into the clinic in the first place. As suggested by Curtis and Abernethy (2015b), it is unrealistic to think that there was no preclinical efficacy and safety data generated, so this can be discounted as a cause. It is almost certainly due to one or more of the following: (1) a lack of translation of the preclinical models to the clinical setting, (2) preclinical experiments that do not address the relevant scientific questions or (3) poorly designed and conducted experiments that are incapable of generating reliable results and conclusions. Issues that undermine scope for translation have been repeatedly highlighted in the literature (Rice et al. 2008; Macleod et al. 2009; Levin and Danesh-Meyer 2010; van der Worp et al. 2010) and are not discussed in this article. However, as a 2014 Nature article by Peers et al. (2014) suggests, the issue of translation can only properly be addressed once the preclinical data is generated reproducibly from trustworthy assays which have quality built in. The focus of this article is to highlight a simple tool that was created to address the issues of reliability, replication, and reproducibility of preclinical assays to ensure that failure in the clinic cannot be attributed to poor decisions made from misleading preclinical data.

It is fair to say that scientists never intentionally run inadequate experiments. However they do occur, in part, as a consequence of inadequate statistical thinking in the experimental design, conduct, and analysis. The ideal solution would be to involve a statistician in every scientific experiment, but in reality the ratio of preclinical statisticians to laboratory scientists is low in the pharmaceutical industry and even more so in academic institutes. Laboratory scientists involved in research are expected to be multiskilled: generating scientific hypotheses, designing, and conducting experiments and analyzing and interpreting the resulting data. Peers et al. (2012) acknowledge that expert statistical input is currently underutilized in the pharmaceutical industry in preclinical studies and its systematic adoption could address the issues of robustness, reproducibility, and quality of preclinical research. It is not realistic to assume that every organization can hire groups of preclinical statisticians and therefore additional tools are necessary to ensure the statistical rigor and validity of preclinical assays.

Within Pfizer, the long-running collaboration of preclinical statisticians and scientists has resulted in the Assay Capability Tool; thirteen questions that accord with new guidelines published for authors of PR&P (Curtis and Abernethy 2015a) and British Journal of Pharmacology (BJP) (Curtis et al. 2015) and guide the development and use of fit for purpose in vivo and in vitro assays.

Table 1 contains the 13 questions and explains why each is important to consider when developing a new assay or adapting an assay described in the scientific literature. The questions provide the scientist with the initial guidance that an experienced statistician would offer and focuses the scientist on two key aspects: the statistical requirements necessary to generate sound data and the capability of those data to influence the decisions made during the research process, for example, progression or otherwise of new compounds for a pharmaceutical company.

The Assay Capability Tool ensures the focus of scientists is placed not only on established good scientific practices but also on the well-known good statistical experimental design principles outlined by Sir David Cox (1958) that are still applicable to modern experimentation. These state: "the requirements for a good experiment are then that the treatment comparisons should as far as possible be free from systematic error, that they should be made sufficiently precisely, that the conclusions should have a wide range of validity, that the experimental arrangement should be as simple as possible and finally that the uncertainty in the conclusions should be assessable." In practice this means emphasizing the importance of identifying and quantifying sources of systematic and random variation in the assay system and using well-established statistical techniques, such as randomization, blocking, and blinding, in the design and conduct of the assays to minimize or eliminate uncertainty and bias. A natural extension requires an assay to be capable of producing reproducible results when it is used repeatedly over long periods of time by different scientists. The majority of questions focus on these good statistical principles.

However, the primary objective for research in the pharmaceutical industry is the identification and validation of new targets and compounds. The data obtained from experiments must enable reliable decisions to be made regarding the future direction of a drug project, either to progress into human clinical studies or to stop further research without regret. Very little focus has been given to this critically important part of the drug discovery process and a more rigorous quantitative approach is needed that moves beyond the standard questions given in recently published guidelines (Kilkenny et al. 2010; Curtis and Abernethy 2015a; Curtis et al. 2015) and checklists (Peers et al. 2014). That is why questions 1 to 3 force both the scientist who is conducting the assay and the project team who will make decisions from the data, to focus on such important information as: what size of effect actually matters; how precisely that effect needs to

Table 1. The assay capability tool.

