Ten simple rules in good research practice for early career researchers

Simon Schwab[1,2], Perrine Janiaud[3], Michael Dayan[4], Valentin Amrhein[5], Radoslaw Panczak[6], Patricia M. Palagi[7], Lars G. Hemkens[3,8,9], Meike Ramon[10], Nicolas Rothen[11], Stephen Senn[12], Eva Furrer[1,2], Leonhard Held[1,2]

[1] Center for Reproducible Science, University of Zurich, Zurich, Switzerland.
[2] Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.
[3] Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland.
[4] Human Neuroscience Platform, Fondation Campus Biotech Geneva, Geneva, Switzerland.
[5] Department of Environmental Sciences, Zoology, University of Basel, Basel, Switzerland.
[6] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.
[7] SIB Training Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland.
[8] Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA.
[9] Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Berlin, Germany.
[10] Applied Face Cognition Lab, University of Fribourg, Fribourg, Switzerland.
[11] Faculty of Psychology, UniDistance Suisse, Brig, Switzerland.
[12] Statistical Consultant, Edinburgh, UK.

Introduction

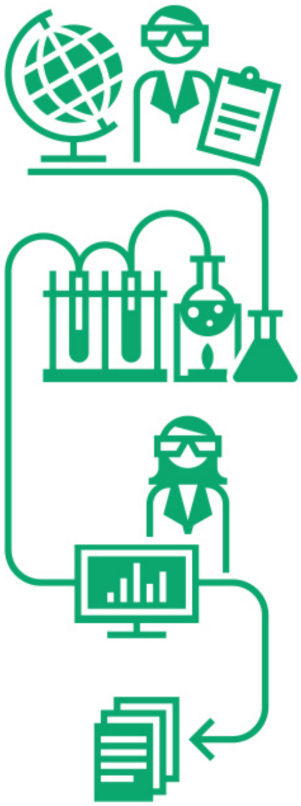
The lack of research reproducibility has caused growing concern across various scientific fields [1–4]. Today, there is widespread agreement, within and outside academia, that scientific research is suffering from a reproducibility crisis [5,6]. Even in an optimal and controlled setting, researchers might inadvertently reach different conclusions when multiple steps of the analysis are decided upon [7,8]. As we move beyond the state of crisis, some of the major causes of irreproducible research have been identified, which allow us to advocate good research practice (GRP). Indeed, powerful solutions are available, for example preregistration of study protocols and statistical analysis plans, sharing of data and analysis code, and adherence to reporting guidelines. Although these and other best practices may facilitate reproducible research and increase trust in science, it remains the responsibility of researchers themselves to actively integrate them into their everyday research practices.

Enhancing research quality and combatting causes of irreproducible research is directly linked to how the next generation of young researchers is trained to conduct scientific work. Contrary to ubiquitous specialized training, cross-disciplinary courses on best practices that enhance the quality of research are often lacking. This is unfortunate, as the intersections between disciplines offer a space for peer-evaluation, mutual learning and sharing best practices. In medical research, interdisciplinary work is inevitable. For example, planning and conducting clinical trials may require experts with diverse backgrounds, including clinical medicine, pharmacy, biostatistics, evidence synthesis, nursing and implementation science. They contribute to study protocols, statistical analysis plans, the conduct of the clinical trial and the generation and interpretation of results. Bringing young researchers with diverse backgrounds together to exchange knowledge and learn about problems and solutions in everyday research practice adds value and improves the quality of research.

This paper aims to provide early career researchers a useful introduction to GRP. The selection of rules is based on feedback from students, our experiences of teaching GRP courses at the University of Zurich as well as on the views of a diverse group of experts from within the Swiss Reproducibility Network (www.swissrn.org). This cross-disciplinary exercise resulted in a list of ten simple rules proposed to foster GRPs for early career scientists. This is neither an exhaustive list nor does it aim to address and systematically summarize the wide spectrum of issues including research ethics and legal aspects (e.g., related to misconduct, conflicts of interests and scientific integrity). We focused on practical advice for early career researchers at the different stages of every-day research: from planning to conduct and reporting of research. For a more comprehensive overview on GRPs we point to the United Kingdom's Medical Research Council's guidelines [9] and the Swedish Research Council's report [10]. While the discussion of the rules may predominantly focus on clinical research, much applies, in principle, to basic biomedical research and research in other domains as well.

The ten proposed rules may serve multiple purposes: an introduction to relevant concepts to improve research; a primer for early career researchers on what to expect in courses related to GRPs; and finally, a starting point for lecturers who are planning such courses to further enhance research quality at their own institutions. The ten rules are grouped according to planning (4 rules), conducting (4

rules), and reporting of research (2 rules); see Figure 1. These principles can (and should) be implemented as a habit in everyday research practice, just like tooth brushing.



