Frameworks for evaluation and integration of data in regulatory evaluations: The need for excellence in regulatory toxicology

Colin Berry

Abstract
Following a number of published expressions of concern about the reliability of experimental science and the implications of non-reproducibility for regulatory toxicology, the European Risk Forum undertook to consider what practises might improve the basis on which regulatory decisions might be made. Guidelines which may be useful in assessments are presented. The document acknowledges the value of the experimental standards used in most regulatory studies but indicates how these may fail to provide the ‘best outcome’ and how imperfect studies based on outdated views of pathophysiology of disease in H. Sapiens may offer little of predictive value.

Keywords
Regulatory toxicology, pathophysiology of disease, reliability of experimental science, non-reproducibility

Introduction
Most toxicologists will have been involved in some aspect of regulatory toxicology, either as the generators of studies in companies, in the interpretation of these studies in regulatory agencies, or in the discussion of the findings and their merits and demerits as independent members of regulatory bodies or as consultants. At the complex interface between what is desirable to inform the public about the ‘safety’ of new xenobiotics and the probable predictive value of the information obtained by testing, many will feel there is an uncomfortable interaction between what is expected of, and what is possible in, testing regimes. Despite our best endeavours, it remains the fact that there are still compounds which have to be removed from use due to unexpected adverse effects. In academic toxicology, the value of many experimental observations will be informative for those who provide regulatory advice to legislators but the nature of experimental produces difficulties in the use of non-protocol controlled studies in evaluations.

The purpose of Regulatory Toxicology is to provide data which enable those responsible for the regulation of the use of pharmaceuticals, agrochemicals, medical devices and chemicals widely used in the environment, to make rational decisions about potential effects of exposure to H. sapiens and to many other species to be made. In many ways, regulatory studies fail to reach the standards required of ‘good science’ for reasons that will be touched on in this document but it is important to realise that there is no aspect of the scientific method that is free of difficulties if the results of any investigation are to be used in a way that extends the primary purpose of an experiment, which is an attempt to justify the hypothesis tested. This basic failure in the use of the scientific method underlies many of the difficulties in the performance and interpretation of regulatory studies.

1 Professor Emeritus of Pathology, Queen Mary and Westfield College, London, UK

Corresponding author:
Colin Berry, Professor Emeritus of Pathology, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.
Email: colin@sircolinberry.co.uk
For many commentators, experimental science is assumed to be a system that provides answers. So some were surprised when in 2013, the front page of The Economist magazine read ‘When Science Goes Wrong’. The accompanying editorial comment was critical of biological sciences, a criticism based on the view that venture capitalists have come to assume that only 50% of academic study results can be replicated. A little before this Begley and Ellis had reported that a study of 53 ‘landmark’ papers in oncology found that only 6 were reproducible, even with co-operation of the original authors.

To those who have spent their lives in experimental biological sciences these observations were not startling. Nevertheless, the questions raised produced a flurry of activity in major journals relating to reproducibility in science. It is not the purpose of this text to rehearse the arguments but it is important to note that it is never easy to repeat an experiment. The work reported by Lithgow et al. makes clear the difficulties involved harmonising results between different laboratories using an invertebrate (planarian worms) with a single-celled food source (bacteria). Whole animal studies with complex diets represent a different scale of problems.

The lack of reproducibility in test results in regulatory toxicology has received little attention. The physiologist Sampson Wright coined the epigram ‘a drug is a substance which when injected into a rat, produces a paper’ and unhappily, much of regulatory toxicology has no better-defined aim, if the notion of an experiment being the test of an hypothesis is taken as the mark of good design. Regulatory toxicology has a bias towards the production of effects rather than the acceptance of a null effect. Studies rarely test mechanisms of the production of an effect; this may follow the demonstration of a change; post hoc rather than propter hoc. Critical reviews of current systems are uncommon and when made may make uncomfortable reading (see Smith and colleagues for a detailed consideration of the NCI studies).

It is thus very important that when regulatory studies are performed that they should conform to a number of critical points in terms of design, execution, record keeping and reporting. These have been established by virtue of decades of experience relating to numbers of animals used to enable a statistically significant result to be arrived at, standards of accommodation, feed characteristics, environmental controls and so on.

The purpose of this document is to identify critical factors in the constraints that should apply to regulatory toxicology and makes clear the reasons why non-compliant studies should not be part of the regulatory process, although they may be of scientific value and point to a need for further properly designed work. It is possible to define what constitutes a good study in this context; those requirements make clear why non-compliant studies should not be made part of the regulatory decision-making process.

