

Errors and error management in biomedical research

Ulrich Dirnagl^{1,2,3,4,5} ORCID ID 0000-0003-0755-6119

René Bernard¹ ORCID ID 0000-0003-3265-2372

¹Department of Experimental Neurology and Center for Stroke Research Berlin (CSB), Charité Universitätsmedizin Berlin, Germany; ²German Center for Neurodegenerative Diseases (DZNE) and ³German Center for Cardiovascular Diseases (DZHK), Berlin sites; ⁴Excellence Cluster NeuroCure, Berlin, Germany; ⁵Berlin Institute of Health, Berlin, Germany

This chapter has been published in the book “How Could This Happen? – Managing Errors In Organizations”, edited by Jan U. Hagen, published by Palgrave Macmillan, Cham in 2018

<https://doi.org/10.1007/978-3-319-76403-0>

The validity of biomedical research results has come under scrutiny that revolves around factors regarding the quality of these results. However, due to the complexity of the experiments involved, errors quite naturally occur frequently. They include quality-compromised devices, errors due to limitations of measurement, protocol deviations, reporting errors, and human errors caused by carelessness or moments of distraction during complex tasks. Additionally, there are errors of unknown cause, which makes it vital to communicate them and keep a record of them for further investigation.

A way of managing these errors is the “Laboratory Critical Incident and Error Reporting System” (LabCIRS), a software tool to record all incidents anonymously and to analyze, discuss, and communicate them. It has been adapted from the Critical Incident and Error Reporting System (CIRS), used in the clinical world to improve patient safety in complex, fast-paced, and often understaffed settings.

LabCIRS is a nonpunitive format devoid of emotional connotations, exclusively focused on how to avoid errors in the future. Errors are reported, viewed, and initially classified. Thereafter, a decision is made about the urgency of actions. After further discussion, measures to be taken in response are determined and entered into the system. They are

presented to the research group in question. A newsletter, issued on a regular basis, summarizes both errors and actions that were taken. No personal information or computer IP is ever recorded, ensuring the confidentiality of the reporter.

LabCIRS could be considered an essential model for any community seeking a systematic error-management strategy to handle quality issues that have become a matter of concern.

Introduction

Errare humanum est - To err is human. Biomedical research, a human enterprise, is no exception in this regard. Ever more sophisticated methodologies probing how complex organisms function in health in disease invite errors on all levels - from designing experiments and studies to the collection of data and the reporting of results. The stakes are high, in terms of resources spent, and professional rewards to be gained for individuals. Up to now the public holds scientists, and in particular those aiming to improve human health, at very high esteem. Nevertheless, the exposure of spectacular cases of fraudulent and irreproducible research ^{1,2} and the realization that biomedical results appear to lack robustness as well as the scientific process rigor, have recently begun to undermine some of that public trust ³.

Even within the research community there is growing concern that a considerable fraction of research is actually waste ⁴ - a notion that is reflected in the fact that a majority of scientists agrees that we are experiencing a significant 'reproducibility crisis' ⁵. The search for causes, and potential remedies, has led to considerable introspection, science has turned its scrutiny upon itself ⁶.

Many factors have been singled out which may be invoked to explain the current concerns about the validity of biomedical research results. Top of the list are exceedingly low statistical power (i.e. sample sizes are too small), as well as low internal validity ⁷. Internal validity encompasses a number of quality factors, most of them related to bias. Bias can be eliminated, or at least controlled, by such measures as randomization, blinding, or prespecification of in- and exclusion criteria.

Clearly, the discussion is revolving around factors which negatively impact on the quality of research - and which may be remedied by structured measures to improve research quality ⁸. The potential contribution of errors to the disappointingly low level of reproducibility and predictiveness of biomedical research, and how scientists deal with these errors, has not yet been considered. This is highly surprising, as error management plays a central role in any structured approach to safeguard quality ('quality management'). It is also safe to assume that due to its multiple levels of complexity, errors must be quite frequent in biomedical research. In the following we explore the types of errors that might occur in biomedicine. We then propose and offer a simple tool to establish a mature error culture in biomedical research.

Systematic (device) errors

Such errors happen when a device is failing in its precision to measure in the desired range, or a complete device malfunction occurs. It is especially difficult to discover such failures if the device does not contain a self-check mechanisms for proper operational range or an internal standard. A prominent example are measuring pipettes operating in the microliter range. In microbiology or cell biology experiments, faulty volume measurements result in serious mistakes and directly lead to false outcome measures that often go undetected.

