Be CLEAR to ensure methodological and data transparency.

Christopher H. George[1], Simon T. Bate[2], S. Clare Stanford[3], Amrita Ahluwalia[4], Michael J. Curtis[5], Farhad Dehkhoda[6], John Kolassa[7], Vanessa Minervini[8], Vandana Nikam[9], Taiwo O. Omotoso[10], Michael Spedding[11], Debbie L. Hay[12]

[1]Medical School, Faculty of Medicine, Swansea University, UK; [2]CMC Statistics, GlaxoSmithKline, UK; [3]Department of Physiology, Pharmacology & Neuroscience, University College London, UK; [4]Faculty of Medicine & Dentistry, Queen Mary University of London, UK; [5]Department of Pharmacology and Therapeutics, School of Cardiovascular and Metabolic Medicine & Sciences, King's College London, UK; [6]School of Biomedical Sciences, The University of Queensland, Australia; [7]Rutgers, the State University of New Jersey, USA; [8] Department of Psychological Science, Creighton University, USA; [9]S. P. Pune University, India; [10]Department of Clinical Pharmacology, University College Hospital, Nigeria; [11]Spedding Research Solutions, France; [12]Department of Pharmacology and Toxicology, University of Otago, New Zealand

Corresponding author:

Debbie L. Hay, Department of Pharmacology and Toxicology, University of Otago, Dunedin, 9016, New Zealand.

email: debbie.hay@otago.ac.nz

Keywords

Experimental design / methodology / data / reporting / transparency

Abstract (50 words)

A central tenet of research articles is that they should accurately describe the experiments performed. Yet, important aspects of experimental design and methods are sometimes omitted, precluding proper interpretation and follow-on studies. To remedy this, we urge researchers to adopt the CLEAR principle (***CL***arity, ***E***valuation, ***A***ssessment, ***R***igour) when reporting research.

Background and context

Advancements in science require the rigorous application of method, process and insight. Each of these depends on foundational discipline, skill, and technique. However, some elements of research practice can be considered more of a craft (e.g., experimental design) and as such researchers require appropriate education to develop such craftsmanship. Unfortunately, evidence suggests that available experimental design training is suboptimal for many researchers [1]. As a consequence, inappropriate designs can be adopted, producing research outcomes that may not be robust.

In a landmark initiative resonant to this issue, Errington and colleagues found that out of 193 key experiments from 53 'high-impact' papers, the methods sections and supplementary information contained insufficient detail to enable the experiments to be repeated [2,3]. The authors concluded that "original papers do not include enough information about the methodology and results". Hence much effort has focussed on improving reproducibility and reliability, prompting a consensus for change.

As part of this effort, good progress has been made in implementing the principles and practicalities of robust experimental design, statistical analysis and reporting in pre-clinical pharmacology. For example, it is acknowledged that improvements in transparency and reproducibility demand the use of adequately-powered group sizes [4] and that blinding and randomization should be incorporated as core elements of study design and analysis [5,6]. Resources have also been developed to help scientists increase the robustness of their work (e.g., Resource Identification Initiative [7]; SciScore [8]; Experimental Design Assistant [9]; reagent validation [10,11]). These resources have been developed by enthusiastic peer-groups and organisations with a common goal of improving research and promoting best practice. However, the level of uptake by the research community is unclear [12].

As IUPHAR's Transparency and Reproducibility Committee we asked ourselves why, despite increased awareness and freely available resources, poor experimental reproducibility and reliability in pre-clinical research persists. One possible explanation is that some researchers simply do not realise that their research paper fails to include

important descriptions of the work and so defies proper interpretation by others. For example, research papers may omit essential experimental design details such as the sex of animals or concentration(s) of drug used, or fail to include precise descriptions of how specific methodologies were implemented and how the experimental design was actually conducted. Perhaps the 'pressure to publish' forces authors into submitting papers that feature incomplete or poorly conceived datasets which have not been fully explored. Some authors may fail to consider the data from all angles (while fully acknowledging the shortcomings of said approach) leading to conclusions that are unintentionally flawed. In this *Forum* article we raise awareness of these issues and consider the importance of *cl*arity, *e*valuation, *a*ssessment of variability and *r*igour of analysis (*CLEAR*) as foundational elements of reporting pre-clinical research. To help achieve this, we use a hypothetical scenario to illustrate the consequences of omitting some often overlooked elements when reporting research (Figure). The example is intentionally flawed, and we argue these issues only become apparent if the *CLEAR* principles are followed.