In any regulatory assessment a Systematic Review of all available data should take place and form part of the data base on which the regulatory decision is made. This review should also follow an approved methodology such as that of the Cochrane Collaboration. In the same way, where epidemiological studies are included in assessments, they should conform to agreed standards. The appendix to the document is derived from the recommendations of the Netherlands Epidemiological Society, a very comprehensive document.

www.nature.com/sdata/policies/data-policies
www.sciencemag.org/authors/science-journals-editorial-policies

Scientific integrity in regulatory studies – Principles and guidelines

1. Introduction

1.1. This text focuses on:

- Scientific assessments used to inform risk management decisions made by governments to provide a high level of protection of human health and the environment;
- The nature of the evidence assessed for this purpose, provided primarily by the findings of regulatory toxicology and epidemiology; and,
- The role of these data in safety testing of new or existing substances, technologies or materials, as well as experimental, ‘investigative’ studies in identifying and evaluating new hazards or in challenging the existing body of risk management knowledge.

1.2. Objectives

- To provide a set of principles and guidelines that, if implemented properly, will help strengthen the integrity, quality, and consistency of scientific assessments used as part of the process of public management of technological risks; and,
- To ensure that opinions derived from scientific assessments are based on the available body of relevant, reproducible, and testable evidence provided by toxicology, and related fields of scientific endeavour, enabling risk managers to base decisions on reliable evidence.

1.3. Challenges

- The complexity of the risks posed by technologies to human health or the environment means that there is rarely a single study or determinative experiment that is capable of resolving all risk management
issues. High quality decisions require the aggregation of multiple sources of evidence.

- In most jurisdictions, risk managers are required to consider all potentially relevant studies. This requirement and the evaluation of what is relevant, poses problems for ensuring scientific integrity.

- Modern industry-funded safety research must satisfy quality standards and controls defined by regulators and must comply with the demanding standards laid down in internationally accepted guidance (e.g. OECD, ICH) – the extent of compliance of these guidelines defines the quality of such studies and assists in the assessment of their value.

- Many experimental, ‘investigative’ studies do not comply with accepted regulatory guidance and internationally accepted standards – a problem increasingly recognised by leading journals and universities. These studies may inform about potential mechanisms of harm but their lack of reproducibility means that they are not useful within the regulatory decision-making process.

- ‘Open Science’ publishing may offer the possibility of speeding up the availability of findings from scientific research. At its best, it makes freely available high-quality studies that have undergone rigorous review (so-called ‘open source’ publishing). However, at its worst, it can trigger waves of social concern about alleged hazards identified by low quality or misleading studies that have not been reviewed independently or replicated.

- Out-of-date studies, some several decades old, continue to influence risk perceptions and hence regulatory interventions. Many older studies fail to reflect modern scientific understanding or standards, or may have been discredited or even retracted. In many instances such studies are unreliable and the raw data are untraceable.

- Questionable research practices are increasingly evident, and major journals (‘Science’ and ‘Nature’) have emphasised the need for documentation of protocols so as to enable studies to be replicated. Errors include outcome reporting bias, selective reporting of research findings, protocol deviations not clearly described, data dredging and citation bias.

- Some traditional bioassays (studies in whole animals), assessing chronic exposures or multiple modes of action or long-term hazards, lack scientific validity. Progress in scientific understanding of the mechanisms of adverse effects, notably in human carcinogenesis, has meant that regulators should exercise caution when considering evidence derived from a methodology that does not reflect the current understanding of the pathogenesis of particular events.

1.4. Coverage

This text sets out principles and guidance in four areas:

- Study Quality (see Section 2)
- Assessment of Studies (Section 3)
- Communication of Scientific Opinions to Risk Managers (Section 4)
- Selection of Experts (Section 5)

2. Study quality

2.1. Principles

- All high-quality studies must meet the basic precepts of the scientific method;
- Study design must be relevant and thereby able to answer the specific question posed by regulations or regulators; and,
- Premature, experimental studies that are not sufficiently tested and controlled should not form part of the body of data used in regulatory scientific assessments.