Therefore, every research laboratory, however small or big they are, need to make an inventory of their devices and lab equipment which provides information regarding warranty status, maintenance cycles, routine checks, and possible calibration procedures. Corresponding event dates need to be listed showing when these tasks were performed last and when they are due. Setting up an automated reminder system can prove very helpful to prevent errors due to lack of calibration. In industry contractors or service agents from device makers usually take over all these tasks for a fee. Academic biomedical research laboratories usually do not possess sufficient funds to outsource maintenance for all devices. Therefore, it is vital that they identify all critical devices in the experimental process and ensure that all of them function properly prior to the experiment. Only then valid results

can be obtained. When noticing that an uncalibrated or otherwise quality-compromised device is being used in an experiment, every researcher needs to note this in his or her laboratory notebook along with results and protocol used. Only then results can be further evaluated and by comparison a decision can be made whether to keep or discard the experiment.

Errors due to limitation of measurement

Many laboratory devices use changes in physical properties (light or ray emission, chemical reaction) of the analyte as proxies for the parameter under study, simply because these are easily detectable and quantifiable. However, there are certain measurements that rely on human judgment, such as color scale matching of pH paper. Another more common measure using human evaluation scales are behavioral scores for laboratory animals. Even though a common description exists, discrepancies in interpretation among lab personnel or misjudgments occur and are major contributors to the large variance of behavioral experiments.

One strategy is replacement of these measurements with objective assays, e.g. calibrated, electronic pH meters. Regarding scoring behavior, training and easily understandable, detailed standard operating procedures are a good starting point. Ideally video examples representing specific scores can not only help during training but also offer better recall during actual scoring. Every training should be followed up by blinded tests to confirm the validity of the practice. In addition, multi-lab comparisons for test procedures can help identifying ambiguity in protocol and therefore ensure reproducibility.

Protocol deviation

Protocol deviation is any non-compliance with an existing protocol, standard operating procedure (SOP), or work instruction. Another irregularity related to this category is known as "protocol drift", which is present when a protocol is executed without or only limited supervision or content checks. If this alteration does not result in immediate experimental

failure or is otherwise noticeable, the practice becomes the norm and is most often only detected by accident, if at all.

To combat this error, several steps can be taken. The SOP can be accompanied by a mandatory checklist list containing key elements of the protocol in a concise manner which need to be checked, or certain data that need to be inserted. Another element concerns mandatory regular checks of the protocol content by a supervisor or other responsible person. Every protocol should contain information regarding this validity check, i.e. an expiration date. Most effective against protocol deviations are internal reviews or method audits.

Reporting errors

Reporting errors can be the result of insufficient or faulty documentation during an experiment. Post-experimental analytical errors, such as statistical errors, also belong to this group. Without access to the original data or the original documentation it is hard to detect such errors, especially when the publications have already been peer-reviewed.

Various publication platforms have emerged, such as *F1000Research*, permitting public post-publication review, transparency of the entire review process, commenting tools and the possibility of versioning of a publication, e.g. in response to a comment. They also require a public deposition of the underlying original data. Data platforms like *FigShare*, *Dryad* or *Mendeley Data* host any research data, including data underlying published reports, which then contain cross references to the deposited data. Consequently, the entire scientific community can scrutinize the validity of raw and summary measures and reuse these data for further analysis, data synthesis, and aggregation.

Another often underutilized tool to prevent reporting errors or unnecessary ambiguity are reporting guidelines. International scientific organization have long recognized that despite peer-review many publications lack vital information for data interpretation or definition of responsibilities in the research and publication process. Two prominent examples for established guidelines are the ARRIVE guidelines, intended to improve the reporting of

research using animals (<http://www.nc3rs.org.uk/arrive-guidelines>), and the ICMJE guidelines to establish best practice and ethical standards in the conduct and reporting of research (<http://www.icmje.org/recommendations/>). Similar to protocols, these guidelines are often accompanied by checklists which make it easier for authors to verify their concordance with these guidelines. Journals are partially to blame for errors on method reporting because often there are strict word limits on specific sections of the manuscript (including the method section), forcing authors to be less specific or use references to other sources that are often not accessible for all readers. Fortunately, an increasing number of journals encourage authors to provide links to the underlying raw data for each figure, some even made this step mandatory.