Planning of research
1. Specify your research question
2. Write and register a study protocol
3. Reduce potential sources of bias
4. Have a data management plan

Conducting of research
5. Ensure research reproducibility
6. Make your research open
7. Justify your sample size
8. Get statistical expertise

Reporting of research
9. Report all findings
10. Follow reporting guidelines

Figure 1: The ten simple rules in good research practice (GRP) grouped in the planning of research, conducting of research and reporting of research.

Planning of Research

Rule 1: Specify your research question

Coming up with a research question is not always simple and may take time. A well-developed research question is relevant, and for a study to be successful it should be narrow and clear. In evidence-based research, prior studies are assessed in a systematic and transparent way to identify a research gap for a new study that answers a question that matters [11]. It is good advice to look for recent papers that provide a comprehensive overview of the current state of research in the field. Particularly, review papers and perspective papers are ideal to start with. For example, imagine a paper titled "SARS-CoV-2 and COVID-19: The most important research questions".[1] While this would be a great kick-off, systematic assessments of research gaps are even better to consider than opinion-based research priorities.

Once a first research question is drafted, it should be further developed and refined. It is natural to be in doubt at the beginning of the process—scientists believe in the theory to carry forward but are sceptical as well [12]. In clinical research and evidence-based medicine there is an approach called

[1] A paper with this title does actually exist.

PICOT (population, intervention, comparator, outcome, and time frame) with a set of criteria that can help framing a research question [13]. From a well-developed research question, subsequent steps will follow, which may include the exact definition of the population, the outcome, the data to be collected, and the sample size that is required. Next, find out if others find the idea interesting as well and whether it might promise a valuable contribution to the field. Some would argue that one should not ask only colleagues or researchers about which questions matter, but primarily the public and patients as well.

As you read these lines you may wonder: How does this apply to research that is less theory-driven, for example to lab experiments that coincidentally find something unexpected? Indeed, Alexander Fleming discovered penicillin in a happy accident. Likewise, Christopher Columbus had a totally different goal in mind as he discovered America. In exploratory research, no or only a vague hypothesis is required. Both exploration and confirmation are needed in science [14], and they have their own strengths and limitations. However, it is important to decide whether the research is exploratory or confirmatory in the first place and not confuse the two [15].

Rule 2: Write and register a study protocol

In clinical research, registration of clinical trials has become a standard since the late 1990 and is now a legal requirement in many countries. Such studies require a study protocol to be registered, for example with ClinicalTrials.gov, the European Clinical Trials Register, or the World Health Organization's International Clinical Trials Registry Platform. Similar effort has been implemented for registration of systematic reviews (PROSPERO). Study registration has also been proposed for observational studies [16] and more recently in preclinical animal research [17] and is now being advocated across disciplines [18,19]. Often the term "preregistration" is used and the process is facilitated by the Open Science Framework (OSF) that supports registration of study protocols across all research fields.

Study protocols typically document at minimum the research question and hypothesis, a description of the population, the targeted sample size, the inclusion/exclusion criteria, the study design, the data collection, the data processing and transformation, and the planned statistical analyses. Preregistering study protocols not only helps reducing publication bias but also prevents hindsight biases that could affect the outcome of a study. Registration of hypotheses, procedures and planned analyses can therefore safeguard honest research [20] and minimize research waste [21,22]. But the registration of studies should be more often under scrutiny by comparing the reported research to what was actually written in the protocol, and not doing so may be a serious issue.

Note that registration does not mean that researchers have no flexibility to adapt the plan as needed. Indeed, new or more appropriate procedures may become available or known only after registration of a study. Therefore, a more detailed statistical analytical plan can be amended to the protocol before the data are observed or unblinded [23,24]. Likewise, registration does not exclude the possibility to conduct exploratory data analyses—they do, however, have to be clearly reported as such.

To go even one step further, registered reports are a novel article type of publishing research to incentivize rigorous study design, methodology and statistical analysis—irrespective of the ultimate study outcome [25]. With registered reports, researchers lay out their hypotheses, research methods,

and analysis plans, and peer reviewers decide whether to accept them before anyone knows the study's results effectively leading to higher quality research [26]. Journals from various disciplines increasingly support registered reports [27] and assign reviewers a more active role where they can identify problems and have an impact on the design and analysis of the study prior to data collection and analysis.

Rule 3: Reduce potential sources of bias

Bias is a distorted view in favor of or against a particular idea or thing. In statistics, bias is a systematic deviation of a statistical estimate from the (true) quantity it estimates. Bias can invalidate our conclusions, and the more bias there is, the less valid they are. For example, in clinical studies, bias may mislead us into reaching a causal conclusion that the difference in the outcomes was due to the intervention or the exposure - although it is actually not true. This is a big concern, and therefore potential sources of bias are usually assessed in clinical trials [28] and observational studies [29,30].