2.2. Guidelines

- The study is conducted following a well-defined protocol – the protocol refers to specifications for the research process;
- The methodology follows appropriate standards applicable to the field of study – such as testing guidelines (ICH, OECD, US EPA, ECHA) and GLP standards;
- The results are relevant and applicable to the hypothesis being tested – conclusions answer the hypothesis. The results are not used to propose hypotheses that were not part of the initial research project;
- The study is designed and reported in such a way that anyone can repeat it using the same methodology and materials. Sufficient detail, so that others can repeat the study, should be part of the protocol and be included in the methods section;
- The study should include a Systematic Review of previously produced related research – the evidence and conclusions are seen in the context of the existing body of evidence on the topic studied;
- All data generated must be critically analysed and the weight of the data generated considered, even if outliers are subsequently excluded. The evidence should be analysed critically, as opposed to a simple data collection;
- All data used in the analysis should be available to any researcher for the purpose of reproducing or extending the analyses;
- A materials and methods section should be included in the study findings – it provides sufficient detail to allow replication of the study;
3. Assessment of studies

3.1. Principles

- The statistical methods used are described with enough detail to enable a qualified reader with access to the original data to verify the results – disclosures of statistical methods meet the standards set out by leading journals, such as ‘Nature’ and ‘Science’;
- All data and the protocol are deposited in an approved repository – it is made publicly available without restriction, excepting reasonable controls related to human privacy or biosafety, and respecting relevant data protection laws;
- The conclusions should be supported by the data gathered, analysed and reported – they are not based on anything other than those data gathered and analysed as part of the study. Conclusions are well founded, based on relevance of the experimental design, statistically significant evidence and causality, when applicable;
- The study has been opened to expert scrutiny – peer review increases the probability that a study is properly conducted and conclusions are based on a credible, reasoned interpretation of the data generated by the study;
- The specific hypothesis and appropriate research methodology are disclosed and clearly explained – the hypothesis sets out the purpose of the research which should form the basis of all scientific activity. Hypotheses are based on scientifically plausible scenarios that a certain effect may occur;
- Funding, affiliations, and additional interests of authors should be disclosed – transparency allows other scientists, policy-makers and the public to better understand the motivation behind a study, as well as the context in which it was performed;
- Novel hypotheses or non-validated methodologies should not influence findings of assessments unless supported by compelling scientific evidence;
- Further tests should only be sought by assessors if it is clear that the results will be relevant to the scientific assessment;
- Assessments should not address or be influenced by economic, social, ethical or other non-scientific factors when characterising risks to human health, public safety or the environment;
- Use of the Precautionary Principle should be limited to the selection of risk management measures. It should not inform or shape assumptions, defaults, methods or procedures used in assessments of scientific studies. Interpretation and use of the Precautionary Principle should follow the European Commission’s Communication12;
- Unless mandated by legislation, findings from assessments should not explicitly recommend or include risk management measures;

3.2. Guidelines

- The assessment methods and procedures correspond to best international practices and accepted standards;
- A Systematic Review is performed to agreed standards (those of the Cochrane Collaboration, for example) to assess quality and relevance of all reliable studies that could inform the outcome of the assessment. This includes all positive and negative studies, and may be a legal requirement;
- The Review gathers all potentially relevant studies; provides a transparent basis for excluding low quality or irrelevant studies; and ‘scores’, using agreed and pre-stated criteria, studies that will form part of the weight-of-evidence review. Assessment of study ‘quality’ recognises compliance with regulatory guidelines; agreed standards of quality for investigative studies; and the ‘power’ of the journal within which an investigative study is published;
- A weight-of-evidence review is then undertaken, using the studies identified by the Review. It examines all relevant and high quality positive and negative studies that meet pre-determined criteria for selection. It meets predetermined standards of quality; and, it is undertaken in a transparent manner, including the provision of a clear formulation of methodology. Assessments based on the weight of available scientific evidence ensure that individual studies of questionable quality or reproducibility do not have a disproportionate impact on risk evaluation or mitigation measures;
- The weaknesses of old or superseded studies are recognised explicitly in the weight-of-evidence review. It treats evidence from older long-term chronic exposures, multiple modes of action or the long-term hazards derived with caution, and an appropriately conservative weight is applied to them. A similar approach is taken to studies that are not compliant with relevant protocols such as GLP. It recognises the limitation of correlation as opposed to causation based on a plausible mechanism of action;
- Relevant uncertainties are systematically identified, analysed and documented;
• The findings of assessments must be consistent with all available high-quality relevant data and knowledge, including positive and negative findings;
• The assessments and their findings must be understandable to experts and be reproducible.