A retraction of a publication is necessary when an error contained it cannot be clarified by a corrigendum, or when important conclusions of the article are affected by the error. Other reasons for retraction may be plagiarism or duplicate/concurrent publishing which will be not be considered further here. In the past, retractions happened ‘quietly’ and were therefore often not noted by the community. This changed in 2010 when an Internet blog service called *Retraction Watch* (www.retractionwatch.com) appeared. This blog aims to cover all retractions of research papers, and to report on the reasons or background of the retraction. Retraction Watch is widely read in the community and by journalists. Importantly, the editors of Retraction Watch also point out commendable retractions. As stated above, to err is human, and to stand by one’s errors, to expose and correct them are important elements of the self correcting quality of science. Hopefully, this will foster the development of an error culture in science.

Errors of yet unknown cause

Laboratory protocols often contain experimental controls or checkpoints in which verification takes place, or a comparison to a known standard. Despite adherence to the protocol and verification of all used reagents, the obtained result may not measure up to the standard. We here classify these errors as “of yet unknown cause”. It is important to communicate these

errors and to keep a record of them. If such an error happens more than once, an underlying systematic error must be suspected which requires further investigation until the source is identified and eliminated.

For instance, cell survival rate in the preparation of a cell culture suspension from neonatal rodent brains is a standard procedure that delivers an anticipated yield of living neuronal cells in culture when the protocol is followed. However, occasionally a large fraction of cells is dead. The causes are sometimes unknown, and an error must be suspected. If a second, similar incident occurs by another experimenter soon thereafter, a structured search for a potential error source is indicated. In our department, a cell toxic impurity of one of the cell culture medium ingredients accidentally introduced by the manufacturer had caused such an error. Only through swift and systematic investigation we were able to minimize the waste of resources, not only in our laboratories, but also for other customers of the manufacturer of the cell culture medium.

Human errors

In any work environment in which people are planning and executing tasks so called 'human errors' are bound to happen. Main causes include carelessness or moments of distraction during complex tasks. Most often human errors in the biomedical lab present themselves as mix-ups, for instance when a wrong reagent is used with a similar appearance, or samples in identical containers get mixed up or when a container received a wrong label (Figure 1).

Examples for negligence include the failure to close the door of a lab freezer which contains important samples, or the introduction of thermolabile equipment into an autoclave. Despite the fact that many lab records are obtained electronically, human errors can occur when electronic documentation of experiments is not saved, or is accidentally overwritten.

Error management

A number of specific measures can help to reduce the potential for error in biomedical research: Critical reagents which are prone to mix-ups should receive a distinctive color-

coding. Critical work steps should be witnessed by another person (known as four-eye-principle). Solutions which assist humans by automatizing repetitive workflow or involve alarms are particularly effective, such as freezers with an active alarm system or automatic centralized data backups concurrent with data generation. Most important, however, is that errors are recorded, evaluated and communicated. In the clinical world such error discussion sessions are known as “Morbidity and Mortality” conferences: Medical doctors regularly present cases that involve human errors in order to come up with preventive measures, and to help others making the same mistakes. Another structured approach to error communication and prevention are anonymous reporting tools known as Critical Incidence Reporting Systems (CIRS). Up to now such a stringent error reporting practice only takes place in highly regulated environments, such as healthcare, aviation, or power plants.

LabCIRS - a simple and free error management tool for biomedicine

In our department we have developed, implemented, and tested a free and simple error management tool for biomedicine. Laboratory Critical Incidence Reporting System (LabCIRS) is an adaptation of the clinical CIRS model. Biomedical lab personnel is encouraged to report any laboratory practices, results or situations that could negatively impact safety, animal welfare, longevity of material and devices in addition to observed protocol deviation or any other error of yet unknown causes ⁹. While this directive seems logic and reasonable, the reality of error handling in biomedical laboratories today is different. Commonly many errors go unnoticed or are not communicated at all because currently there is no mandatory error management system for preclinical research laboratories. Errors have a negative connotation and there is the fear for personal liability and humiliation which may hinder biomedical lab personnel to report such incidents. We hope with the introduction and free provision of LabCIRS that this will change, as CIRS has changed the error communication culture and has become a standard in the medical field.