There are different types of biases that can occur in a study [31]. A well-known and frequent example is selection bias which is present when participants differ systematically from the population of interest, or allocation bias in RCTs when the compared groups differ systematically in some important aspect beside the intervention or exposure. Information bias, another frequent example of bias, arises from systematic differences in the collection, recall, recording or handling of data. An example of information bias was observed when in 1998 an alleged association between the measles-mumps-rubella (MMR) vaccine and autism was reported. The reason was that parents of autistic children tended to recall the onset of autism after MMR more often than other parents [32]. But many other biases exist; for a comprehensive overview see the Oxford University's Catalogue of Bias [33].

The prevention or at least reduction of bias merits a "ten simple rules" article on its own. There is a large body of literature on bias, and the previously mentioned catalogue of bias [33] also provides preventive steps. To minimize selection bias, the choice of a control group or control condition is critical and affects the inference that can be drawn. For example, in a case-control study, the control group should be representative of the population that produced the cases. Allocation bias in randomized controlled trials can be addressed by allocation concealment. Prevention of information bias can be achieved by using objective rather than subjective measurements or blinded outcome assessments. Standardized data collection is also critical by using the same method for data collection in both groups and not, for example, collecting data by an interview in one group and by phone in the other. Even procedures that may seem "simple" such as measuring blood pressure can be affected by serious bias. Standardized operating procedures (SOP) and electronic lab notebooks [34] additionally help to follow well-designed protocols for data collection and handling. Despite the failure to mitigate bias in studies, complete descriptions of data and methods can at least allow the assessment of potential bias.

Rule 4: Have a data management plan

In 2020, two COVID-19 papers were retracted by leading medical journals after concerns about the data were raised. Only when data are available, the result can be verified, which increases trust in science. Today, data is more often recognized as a key outcome of research along the research paper.

Therefore, it is important to develop a strategy for the lifecycle of research data, including suitable infrastructure for long-term storage.

The data lifecycle is described in a data management plan: a document that describes what data will be collected during research, how the data will be organized, stored, handled and protected during and after the end of the research project. A number of research funders require a data management plan in research grant submissions, and publishers like PLOS encourage authors to do so as well.

The FAIR data principles (Findable, Accessible, Interoperable and Reusable) promote maximal use of research data and enable computational systems to access and reuse data with minimal human intervention [35]. FAIR principles require the data to be retained, preserved, and shared preferably with an immutable unique identifier and a clear usage licence. Appropriate metadata will help other researchers (or machines) to discover, process and understand the data.

Data should as well be openly accessible in a public repository (such as Dryad, figshare, OSF, Zenodo or your institutional repository). However, in some instances, authors may not be able to make their data publicly available for legal or ethical reasons. In such cases, a data user agreement can indicate the conditions required to access the data. Journals highlight what are acceptable and what are unacceptable data access restrictions, and often require a data availability statement.

Interoperability and reusability can amplify the scientific impact of the data and code produced during a research project. It can be helpful for early career researchers to reuse the data of another person from the same lab, or even for the same person after a long break on that project. Organizing the study artefacts in a structured way greatly facilitates the reuse of data and code within and outside the lab, enhancing collaborations and maximizing the research investment.

Support and courses for data management plans are usually available at universities, and another ten simple rules paper for creating a good data management plan is dedicated to this topic [36].

Conducting of research

Rule 5: Ensure research reproducibility

Questionable research practices can lead to exaggerated findings and false conclusions and thus lead to irreproducible research [37]. Often questionable research practices are used with no bad intentions (as opposed to fraud and data fabrication, which may be less common), but because researchers lack the training to perform research of the highest quality. This becomes evident when methods sections explicitly describe such procedures, for example, to increase the number of samples until statistical significance was reached that supports the hypothesis. Thus, teaching GRPs may be (at least in theory) highly effective to enhance the quality of research.

Several questionable research practices (QRPs) have been named [37]: 1. low statistical power, 2. pseudoreplication, 3. repeated inspection of data, 4. *p*-hacking, 5. selective reporting, 6. hypothesising after the results are known (HARKing) and 7. publication bias [38].

The first two QRPs, low statistical power and pseudoreplication, can be prevented by proper planning and designing of studies, including sample size calculation and appropriate statistical methodology to avoid treating data as independent when in fact they are not. Statistical power is not equal to reproducibility, but statistical power is a precondition of reproducibility as the lack thereof can result in false negative as well as false positive findings (see rule 7).

In fact, when early career researchers write a study protocol and statistical analysis plan, many of these QRPs are addressed. Preregistration, as described in rule 2, is considered best practice for this purpose. However, many of these issues can additionally be rooted in institutional incentives and rewards. Thus, a wider "culture change" is needed beyond what young researchers can achieve. Both funding and promotion are often tied to the quantity rather than the quality of the research output. At universities, still only few or no rewards are given for writing and registering protocols, sharing data, publishing negative findings, and conducting replication studies.

Rule 6: Make your research open

For many people, transparency is a natural feature of science, and adding "open" to the word "science" would seem simply redundant. However, in reality, science often lacks transparency. Several universities and research funders already implemented open science roadmaps to advocate free and public science as well as open access to scientific knowledge, with the aim of further developing the credibility of research. Open research allows more eyes to see it and critique it, similar to a principle to spot bugs in software (Linus's law).