Clarity on the experimental design and methodologies

In the scenario described in this article, a laboratory measures hormone X concentration in rats; (i) under baseline conditions, (ii) after the infusion of a vehicle and then (iii) following drug treatment (Figure, panel A), where sequential sampling of each rat under the three conditions implies the drug effect can be assessed against the within-rat variability. A statistical power analysis revealed that N=6 rats should be sufficient to identify any drug related effects. However, to understand reproducibility, the laboratory decided to repeat the experiment three times, and hence 18 rats were ordered and then randomly separated into three sets of N=6 (Figure, panel A). Blinded measurements of the hormone were done each time.

When publishing research, it should be stated whether the experiments are 'exploratory' (pilot or hypothesis-generating) or 'confirmatory' (i.e., test a precise prediction) [13]. While both types of experiment are an important part of any scientific programme of work, for the latter type, the researcher should be able to predict the results, and the study should have been appropriately powered with the false positive/negative risks considered and controlled. For this scenario, as the laboratory is still trying to understand whether any additional sources of variability may affect the reproducibility of their experiments, they should clearly state that this is work is 'exploratory'.

Regardless of whether the experiment is exploratory or confirmatory, detailed reporting of experimental procedures is essential and typically dominates the Methods section of original research papers. Most researchers recognise the value of transparently disclosing the individual results (e.g., all datapoints on a scatter plot), as has been done in

the Figure [14]. However, neither the description of the methods, nor the detail with which the results are shown, address underlying issues on how the experiments were designed and conducted, and how this might influence the conclusions. While some funders and journals encourage publication of the full experimental plan and design in supplements, there is often a lack of transparent explanation of how that plan was configured, delivered, and how the experiments were *actually* carried out.

For the purposes of this scenario, all protocols, materials and reagents used in each experiment are identical and hence any inter- and intra-assay variability should be minimal. Figure panel A describes an intentionally imperfect experimental design; a detailed consideration of how to resolve the limitations of this design are beyond the scope of this article. However, this could include use of a crossover design, where the order of testing the vehicle and drug would be varied across animals to enable any time-related effects to be accounted for in the statistical analysis. Nevertheless, Figure panel A illustrates the ethos of *CLEAR*: only when reported with *clarity* do design shortcomings become evident.

_E_valuation of experimental outcome

The laboratory finds different effects of the drug on the concentration of hormone X in each experiment: no effect (Experiment 1); decrease (Experiment 2); increase (Experiment 3) (Figure, panel B). These seemingly disparate findings cast doubt on the reproducibility of the experimental conclusions and highlight the problem that inter-experimental results, even within the same laboratory, can be different. To better understand their data, as all experiments are conducted the same way and with identical reagents, the three sets of results were pooled together and an assessment of all the data generated was performed. In theory, while this potentially will lead to over-sensitive statistical tests, as it contradicts the initial power analysis choice of sample size, the combined data set (Figure, panel C) reveals that any conclusions based on the statistical analysis of each experimental set of N=6 rats are potentially misleading. Care must also be taken with the statistical test performed. It is important to consider the appropriateness of a paired t-test, given that ‘Experiment’ should be included as a blocking factor in the analysis. Additionally, the variability in the responses to the drug is greater than to the vehicle (Figure, panel B), and therefore a mixed model that allows for unequal variances would be more appropriate.