A CIRS system was first described by Flanagan in 1954 ¹⁰, by 1978 introduced in anesthesiology ¹¹, and is today a worldwide established mechanism that improves patient

safety in complex, fast-paced and often understaffed clinical settings. What made it so attractive to adapt this system for the preclinical world was the fact that all incidents are recorded anonymously, analyzed, discussed and communicated (Figure 1). These reports possess a non-punitive format devoid of any emotional connotation of the incident, entirely focussing on how to avoid the described error in the future. As most communication and reporting nowadays is electronic, we developed a freely available software package called LabCIRS. LabCIRS permits safe and anonymous lab-related error reporting via an intranet network environment. No personal information or even computer IP is ever recorded, assuring confidentiality of the reporter. LabCIRS has common login credentials for every department member which not only allows error reporting. Everyone can view the entire history of all errors. Within the first two years since the introduction of LabCIRS in the Department of Experimental Neurology of the Charité 49 incidents have been entered in the system. Of these 4 were device errors, 18 protocol deviations, 18 human errors, 6 error of unknown cause. In addition, 3 injuries were reported. This is a testimonial that LabCIRS not only works as intended but is accepted by the lab members of the department.

After entries are made into LabCIRS, errors are viewed and initially classified by our quality management officer and a decision is made about the urgency of actions. Most errors are then discussed among the quality representatives of all research groups. They decide if and which measures need to be taken in response. These measures are then also entered into LabCIRS. All newly reported LabCIRS errors are presented during a weekly lab meeting. A monthly newsletter to all department members summarizes all LabCIRS errors of the past weeks and the actions that were taken. Most reporters of incidents now reveal their identity, a further indication that the system is accepted and that reporting of errors has no ill consequences for employees. We sincerely hope that LabCIRS is adopted by many laboratories and departments of the academic research community and will contribute to the development of a much needed open error culture in the preclinical research laboratory.

Summary and Outlook

As any other scientific field, biomedical research is susceptible to various types of errors. Compared to the aviation industry, clinical medicine, or radiation safety, most errors in the biomedicine do not have potentially life-threatening consequences. Therefore, governmental oversight in biomedical research is minimal and most commonly restricted to work with genetically-modified organisms, animal welfare, occupational health related safety, and environmental protection. Nevertheless, errors in biomedicine may cause major waste of resources, and potentially harm patients if clinical studies are based on erroneous or faulty results. Despite the existence of protocols and guidelines in biomedical laboratories, most errors simply 'occur', and are not systematically followed up for future prevention. Policies, established procedures, or tools regarding error management are virtually non-existent in this environment. We posit that the biomedical research community, which is currently undergoing a 'reproducibility crisis' to which quality issues may contribute substantially, should develop and implement minimal standards for quality management which includes systematic error management strategies¹²⁻¹⁴.