Many early career researchers already know about open science and even practice it. Open science has a wide scope, but one particular way to understand open science is to use it as a tool to document research output, and advance research and researchers' careers, especially for early-career researchers. The increased visibility, retrievability, and citations of datasets can all help career building [39]. Therefore, institutions should provide necessary training, and hiring committees and journals should align their core values with open science, to attract excellent researchers who aim for transparent and credible research [40].

Open science makes the process of producing evidence and claims transparent and accessible to others [41]. Only with open science, readers can fully and appropriately assess the performed research. As science often progresses incrementally, making a study protocol and making data and methods readily available is crucial to facilitate knowledge building. Researchers often need to re-create the work of others, especially if they have no access to it. For instance, computational notebooks using free software facilitate open access and reproducibility of the main results (e.g., R Markdown, Jupyter). Making both data and code open thus minimizes waste of funding resources and accelerates science, and consequently, funding agencies strongly support open science practices.

Rule 7: Justify your sample size

Lab researcher: *"We always used 3 patients. That is what we have published before."*
Biostatistician: *"Three patients per group or in total?"* (Source: https://youtu.be/PbODigCZqL8)

Box 1: A hypothetical conversation between lab researcher and biostatistician.

Too small studies can undermine research [42], but what is a small study? For one research field, a total of 30 patients may be average, but for another, 1,000 patients may be common. How large the sample size needs to be depends on the magnitude of effect you are interested in[2], and how large the probability to obtain a "statistically significant" effect should be, given there really is an effect (among other factors). This concept is based on the philosophy of statistical hypothesis testing and is known as statistical power: the probability to reject the null hypothesis when it is in fact false. The statistical power required by funding agencies is often 80% or larger. A worthwhile alternative for planning the sample size that puts less emphasis on null hypothesis testing is based on the desired precision of the study; for example, one can calculate the sample size that is necessary to obtain a desired width of a confidence interval for the targeted effect [43,44].

Early-career researchers in our GRP courses often identify sample size as an issue in their research. For example, they say that they work with a low number of samples due to slow growth of cells, or they have a limited number of patient tumor samples (or primary cell lines) but at the same time a high variation within the tumor samples. The problem here is that if your sample size is too low, your study has low statistical power and has a high risk of committing Type II errors (i.e., a false negative result). In other words, you are unlikely to find a result even if there is a true effect.

But there are also other issues with small studies. When an effect from a small study was selected for drawing conclusions because it is statistically significant, low power reduces the probability that the estimated effect size comes close to the true effect size [42]. The reason is that with low power, studies that due to sampling variation find larger (overestimated) effects are much more likely to be significant than those that happen to find smaller (more realistic) effects. Thus, in such situations effect sizes are often overestimated. The general phenomenon that small studies often report more extreme results than large studies is called the small-study effect [45]. In any case, an underpowered study is a problematic study, no matter the outcome.

Low-powered studies can be avoided by performing a sample size calculation to find out the required sample size of the study (unless your research is purely exploratory). As a solution and to support larger studies, team science and collaboration is crucial when one's own resources are limited.

Rule 8: Get statistical expertise

All early-career researchers have specialist training in their field, but basic training in statistics largely varies across fields. It is not surprising that in the small surveys we conduct in our GRP courses, early career researchers seem to find good statistical methodology the most important training to enhance their own research (Figure 2). Nurses, pilots and many other professionals have mandatory continuing development and training. Likewise, researchers could profit from better training in statistics and GRP [46].

[2] Not all research looks at effects and casualty. In diagnostic testing, one would be interested in an anticipated minimal sensitivity or specificity.

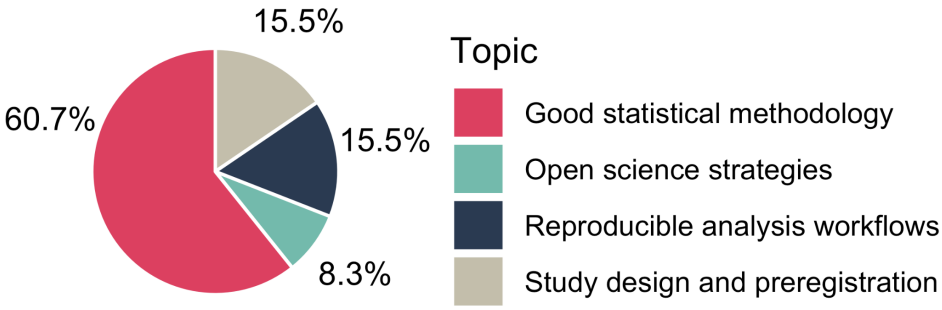


Figure 2: Early-career researchers (N=84) at University of Zurich that participated in the Good Research Practice (GRP) course responded to the question "*What training would you find most important to enhance your personal research?*".