So how do we explain what is going on? Was the initial power analysis, that identified N=6 per group as sufficient, incorrect? Ideally, the number of experimental units included in a study is calculated using a power analysis which makes assumptions about expected variance and biologically relevant effect size. However, a single pre-clinical research study rarely incorporates a sample size that features every conceivable source of variability (an N of between 5 and 10 is common in publications). The Figure also highlights a potential pitfall

that can occur when randomly sub-sampling (N=6) from a larger (perhaps more appropriate) sample-set (N=18) where the random sampling can result in sub-samples that are not representative of the 'true' effect. These issues also illustrate the need for a careful *evaluation* of all the information contained in a dataset.

***A**ssessment of* the variability

Research will be improved if there is careful and systematic assessment of where the intrinsic and extrinsic variability in an experimental system lies [15]. The type of experimental design will influence both the understanding of the sources of variability and the conclusions that can be drawn from the data generated [13,16]. Since the experimental design and analysis are clearly described in Figure panel A, and many of the variables that could affect the data have been (artificially) reduced, Figure panels B and C should provoke the question of why the effect of the drug is apparently so variable across the three repeated experiments? Were there additional sources of variability influencing the results in different ways across the three repeated confirmatory experiments?

The data in Figure panel B shed light on the underlying relationships between the variables that influence the results but clearly all sources of variability are not accounted for. This further emphasises the need to describe this study as 'exploratory' rather than 'confirmatory' and additional sources of variability should be investigated. The next step is to use these findings to inform subsequent investigations that aim to identify unaccounted sources of variability prior to running a 'hypothesis-confirming' experiment [12].

Finally, Figure panels B and C report absolute values of hormone X concentration for the purposes of transparency. This is the preferred approach because this enables better *assessment of the variability*. Also, any transformation of the raw data to reduce its apparent variability, such as 'normalisation' (percentages or incremental changes), is potentially problematic because such transformed data can be distorted by variability of the comparator values as well as 'thresholding' and 'cut-offs', which increase the likelihood of misrepresenting the real findings of the experiment and the conclusions.

***R**igour* in data analysis

It is standard practice for researchers to report only the effects of an experimental treatment compared to a control/vehicle group. In this context, the laboratory ignored the baseline concentration of hormone X and only reported the results of a paired t-test comparing the 'vehicle' mean to 'drug treated' mean (grey shaded areas, Figure panels B and C). By using a paired t-test the researchers neglect other factors that might influence those treatment or control measurements (i.e., baseline levels) which gives an incomplete picture of the experiment. Since this baseline measurement might fundamentally influence the rats'

response to vehicle or drug, it should be incorporated into statistical analysis and interpretation. To explore this explicitly, Figure panel D plots hormone X concentrations at baseline against measurements following vehicle- and drug-infusion. Visualising the data this way reveals that the starting baseline hormone concentration affects the outcome, something that is not evident from the data reported in Figure panels B and C. This highlights why care must be taken in identifying the best way to visualise data to uncover underlying effects. Figure panels C and D contain the same information and yet the different way they are presented reveals different aspects of the underlying relationships, with Figure panel D showing the baseline/vehicle/drug relationships more effectively. We emphasise that *rigour* in data analysis is essential to understand these effects.

Conclusions

Researchers must provide an unambiguous description of *how* an experiment was conducted and the data analysed; this information is as important as the justification for *why* an experiment was conducted. Original research papers should explicitly disclose sufficient detail so that readers can understand, appreciate and repeat the work. Researchers must ensure that any conclusions acknowledge and contextualise variability and are made only following rigorous scrutiny of the data. When preparing their submission, authors should scrutinise their draft manuscript and task themselves with checking whether they could repeat and interpret their own work, based on what they have written.

CLEAR aims to close the gap on the long-standing problem of poor experimental reproducibility and reliability and we urge researchers to adopt the principles of ***CL***arity, ***E***valuation, ***A***ssessment, ***R***igour. This article illustrates the value of enhanced descriptions of experimental design, methodological approach and data evaluation for gaining a greater understanding of research outcomes.

CLEAR will facilitate robust discussion on the strengths and limitations of published work and should be used to strengthen collaboration between authors, reviewers, journal editors and publishers to improve transparency in research reporting.