4. Communication of scientific opinions to risk managers

4.1. Principles
• Communication of the findings of assessments to risk managers should be understandable, clear, and supported by the data gathered;
• High quality communication contributes to transparency and public trust in risk analysis: characteristics of good regulatory governance;
• Sufficient explanation and evidence should be provided to enable a similarly qualified and equipped scientist to reproduce the findings and conclusions.

4.2. Guidelines
• The communication of the findings of assessments ensures that risk managers are fully aware of the meaning of scientific advice; methodology and evidence on which conclusions are based; and, limitations of the validity of conclusions, including relevant uncertainties;
• The overall reporting ensures that there is transparency in all aspects, including data, study design, information, calculations, assumptions, and methodologies;
• The strategies and processes for identifying and acquiring studies, information, and data are documented and transparent;
• The criteria used for critically evaluating studies, data and information, along with their application, are fully explained and transparent;
• The limitations related to the data, studies, and information used in the assessment are explained, and gaps in the state of scientific knowledge are highlighted;
• The evidence and expert judgement are properly presented, explained, and documented, including methodologies used to reconcile inconsistencies in scientific data;
• The limitations of novel hypotheses or non-validated methodologies are acknowledged and documented;
• The reporting of uncertainties avoids hypothetical speculation, recognises that uncertainty is inherent to the nature of scientific evidence, and identifies resolvable issues that lie within the scope of existing requirements and require further investments in science. Where appropriate contextual information is provided;
• Sufficient information is provided on the data, information, and studies to allow a clear understanding of the rationale of the opinion;
• Any dissenting opinions are noted and reported, along with an accompanying rationale;
• New evidence that might alter the conclusions reached in the assessment is highlighted;
• Value judgements are avoided, including the framing of risks and commentaries on the social or political acceptance of risk, and the opinion focuses solely on scientific evidence and scientific advice.

5. Selection of experts

5.1. Principles
• The primary objective of any selection process is to ensure that the best available experts undertake scientific assessments. They should meet accepted standards for the determination of their expertise and the relevance of that experience to the issues to be considered.
• Bias, or the failure to act impartially and in the public interest, can result from conflicts of interest. These are multiple and encompass materialistic factors (such as financial gain), beliefs and ideologies, political affiliations, and personal factors, including ambition, family history power and status. They are part of the human condition.
• Appropriately qualified experts should not be excluded from joining scientific committees or panels simply because they have one or more demonstrable conflict of interest.
• Rigorous, fair, and transparent processes should be employed to identify and disclose all forms of material conflict of interest that are likely to be relevant to the specific work of the expert group, committee, or panel.
• Genuine scientific disagreement, if based on well-founded scientific evidence, does not constitute a conflict of interest. Evidence of intellectual debate and differences of opinion are part of the scientific process but so is the resolution of these difficulties in the light of new evidence.
• Stakeholders of all types should be encouraged to make use of high-quality scientific evidence and advice when informing their respective positions and this information should be made available to inform broader societal debates.
• Undertaking paid work for industry or for activist groups (or research institutes that pursue a specific social or political agenda) is not, on its own, grounds for exclusion from serving on advisory groups, panels or committees.