It would help if more researchers were familiar with correct interpretations and possible misinterpretations of statistical tests, *p*-values, confidence intervals, and statistical power [47]. Specifically, the traditional dichotomisation into significant yes/no is seldom appropriate and we instead recommend interpreting p-values in a quantitative way in terms of the evidence against the null hypothesis [48]. A p-value around 0.05 provides only weak evidence, as is best illustrated by the associated replication power: the probability that a hypothetical replication study of the same design will lead to a statistically significant result is only 50% [49], and is even lower in the presence of publication bias and regression to the mean (the phenomenon that replication effect estimates are often smaller than the estimates in the original study) [50]. Claims of discovery of new effects should therefore be based on smaller p-values providing stronger evidence against the null hypothesis [51].

Furthermore, a study result and its associated *p*-value and confidence interval are just the tip of the iceberg [52]. The statistical analysis has many stages, for example data processing, cleaning, transformation, addressing missing data, and statistical inference. Errors and pitfalls can creep in at any stage, and even a tiny error can have a big impact on the result [53].

Training in basic statistics, statistical programming and reproducible analyses, and better involvement of data professionals in academia is necessary. University departments sometimes have statisticians that can support researchers. Difficulties and statistical pitfalls can start already with study design and sample size calculation [54]. Therefore, not surprisingly, statisticians often urge researchers to get them involved early in the process and on an equal footing, and not just at the end of a project to perform the final data analysis [55].

Reporting of research

Rule 9: Report all findings

Researchers, reviewers or publishers sometimes find study results not interesting or worth publishing. As a consequence, studies are not published, also known as the file drawer effect [56], or outcomes

and analyses are only selectively reported in research reports [57]. This phenomenon is more likely to occur when study results were found to be statistically non-significant.

Often a statistically non-significant result is interpreted as a "null" finding, but this is problematic. It is important to understand that a non-significant finding doesn't necessarily mean a null effect (absence of evidence is not evidence of absence, see [58]). Perhaps the study was just underpowered and the cumulative evidence from multiple single studies can indeed provide sufficient evidence for the research hypothesis. Another argument is that a confidence interval that contains the null value often also contains non-null values of high practical importance. Only if all the values inside the confidence interval are deemed unimportant (from a practical perspective), then it may be fair to describe a result as a null finding [59]. For example, clinical equivalence studies provide a formal framework to establish that the efficacy of two treatments is equivalent and require pre-specification of an interval of equivalence [60].

We should never report 'no difference' or 'no association' just because a p-value is larger than 0.05 or, equivalently, because a confidence interval includes the "null" value representing no difference between the groups [59]. However, studies that report statistically non-significant results often use misleading reporting strategies ("spin") to claim that the experimental treatment is generally beneficial, despite a statistically non-significant difference for the primary outcome [61,62].

The extent of publication bias in the literature is illustrated by the overwhelming frequency of statistically significant findings [63]. For example, a study extracting p-values from MEDLINE abstracts and full-texts available in PubMed Central showed that 96% of the text-mined records reported at least one statistically significant p-value [64], which seems implausible in the real world. Another study plotted the distribution of more than one million z-values from Medline – the huge gap from −2 and 2 is striking, showing particularly clearly that statistically non-significant results are often unpublished [65]. In addition, an overview of studies investigating the likelihood of being published showed that positive studies (i.e., statistically significant, perceived as striking or showing a beneficial effect) were 4 times more likely to get published than negative studies [66].

Findings that are not being published have a tremendous impact on the research ecosystem, distorting our knowledge of the scientific landscape by perpetuating misconceptions, and jeopardizing researchers and the public trust in scientific findings. In clinical research, publication bias can mislead care decisions and harm patients, e.g., when treatments appear useful despite minimal or even absent benefits reported in studies that were not published and thus are unknown to physicians [67]. Moreover, publication bias also directly affects the formulation and proliferation of scientific theories, which are taught to students and young researchers, thereby perpetuating biased research from the core. It has been shown in modeling studies that unless a sufficient proportion of negative studies are published, a false claim can become an accepted fact [68] and the false positive rates influence trustworthiness in a given field [69].

In sum, "null" findings are undervalued. They need to be more consistently reported at the study level or systematically investigated at the systematic review level. Researchers have their share of responsibilities, but there is clearly a lack of incentives from promotion and tenure committees, journals and funders.

Rule 10: Follow reporting guidelines

Study reports need to faithfully describe the aim of the study, what was done including potential deviations from the original protocol, and what was found. Yet evidence of discrepancies between protocols and research reports and insufficient quality of reporting are numerous [57,70–74]. Reporting deficiencies clearly jeopardize our ability to clearly communicate findings, replicate studies, make informed decisions and build on existing evidence, wasting time and resources invested in the research [75].