214

215 **Table. A *CLEAR* action plan**

What		**Action plan**
CLarity	Transparently explain how an experimental plan was configured, delivered, and *actually* carried out.	• Carefully review methods to check you could reproduce your own work, based on the information provided • Identify experiments as exploratory (pilot or hypothesis-generating) or hypothesis-confirming. • Provide all elements that are integral to data interpretation, using (peer-reviewed) supplementary methods where main manuscript word limits preclude extensive methods descriptions. • Show raw data wherever possible. • As randomisation and blinding of the experimental design & execution is a foundational feature of good practice, assessments of variability / statistical analysis should be blinded wherever possible.
Evaluation	Provide a detailed scrutiny of all the data.	
Assessment	Extensively consider sources of variation and their impacts to accurately interpret the results and inform future experimental planning.	
Rigour	Ensure conclusions are appropriately robust and reflect all elements of the data.	

216

217

218

219

220

221

222

223

Figure. Illustrating the CLEAR principle using a hypothetical scenario and simulated dataset.

A. Sequential blood samples were taken from three sets of N=6 rats under baseline conditions (black), following the infusion of vehicle (green) and after drug challenge (blue). Samples are stored at -80°C until use. The identifier of each rat is given in orange.

B. The concentrations of hormone X in baseline, vehicle- and drug-treated samples from each set of N=6 rats is measured. Statistical testing of hormone X concentrations between vehicle and drug-treated samples was performed using a paired t-test (grey shaded boxes; *, $p<0.05$ declared statistically significant).

C. Combined data from all the experiments (N=18 rats).

D. The concentrations of hormone X in vehicle- and drug-treated samples plotted against the baseline concentration. Red and grey arrows depict negative and positive differences between the vehicle and drug, respectively.

Declaration of competing interests

The authors are members of the Transparency and Reproducibility Committee (TRC) of the International Union for Basic and Clinical Pharmacology (IUPHAR). The primary aim of this Committee is to drive improvements in awareness and adherence to best practice in transparency and reproducibility in Pharmacology research practice. Debbie Hay is an editorial advisory board member of Trends in Pharmacological Sciences.

Hyperlinks to useful resources

SciScore

Research Resource IDentification (RRID)

Only Good Antibodies

Experimental Design Assistant (EDA)