5.2. Guidelines
• Committees or panels should be institutionally independent and separated from political influence;
• Committees or panels are constituted so as to ensure that decision-makers have access to an appropriate range of relevant different types of scientific
expertise from different scientific disciplines and relevant practical technical expertise;
• As a general rule, committees or panels undertaking scientific assessments seek to manage conflicts of interest rather than exclude appropriately qualified experts;
• Experts are only excluded from specific scientific assessments if one of the two following conditions are met: (i) there is clear and substantial evidence of predetermination; or, (ii) there is a credible likelihood of direct, material financial gain;
• Experts selected to carry out scientific assessments commit formally to act impartially, and in the public interest;
• Whilst respecting intellectual debate and commercial confidentiality, there is a presumption of openness throughout the process;
• Outcomes of scientific assessments are subject to independent peer review. All draft assessments should be reviewed procedurally, whilst significant assessments should be subject to an additional substantive review.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes
1. In general, this term refers to studies designed to test hypotheses that are not performed to agreed international protocols.
2. Scientific assessments combine a critical evaluation of evidence (including data) with expert judgement. They are evaluations of a body of scientific or technical knowledge that typically synthesises multiple factual inputs, data, models, assumptions and best professional judgements to bridge uncertainties in the available information. Such assessments include but are not limited to state-of-science reports; technology assessments, weight-of-evidence analyses, meta analyses, health, safety or ecological risk assessments; toxicological characterisations of substances; integrated assessment models; hazard determinations, or exposure assessments (Source: derived from a definition used by the US Office of Management and Budget).
3. The scientific method is an empirical method of acquiring knowledge. It involves careful systematic observation and experimentation, inductive and deductive reasoning and the formation and testing of hypotheses. Rigorous scepticism about what is observed must be applied, as cognitive assumptions may affect the interpretation of the observations. For discussion, see Popper KR. The logic of scientific discovery. London: Hutchinson, 1968.
4. This is, for example, a particular problem with findings derived from small-scale animal studies or evaluation of exposure to chemicals at unrealistically high doses.
5. These guidelines apply primarily to investigative studies undertaken to identify new hazards or to challenge the existing body of risk management knowledge. Safety testing undertaken or funded by producers of substances, materials or technologies must comply with legal requirements. Many of these are set out in substantive guidance by regulators or international organisations.
6. Exhaustive information about Good Laboratory Practice (GLP) can be found on the websites of the OECD and the European Commission. These sites define a set of rules and criteria for the manner in which non-clinical health and environmental safety studies are planned, performed, reported, and archived.
7. See, for example, Smith V, Devane D, Begley CM, et al. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011; 11: 15.
8. This is a requirement for studies used to inform regulatory decision-making.
9. This may not be possible in some cases, if studies are given confidential status by regulators, thereby protecting data and intellectual property.
10. Weight-of-evidence is an expert process of collecting evidence, assessing, integrating and weighing it to reach a conclusion, with a predefined degree of confidence, on a particular problem. It meets predetermined standards and considers all relevant evidence, both positive and negative, taking into account factors such as strength, relevance, and quality. It identifies conclusions that are best supported by the available body of evidence. It should always be used when scientific questions can only be answered using several different pieces of evidence.
11. Guidance on the use of weight of evidence approach in scientific assessments is found in EFSA J 2017; 15(8). DOI: 10.2903/j.efsa.2017.4971.
12. European Commission. Communication from the Commission on the Precautionary Principle (2000, COM (2000) 1).
13. This condition is satisfied if there is substantial evidence of personal beliefs, commitments, perspectives, or intensely advocated policy positions that suggest predetermination, where an advisor is committed to a particular point of view and unwilling or reasonably perceived to be unwilling to consider other perspectives or relevant evidence to the contrary (Source: derived from a definition used by the US National Academies of Science).
14. For this condition to be satisfied there must be credible likelihood that direct, material financial gain for the advisor of his/her immediate family or employer, be that industry, academic institution or NGO, will result from a certain opinion.
15. See Swaen GMH, Langendam M, Weyler J, et al. Responsible Epidemiological Research Practice: a guideline
Epidemiological studies

1. Epidemiology is the study of the pattern of disease in humans:
   - Its most effective role is to conduct investigations of occupational and other restricted settings where exposures and externalities are amenable to measurement and to provide tentative clues about causation that other scientific investigations may illuminate;
   - It is important to emphasise that epidemiology identifies associations not mechanisms of causation;
   - The weight that might be attached to these associations varies, and the way in which they are evaluated should follow internationally accepted guidelines.

2. Guidelines:
   - The aim of study is defined in advance and published – it includes a specific hypothesis or research question to be tested;
   - The study is based on a detailed research protocol that describes the study and is deposited publicly prior to commencement – the level of detail allows another study group to carry out the study as intended;
   - The study is carried out in accordance with the protocol and all deviations are documented and reported;
   - The statistical analysis is carried out in accordance with the study protocol – additional analyses undertaken but not foreseen when the study protocol was registered are identified as such in the final report;
   - No changes are made to the raw dataset once the statistical analyses commence;
   - The final report is an accurate, balance, and concise reflection of the study – it also describes its limitations and any deviations from the protocol;
   - All results are properly reported and disclosed – the study is rigorously documented and archived such that a trained scientist, not necessarily an epidemiologist, can reconstruct how the study was documented;
   - Once the report is finalised, the raw data files and final dataset used for statistical analysis is stored securely – a full explanatory data description is stored and the privacy of subjects is protected.