Reporting guidelines aim to provide the minimum information needed on key design features and analysis decisions, ensuring that findings can be adequately used and studies replicated. In 2008, the EQUATOR[3] (Enhancing the QUAlity and Transparency of health Research) network was initiated to provide reporting guidelines for a variety of study designs along with guidelines for education and training on how to enhance quality and transparency of health research. Currently, there are 468 reporting guidelines listed in the network, mostly focusing on health research; see some examples in Table 1. Furthermore, following the ICMJE recommendations, medical journals are increasingly endorsing reporting guidelines [76], in some cases making it mandatory to submit the appropriate reporting checklist along with the manuscript.

The use of reporting guidelines and journal endorsement have led to a positive impact on the quality and transparency of research reporting, but improvement is still needed to maximize the value of research [77,78].

Table 1: Examples of reporting guidelines for different study types, see https://www.equator-network.org/ for the complete collection.

Name	Study type
CONSORT	Randomised trials
STROBE	Observational studies
PRISMA	Systematic reviews
SPIRIT	Study protocols
STARD/TRIPOID	Diagnostic/prognostic studies

Conclusions

Good research practice involves many topics. Here, we focus on practical guidelines directed to early career researchers in planning, conducting and reporting of research. Others have aligned GRP with similar topics [79,80]. Even though we provide ten simple rules, the word "simple" should not be taken lightly. Putting the rules into practice usually requires effort and time, especially at the beginning of a research project. However, time can also be redeemed, for example, when certain

[3] https://www.equator-network.org/

choices can be justified to reviewers by providing a study protocol, or when data can be quickly reanalyzed by using computational notebooks and dynamic reports.

Researchers have field-specific research skills, but sometimes lack the awareness of best practices that are implemented in other fields (e.g., clinical research). Cross-disciplinary exchange should enhance self-correcting processes and best practices. Universities should provide GRP courses to train the next generation of scientists. They are an important building block to combat irreproducible research; however, GRP courses are not sufficient to drive a culture change. Many of the proposed solutions are still being discussed and the search for better solutions will and should be an ongoing task in science.

Acknowledgments

This article was written along the good research practice (GRP) courses[4] provided for early career researchers at the University of Zurich. All materials from the course are open and available at osf.io/t9rqm. We appreciated the discussion, development and refinement of this article within the working group "training" of the SwissRN (www.swissrn.org). We are grateful to Philip Bourne who gave some valuable comments on an earlier version of the manuscript.

Funding

Simon Schwab received funding from SfwF (Stiftung für wissenschaftliche Forschung an der Universität Zürich; grant no. STWF-19-007).

Competing interests

The authors have declared that no competing interests exist.

References

1. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nature reviews. Drug discovery. 2011. p. 712. doi:10.1038/nrd3439-c1

2. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science. 2015;349: aac4716. doi:10.1126/science.aac4716

3. Camerer CF, Dreber A, Holzmeister F, Ho T-H, Huber J, Johannesson M, et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour. 2018;2: 637–644. doi:10.1038/s41562-018-0399-z

4. Camerer CF, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, et al. Evaluating replicability of laboratory experiments in economics. Science. 2016;351: 1433–1436.

[4] https://www.crs.uzh.ch/en/training/GoodResearchPractice.html

doi:10.1126/science.aaf0918

5. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533: 452–454. doi:10.1038/533452a

6. Bespalov A, Barnett AG, Begley CG. Industry is more alarmed about reproducibility than academia. Nature. 2018;563: 626. doi:10.1038/d41586-018-07549-w

7. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science. 2018;1: 337–356. doi:10.1177/2515245917747646

8. Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature. 2020. doi:10.1038/s41586-020-2314-9

9. Medical Research Council, MRC. Good research practice: principles and guidelines. 2012. Available: https://mrc.ukri.org/research/policies-and-guidance-for-researchers/good-research-practice/

10. Swedish Research Council. Good Research Practice – What Is It? 2006. Available: https://lincs.gu.se/digitalAssets/1268/1268493_Good_Research_Practice.pdf

11. Robinson KA, Brunnhuber K, Ciliska D, Juhl CB, Christensen R, Lund H, et al. Evidence-Based Research Series-Paper 1: What Evidence-Based Research is and why is it important? J Clin Epidemiol. 2021;129: 151–157. doi:10.1016/j.jclinepi.2020.07.020

12. Erren TC, Cullen P, Erren M, Bourne PE. Ten simple rules for doing your best research, according to Hamming. PLoS Comput Biol. 2007;3: 1839–1840. doi:10.1371/journal.pcbi.0030213

13. Fandino W. Formulating a good research question: Pearls and pitfalls. Indian J Anaesth. 2019;63: 611–616. doi:10.4103/ija.IJA_198_19

14. Tukey JW. We need both exploratory and confirmatory. Am Stat. 1980;34: 23–25. doi:10.1080/00031305.1980.10482706

15. Schwab S, Held L. Different worlds Confirmatory versus exploratory research. Significance. 2020;17: 8–9. doi:10.1111/1740-9713.01369