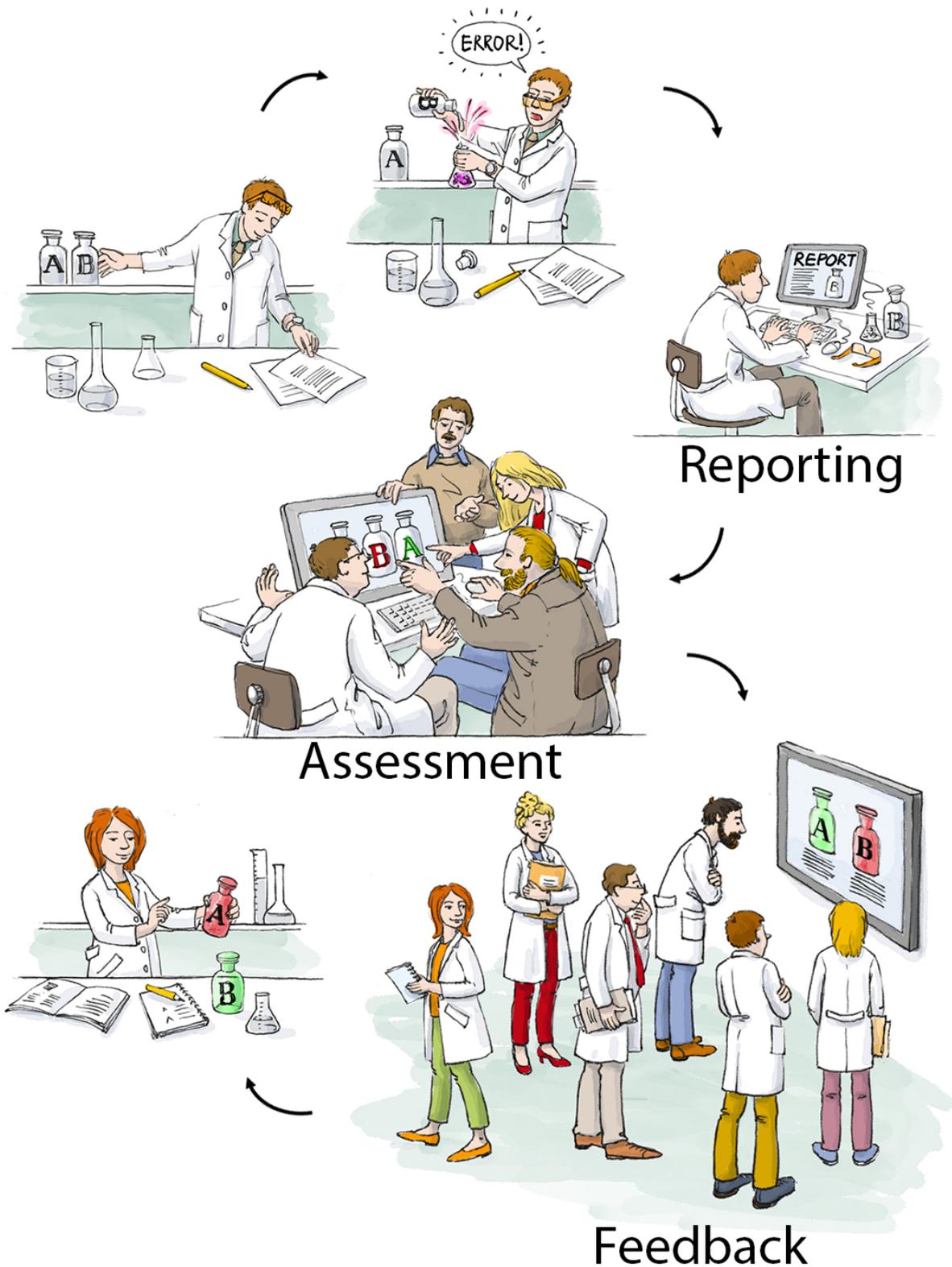


Figure 1: Illustration of an ideal process of error handling in the biomedical laboratory (taken from Dirnagl et al. "A Laboratory Critical Incident and Error Reporting System for Experimental Biomedicine" PloS Biology 2016; in press) A researcher mistook two faintly

labeled reagents A and B, which ruined his experiment. “Reporting”: Entry of the incident into LabCIRS. “Assessment”: A group of experts (scientists and technicians) reviews the error, and takes preventive action by color labelling of the reagents. “Feedback”: The errors as well as the measure to prevent it in the future are communicated to the entire laboratory.

Acknowledgements: The authors would like to thank Sebastian Major, Ingo Przesdzing, and Claudia Kurreck for developing, implementing, and testing LabCIRS.

Funding statement: UD would like to acknowledge the financial support of the German Federal Ministry of Education and Research (BMBF 01 EO 08 01) and the Herman and Lilly Schilling Foundations. The funders had no role in the decision to publish, or preparation of the manuscript.

Competing interests: None of the authors have any competing interests to declare.

References:

1. Cyranoski, D. Verdict: Hwang’s human stem cells were all fakes. *Nature* **439**, 122–123 (2006).
2. Obokata, H. *et al.* Retraction: Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* **511**, 112–112 (2014).
3. Carey, B. Science, Now Under Scrutiny Itself. *New York Times* p. D1. New York City. Retrieved from http://www.nytimes.com/2015/06/16/science/retractions-coming-out-from-under-science-rug.html?_r=1 (2015).
4. Macleod, M. R. *et al.* Biomedical research: increasing value, reducing waste. *Lancet* **383**, 101–104 (2014).
5. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
6. Ioannidis, J. P. A., Fanelli, D., Dunne, D. D. & Goodman, S. N. Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biol.* **13**,

- e1002264 (2015).
7. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
 8. Baker, M. How quality control could save your science. *Nature* **529**, 456–8 (2016).
 9. Dirnagl, U., Przesdzing, I., Kurreck, C. & Major, S. A laboratory critical incident and error reporting system for experimental biomedicine. *PLoS Biol.* In Press (2016).
 10. Flanagan, J. C. The critical incident technique. *Psychol. Bull.* **51**, 327–58 (1954).
 11. Cooper, J. B., Newbower, R. S., Long, C. D. & McPeck, B. Preventable anesthesia mishaps: a study of human factors. *Anesthesiology* **49**, 399–406 (1978).
 12. Riedl, D. H. & Dunn, M. K. Quality assurance mechanisms for the unregulated research environment. *Trends Biotechnol.* **31**, 552–4 (2013).
 13. Davies, R. Good Research Practice: It is Time to do What Others Think we do. *Quasar* 21–23 (2013).
 14. Begley, C. G., Buchan, A. M. & Dirnagl, U. Robust research: Institutions must do their part for reproducibility. *Nature* **525**, 25–27 (2015).