Question to consider	Why it is important
Aligning assay capability with research objectives	
Q1: Are the scientific objectives for running the assay recorded in a protocol/SOP?	The scientific questions to be answered, the measurements to be obtained and analysed along with their required precision (as defined by, e.g., a standard error or confidence limits) must be stated in the protocol/standard operating procedure (SOP) to prevent data dredging and misinterpretation of the results
Q2: What will a successful assay outcome look like in order to guide decision making?	Prespecifying decision criteria leads to crisp decisions and ensures unbiased interpretation of results. State the primary endpoint and state the minimum response or effect required. As all assay results include inherent uncertainty, it is also necessary to state the level of uncertainty that can be tolerated for acceptable decision making
Q3: Is the experimental design, as described in the protocol/SOP, aligned closely with the objectives?	The design and conduct should be addressed in light of the objectives. Once the objectives and definitions of success are defined the assay must be designed so that the analysis can deliver the objectives. Consultation with a statistician is highly recommended if at all possible
Enabling assay capability by managing variation	
Q4: Are the assay's development and validation fully documented?	Describe the work done in order to verify that the assay is fit for purpose. Identify key learnings/issues/concerns arising from experiments done during assay development. Assay developers should document validation runs using positive and negative controls and standard compounds to provide benchmarks and reassurance to the users of the resulting data
Q5: Have the sources of variability present in the assay been explored?	All assays exhibit variability and it is important to know what the sources of variability are and their relative sizes. The major sources of variation and the statistical methods that will be used for their control should be summarized in the assay protocol/SOP. Understanding and controlling the sources of variability in an assay are critical to achieving the required precision as captured in the standard errors and confidence intervals for the key endpoints
Q6: Is the proposed sample size/level of replication fit for purpose?	An assay that enables a crisp decision requires sufficient, but not excessive, precision. Sample size should always be based on what is known about the assay's variability in the laboratory where it will be run and the quantitative definition of what a successful assay outcome will look like. Relying on historical precedent or published values should not be the default strategy
Q7: Is there a comprehensive protocol/SOP detailing study objectives, key endpoints, experimental design, methods of analysis, and a timetable of activities?	A comprehensive assay protocol/SOP supports efficient decisions by specifying the methods to be used to control variation (e.g., randomization, blocking, use of covariates, and blinding). It helps to ensure uniformity in assay execution resulting in assay results that are reproducible and comparable from one run to another. It promotes transparency by documenting the actual conditions of the assay
Q8: How is assay performance monitored over time? What is the plan for reacting to signs of instability?	Repeated assay use should be tracked to detect changing conditions that may affect the interpretation of the results and to understand the natural variability in the assay. Quality control (QC) charts are useful to monitor the consistency of controls or standards over time. Ongoing monitoring is necessary to understand any changes and their implications for interpretation of the results and to trigger remediation when necessary
Objectivity in assay conduct	
Q9: Are inclusion/exclusion criteria for the assay specified in the protocol/SOP?	Criteria for the inclusion/exclusion of animals, cells, plates etc. in an assay should be predefined and clearly stated in the protocol/SOP. This ensures all the appropriate data are collected and eliminates selection bias
Q10: Is the management of subjectivity in data collection and reporting defined in the protocol/SOP?	There is a need to ensure that the scientist remains unaware of the treatment applied to the experimental unit. Even when the assay measurement is obtained automatically without human intervention there is possibility for bias. The use of randomization and blinding is highly recommended. Studies of a long duration should be blocked to ensure that no bias is introduced by changing conditions over time
Q11: If the raw data are processed (e.g., by summarization or normalization) prior to analysis, is the method for doing this specified in the study protocol/SOP?	Methods of processing raw data prior to statistical analysis should be clearly stated in the assay protocol/SOP. For example, is it the raw response data, change from baseline or log transformed data that are to be analyzed; or are the raw data summarized into an area under the curve or average? This ensures that assay methods and results can be reproduced and validated
Q12: Are rules for treating data as outliers in the analysis specified in the protocol/SOP?	Rules for treating data as outliers should be clearly stated in the assay protocol/SOP. Rules should be in place for the removal of individual data points, whole animals/plates and dose groups as required. This ensures all the appropriate data are analyzed and eliminates selection bias
Q13: Is the analysis specified in the study protocol/SOP? Is it fit for purpose?	The statistical analysis must reflect the study design and assay objectives. Inappropriate statistical analyses can result in misleading conclusions and a false sense of precision. Consultation with a statistician is highly recommended if at all possible

be estimated; and is the experiment designed appropriate to actually deliver answers to the key questions.

The Assay Capability Tool details an assay's current capabilities and limitations. The questions are grouped into three domains: decision making (questions 1–3); managing variability (questions 4–8); and objectivity during conduct (questions 9–13). Each can be assigned a low, medium, or high rating to indicate the level of confidence that can be afforded when making important decisions using data from that assay. An assay will generate data irrespective of the use of the tool, but using it and rating the domains provides transparency on the appropriate interpretation of the assay's results in the light of its current capability. For example, a newly developed assay may have lower ratings than a well-established one simply because less is known about the sources of variation and the assay's reproducibility. Therefore data from that assay should be down weighted accordingly in the decision-making process. This does not mean that assays with low ratings are 'bad' assays, rather it is an honest assessment of the assay's capability to deliver reliable results in its current state and as such the tool will help to identify a path forward for improvement.

Added Value of the Assay Capability Tool

In 2010 the ARRIVE (Animal Research: Reporting In Vivo Experiments) guidelines (Kilkenny et al. 2010) were first published aiming to improve preclinical reporting standards and were widely endorsed by funding agencies, publishers, and their journals. However, Baker et al. (2014) analyzed papers in Nature and PLOS journals in the two years before and after the guidelines were published, and suggest there had been little improvement in reporting standards and that authors, referees, and editors generally are ignoring guidelines.