ARRIVE

UK Reproducibility Network

UK research integrity office

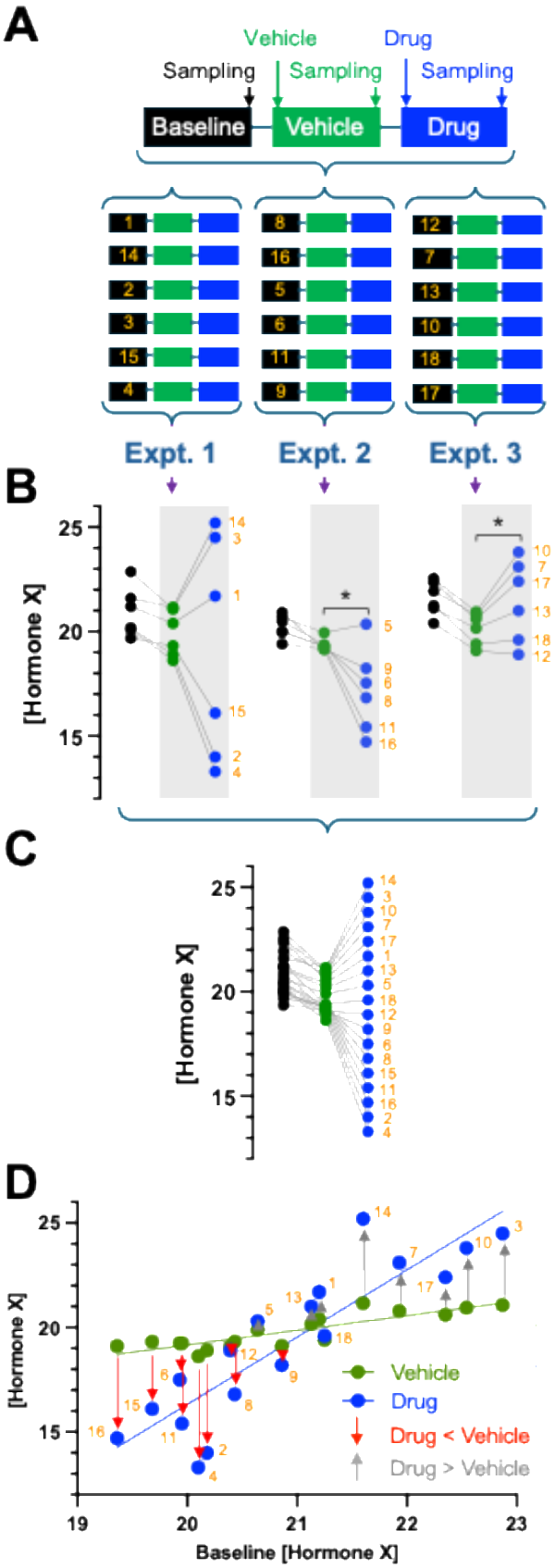
Centre for Open Science

References

1. Fry, D. *et al.* (2025) Teaching experimental design: outputs from the FELASE Working Group. *Lab. Animals* In Press.

2. Errington, T.M. *et al.* (2021) Investigating the replicability of preclinical cancer biology. *eLife* 10, e71601

3. Errington, T.M. *et al.* (2021) Challenges for assessning replicability in preclinical cancer biology. *eLife* 10, e67995
4. Curtis, M.J. *et al.* (2018) Experimental design and analysis and their reporting II: updated and simplified guidance for authors and peer reviewers. *Br. J. Pharmacol.* 175, 987-993
5. Curtis, M.J. *et al.* (2022) Planning experiments: updated guidance on experimental design and analysis and their reporting III. *Br. J. Pharmacol.* 179, 3907-3913
6. Curtis, M.J. *et al.* (2025) Guidance on the planning and reporting of experimental design and analysis. *Br. J. Pharmacol.* 182, 1413-1415
7. Bandrowski, A. *et al.* (2016) The Resource Idenfication Initiative: a cultural shift in publishing. *Neuroinformatics* 14, 169-182
8. Menke, J. *et al.* (2020) The rigor and transparency index quality metric for assessing biological and medical science methods. *iScience* 23, 101698
9. Percie du Sert, N. *et al.* (2016) The Experimental Design Assistant. *PLoS Biol.* 15, e2003779
10. Kwon, D. (2024) The quest to rid labs of reagents that ruin experiments. *Nature* 635, 26-28
11. Uhlen, M. *et al.* (2016) A proposal for validation of antibodies. *Nat. Methods* 13, 823-827
12. Hair, K. *et al.* (2019) A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARUS). *Res. Int. Peer Rev.* 4, 12
13. Bate, S.T. *et al.* (2025) A three-stage strategy for conducting an experimental investigation: a recommendation to improve the reproducibility of reported conclusions. *J. Psychopharm.* 39, 301-312
14. George, C.H. *et al.* (2017) Updating the guidelines for data transparency in the British Journal of Pharmacology - data sharing and the use of scatter plots instead of bar charts. *Br. J. Pharmacol.* 174, 2801-2804
15. van Dongen, N. and Sikorski, M. (2021) Objectivity for the research worker. *Eur. J. Phil. Sci.* 11, 93
16. Festing, M.F. *et al.* (2016) *The design of animal experiments: reducing the use of animals in research through better experimental design.* Sage Publications Ltd.


A
Vehicle
Drug
Sampling
Sampling
Sampling
Baseline
Vehicle
Drug
1
14
2
3
15
4
8
16
5
6
11
9
12
7
13
10
18
17
Expt. 1
Expt. 2
Expt. 3
B
[Hormone X]
25
20
15
*
*
C
[Hormone X]
25
20
15
D
[Hormone X]
25
20
15
19
20
21
22
23
Baseline [Hormone X]
Vehicle
Drug
Drug < Vehicle
Drug > Vehicle

Figure