16. Loder E, Groves T, MacAuley D. Registration of observational studies. BMJ . British Medical Journal Publishing Group; 2010. p. c950. doi:10.1136/bmj.c950

17. van der Naald M, Wenker S, Doevendans PA, Wever KE, Chamuleau SAJ. Publication rate in preclinical research: a plea for preregistration. BMJ Open Science. 2020;4: e100051. doi:10.1136/bmjos-2019-100051

18. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A. 2018;115: 2600–2606. doi:10.1073/pnas.1708274114

19. Nosek BA, Beck ED, Campbell L, Flake JK, Hardwicke TE, Mellor DT, et al. Preregistration Is Hard, And Worthwhile. Trends Cogn Sci. 2019;23: 815–818. doi:10.1016/j.tics.2019.07.009

20. Bradley SH, DeVito NJ, Lloyd KE, Richards GC, Rombey T, Wayant C, et al. Reducing bias and improving transparency in medical research: a critical overview of the problems, progress and suggested next steps. J R Soc Med. 2020;113: 433–443. doi:10.1177/0141076820956799

21. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374: 86–89. doi:10.1016/S0140-6736(09)60329-9

22. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383: 101–104. doi:10.1016/S0140-6736(13)62329-6

23. Yuan I, Topjian AA, Kurth CD, Kirschen MP, Ward CG, Zhang B, et al. Guide to the statistical analysis plan. Paediatr Anaesth. 2019;29: 237–242. doi:10.1111/pan.13576

24. Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. JAMA. 2012;308: 773–774. doi:10.1001/jama.2012.9502

25. Chambers C. What's next for Registered Reports? Nature. 2019;573: 187–189. doi:10.1038/d41586-019-02674-6

26. Soderberg CK, Errington TM, Schiavone SR, Bottesini J, Thorn FS, Vazire S, et al. Initial evidence of research quality of registered reports compared with the standard publishing model. Nature Human Behaviour. 2021; 1–8. doi:10.1038/s41562-021-01142-4

27. Chambers CD, Mellor DT. Protocol transparency is vital for registered reports. Nature human behaviour. 2018. pp. 791–792. doi:10.1038/s41562-018-0449-6

28. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343: d5928. doi:10.1136/bmj.d5928

29. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355: i4919. doi:10.1136/bmj.i4919

30. Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. Syst Rev. 2018;7: 242. doi:10.1186/s13643-018-0915-2

31. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32: 51–63. doi:10.1016/0021-9681(79)90012-2

32. Andrews N, Miller E, Taylor B, Lingam R, Simmons A, Stowe J, et al. Recall bias, MMR, and autism. Arch Dis Child. 2002;87: 493–494. doi:10.1136/adc.87.6.493

33. Catalogue of Bias - Catalog of Bias. 27 Mar 2017 [cited 15 Nov 2021]. Available: https://catalogofbias.org/

34. Kwok R. How to pick an electronic laboratory notebook. Nature. 2018;560: 269–270. doi:10.1038/d41586-018-05895-3

35. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:

160018. doi:10.1038/sdata.2016.18

36. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015;11: e1004525. doi:10.1371/journal.pcbi.1004525

37. Bishop D. Rein in the four horsemen of irreproducibility. Nature. 2019;568: 435. doi:10.1038/d41586-019-01307-2

38. Held L, Schwab S. Improving the reproducibility of science. Significance. 2020;17: 10–11. doi:10.1111/j.1740-9713.2020.01351.x

39. McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How open science helps researchers succeed. Elife. 2016;5. doi:10.7554/eLife.16800

40. Schönbrodt F. Training students for the Open Science future. Nat Hum Behav. 2019;3: 1031. doi:10.1038/s41562-019-0726-z

41. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nature Human Behaviour. 2017;1: 0021. doi:10.1038/s41562-016-0021

42. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14: 365–376. doi:10.1038/nrn3475

43. Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. Epidemiology. 2018;29: 599–603. doi:10.1097/EDE.0000000000000876

44. Bland JM. The tyranny of power: is there a better way to calculate sample size? BMJ. 2009;339: b3985. doi:10.1136/bmj.b3985

45. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000;53: 1119–1129. doi:10.1016/s0895-4356(00)00242-0

46. Norris D, Dirnagl U, Zigmond MJ, Thompson-Peer K, Chow TT. Health tips for research groups. Nature. 2018;557: 302–304. doi:10.1038/d41586-018-05146-5

47. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31: 337–350. doi:10.1007/s10654-016-0149-3

48. Bland M. An Introduction to Medical Statistics. Oxford University Press; 2015. Available: https://play.google.com/store/books/details?id=fKgXCgAAQBAJ

49. Goodman SN. A comment on replication, p-values and evidence. Stat Med. 1992;11: 875–879. doi:10.1002/sim.4780110705

50. Held L, Pawel S, Schwab S. Replication power and regression to the mean. Significance. 2020;17: 10–11. doi:10.1111/1740-9713.01462

51. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, -J. Wagenmakers E, Berk R, et al. Redefine statistical significance. Nature Human Behaviour. 2017;2: 6–10. doi:10.1038/s41562-017-0189-z

52. Leek JT, Peng RD. Statistics: *P* values are just the tip of the iceberg. Nature. 2015;520: 612. doi:10.1038/520612a

53. Schwab S, Held L. Statistical programming: Small mistakes, big impacts. Significance. 2021;18: 6–7. doi:10.1111/1740-9713.01522

54. Sullivan LM, Weinberg J, Keaney JF Jr. Common Statistical Pitfalls in Basic Science Research. J Am Heart Assoc. 2016;5. doi:10.1161/JAHA.116.004142

55. Kass RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N. Ten Simple Rules for Effective Statistical Practice. PLoS Comput Biol. 2016;12: e1004961. doi:10.1371/journal.pcbi.1004961

56. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull. 1979;86: 638–641. doi:10.1037/0033-2909.86.3.638

57. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291: 2457–2465. doi:10.1001/jama.291.20.2457

58. Altman DG, Martin Bland J. Statistics notes: Absence of evidence is not evidence of absence. BMJ. 1995;311: 485. doi:10.1136/bmj.311.7003.485

59. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567: 305–307. doi:10.1038/d41586-019-00857-9

60. Matthews JNS. Introduction to Randomized Controlled Clinical Trials. 2nd ed. CRC Press LLC; 2017. Available: https://play.google.com/store/books/details?id=nYMCtAEACAAJ

61. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010;303: 2058–2064. doi:10.1001/jama.2010.651

62. Khan MS, Lateef N, Siddiqi TJ, Rehman KA, Alnaimat S, Khan SU, et al. Level and Prevalence of Spin in Published Cardiovascular Randomized Clinical Trial Reports With Statistically Nonsignificant Primary Outcomes: A Systematic Review. JAMA Netw Open. 2019;2: e192622. doi:10.1001/jamanetworkopen.2019.2622

63. Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012;90: 891–904. doi:10.1007/s11192-011-0494-7

64. Chavalarias D, Wallach JD, Li AHT, Ioannidis JPA. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. JAMA. 2016;315: 1141–1148. doi:10.1001/jama.2016.1952

65. Zwet EW, Cator EA. The significance filter, the winner's curse and the need to shrink. Stat Neerl. 2021. doi:10.1111/stan.12241

66. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev. 2009; MR000006. doi:10.1002/14651858.MR000006.pub3

67. Egger M, Smith GD. Bias in location and selection of studies. BMJ. 1998;316: 61–66. doi:10.1136/bmj.316.7124.61

68. Nissen SB, Magidson T, Gross K, Bergstrom CT. Publication bias and the canonization of false facts. Elife. 2016;5. doi:10.7554/eLife.21451

69. Grimes DR, Bauch CT, Ioannidis JPA. Modelling science trustworthiness under publish or perish pressure. R Soc Open Sci. 2018;5: 171511. doi:10.1098/rsos.171511

70. Goldacre B, Drysdale H, Dale A, Milosevic I, Slade E, Hartley P, et al. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. Trials. 2019;20: 118. doi:10.1186/s13063-019-3173-2

71. Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ. 2005;330: 1049. doi:10.1136/bmj.38414.422650.8F

72. Koensgen N, Rombey T, Allers K, Mathes T, Hoffmann F, Pieper D. Comparison of non-Cochrane systematic reviews and their published protocols: differences occurred frequently but were seldom explained. J Clin Epidemiol. 2019;110: 34–41. doi:10.1016/j.jclinepi.2019.02.012

73. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ. 2004;329: 883. doi:10.1136/bmj.38250.571088.55

74. Li G, Abbade LPF, Nwosu I, Jin Y, Leenus A, Maaz M, et al. A systematic review of comparisons between protocols or registrations and full reports in primary biomedical research. BMC Med Res Methodol. 2018;18: 9. doi:10.1186/s12874-017-0465-7

75. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383: 267–276. doi:10.1016/S0140-6736(13)62228-X

76. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal “Instructions to Authors” in 2014. Trials. 2016;17: 301. doi:10.1186/s13063-016-1408-z

77. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev. 2012;1: 60. doi:10.1186/2046-4053-1-60

78. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J, et al. Relation of completeness of reporting of health research to journals’ endorsement of reporting guidelines: systematic review. BMJ. 2014;348: g3804. doi:10.1136/bmj.g3804

79. Sarafoglou A, Hoogeveen S, Matzke D, Wagenmakers E-J. Teaching Good Research Practices: Protocol of a Research Master Course. Psychology Learning & Teaching. 2020;19: 46–59. doi:10.1177/1475725719858807

80. Kabitzke P, Cheng KM, Altevogt B. Guidelines and Initiatives for Good Research Practice. Handb Exp Pharmacol. 2019. doi:10.1007/164_2019_275