However, more recent announcements such as the new PR&P (Curtis and Abernethy 2015a) and BJP (Curtis et al. 2015) author guidelines, the new author checklist for Nature publications (Nature 2013), the US National Institutes of Health implementation of new training modules, checklists, and transparency requirements (Collins and Tabak 2014) and the new statistics board of reviewing editors at Science (McNutt 2014) give hope that the reliability, replication, and reproducibility issues are now widely recognized and that measures are being put in place to tackle them. So is the Assay Capability Tool just another checklist being added to an ever growing list of potential solutions?

It has become clear to us during our internal implementation and through discussions with statisticians and scientists at external conferences, that there are few, if any, solutions that address not just the data generation and analysis process, but also the decision-making process. The novel aspect that sets this tool aside from all the other guidelines and checklists is its focus on how the resulting data will be *used,* as well as how it will be generated. It asks the scientist to take a step back before starting their experiment to assess the assay's ability to meet the needs of the research project: does this assay enable good decision making and ensure unambiguous interpretation of the results? Then once convinced of the value of the experiment, they can implement the remainder of the tool to conduct a well-designed experiment, followed by an appropriate analysis which leads to unambiguous conclusions that add real value to the research project. So it is not just another a checklist of statistical design principles, but a very practical document intended to quantify the capability of the assay to deliver data that can be used with the appropriate level of confidence in any decision making, regardless of whether the output is intended for publication or not. This will hopefully reduce the overemphasis of results from assays which are being used without understanding the limitations of the generated results.

The value of the Assay Capability Tool is becoming recognized externally. It was first shared in a paper by Miranda et al. (2014) when describing the development of a preclinical physiological assay to test modulation of knee joint pain in the spinal cord. That paper contained a case study of its use and was a highlighted publication on the National Centre for the 3Rs September e-newsletter (http://www.nc3rs.org.uk/nc3rs-e-newsletter-september-2014). The ABPI cited it in a new members' guide to the Concordat on openness in animal research as an example of a successful collaboration of scientists and statisticians. The Royal Statistical Society has also recognized the team behind its development and implementation as joint winners of their 2015 award for Statistical Excellence in the Pharmaceutical Industry.

The Assay Capability Tool does not replace the statistician, but it can help to facilitate the required discussions between scientists, statisticians, and decision makers. Although it was designed for preclinical research scientists within Pfizer, the tool can be applied to any assay. We consider that this tool provides a robust method of minimizing the risk of nonreproducible research. It can improve the robustness of preclinical research and is central to embedding statistical excellence and quantitative decision making into all scientific research.

When reflecting specifically on the issue of replication highlighted within this issue of PR&P, it is clear that focus must be placed on the quality and integrity of the assay(s) underpinning any publication. We suggest that following a well-defined process during assay development and use such as that laid out within the Assay

Capability Tool means that whatever the result, positive or negative, the researcher can have confidence to publish their findings.

Disclosure

None declared.

References

Baker D, Lidster K, Sottomayor A, Amor S (2014). Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. PLoS Biol 12: e1001756.

Collins FS, Tabak LA (2014). Policy: NIH plans to enhance reproducibility. Nature 505: 612–613.

Cox DR (1958). Planning of experiments. Wiley, New York, NY.

Curtis MJ, Abernethy DR (2015a). Revision of instructions to authors for pharmacology research and perspectives: enhancing the quality and transparency of published work. Pharmacol Res Perspect 3: e00106.

Curtis MJ, Abernethy DR (2015b). Replication – why we need to publish our findings. Pharmacol Res Perspect DOI:10.1002/prp2.164

Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Giembycz MA, et al. (2015). Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol 172: 3461–3471.

Hayes A (2015). Key role of publication of clinical data for target validation. Pharmacol Res Perspect DOI:10.1002/prp2.163

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol 8: e1000412.

Levin LA, Danesh-Meyer HV (2010). Lost in translation: bumps in the road between bench and bedside. J Am Med Assoc 303: 1533–1534.

Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. (2009). Good laboratory practice: preventing introduction of bias at the bench. Stroke 40: e50–e52.

McNutt M (2014). Raising the bar. Science 345: 9.

Miranda JA, Stanley P, Gore K, Turner J, Dias R, Rees H (2014). A preclinical physiological assay to test modulation of knee joint pain in the spinal cord: effects of oxycodone and naproxen. PLoS ONE 9: e106108.

Nature Editorial (2013). Announcement: reducing our irreproducibility. Nature 496: 398.

Peers IS, Ceuppens PR, Harbron C (2012). In search of preclinical robustness. Nat Rev Drug Discovery 11: 733–734.

Peers IS, South MC, Ceuppens PR, Bright JD, Pilling E (2014). Can you trust your animal study data? Nat Rev Drug Discovery 13: 560.

Rice AS, Cimino-Brown D, Eisenach JC, Kontinen VK, Lacroix-Fralish ML, Machin I, et al. (2008). Animal models and the prediction of efficacy in clinical trials of analgesic drugs: a critical appraisal and call for uniform reporting standards. Pain 139:243–247.

van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. (2010). Can animal models of disease reliably inform human studies? PLoS Med 7: e1000245.