Academia represents a paradox. On the one hand, many of the methods and techniques used across disciplines are cutting edge and constantly evolving. However, at the same time, our underlying cultures and working practices remain rooted in the 19th-century model of the independent scientist. Research groups are effectively small, artisanal businesses, each crafting outputs—often exquisite, but the product of the unique skills and processes of that group. Is there room for improvement?

Certainly, there is scope for a more coordinated approach. For example, a degree of standardization in processes and data standards—delivered via foundational and ongoing training—could serve to ensure a common minimum standard, as well as increasing interoperability across the sector. This, in turn, could drive improvements in research quality. But change will require a coordinated approach that recognizes the systems nature of the challenge.

**Systematizing Effective Practice**

What skills are lacking, or exist but are not standardized across the sector? This depends very much on the specific disciplines. For example, in genomics the use of common genotyping chips and imputation platforms enforces a degree of standardization on the data generated and enhances interoperability. Indeed, the widespread availability of summary data from genome-wide association studies, all in a (broadly) common format, has supported the growth of entire methodologies that rely largely or entirely on these data. However, even here there is room for improvement—data are not shared as commonly as they could be, and notation remains non-standard or ambiguous in some cases, which leads to the potential for error.

In other disciplines, however, including many laboratory disciplines *(in vitro, in vivo, animal, human, etc.)* there may be approaches considered standard within a group or laboratory but not across the discipline. Delivering training in basic skills—from file-naming conventions to data curation (including formatting of data files, the use of data dictionaries and codebooks, the need for metadata, the importance of using non-proprietary formats, and other principles of FAIR data, etc.) could go a long way to creating common, sector-wide standards in qualitative disciplines. This training could be built into undergraduate and postgraduate training, not least because these represent key transferable skills relevant to a wide range of careers and industries.

These foundational skills could be considered the base from which to drive forward research quality. This might include training in more discipline-specific methodological skills, and again this would benefit from a degree of standardization of expectations. Here, reporting guidelines may provide a valuable framework. While these are generally intended to describe aspects of a study that should be reported, the importance of many items included arises from their centrality in assessing the validity of a research claim, and thus they speak to important aspects of research design, conduct, and analysis as well as to research reporting. The EQUATOR network ([https://www.equator-network.org](https://www.equator-network.org)) provides reporting guidelines for a range of methodologies across biomedicine—from preclinical animal studies to randomized controlled trials. They therefore provide a framework for methodological training in disciplines where an appropriate guideline exists.

Reporting guidelines are far from perfect—there remain gaps in the coverage as well as duplication (where multiple guidelines exist for similar study designs). Nevertheless, they represent an opportunity for developing a more coordinated approach. They also illustrate the systems nature of the problem. At present, these guidelines are mandated or encouraged by many (but not all) journals, but again focus on the reporting stage when they could perhaps more valuably be incorporated at the study design stage. And this suggests a role for funders, by requiring something similar at the grant submission stage, as well as a role for institutions in terms of the training they offer at undergraduate and postgraduate levels (and beyond).

**Embedding Effective Practice as Standard Practice**

Delivering the highest standards in research is not achieved or sustained without considerable effort. As with many other aspects of our lives, the last 40 years has seen remarkable growth in the measurement of academic performance, ostensibly to improve that...
performance. Surrogate outcomes and metrics have been given a prominent and critical role. A tool that was originally developed to help librarians choose journal subscriptions—the journal impact factor—was repurposed to evaluate individual research outputs and the performance of entire institutions and to inform funding decisions for both. Individual researchers have left the profession, and entire departments have been closed, because of a lack of perceived success under the prevailing “rules of the game.”

Incentive structures are critical; they must be appropriate and in turn allow us to properly evaluate performance without distorting the very thing we seek to improve. Evaluation needs to be rigorous, transparent, multidimensional, and resistant to “gaming” (i.e., manipulations intended to improve performance scores, rather than underlying performance). At present, quality evaluations are too blunt, too infrequent, and too unwieldy to provide focused motivation to drive improvement. Many—for example the UK Research Excellence Framework—require considerable time and effort in their preparation and assessment. Moreover, the imprecision in these evaluations encourages “gaming.”

So how might we provide incentives to systematizing effective practice? The conventional approach of “quality through evaluation” could play a central role, but this would require that any evaluation be richer, deeper, and timelier than is currently the case. The UK Audit Commission has suggested that the desired characteristics of performance measures include that they should avoid perverse incentives and be relevant, clearly defined, verifiable, cost effective, and statistically valid. Our measures of performance should manifest as many of these characteristics as possible.

At present they do not. In certain fields where there is consensus around markers of quality, it is possible to use text-mining approaches to evaluate reporting of key methodological variables. This would complement existing approaches and these tools, and their reporting, might usefully be developed further. In fields where there is no community-derived consensus for markers of methodological quality, this raises the slightly awkward question of what it is that is being measured at present and provides a strong motivation for the development of such consensus.

Conclusion

An alternative to “quality through evaluation” is “quality by design.” If an institution or research group has robust systems in place to manage research quality, then it seems likely that quality will improve more rapidly and more sustainably than research quality for an institution or research group that does not. Moreover, if those systems are shared across institutions, this will not only ensure that quality will improve across those institutions but also that this approach will improve interoperability across those institutions, making it easier, for example, for early-career researchers to move seamlessly between them.

Moreover, while it “seems likely” that these approaches will have the desired effects, this remains a series of hypotheses rather than a set of interventions supported by robust evidence. While increasing incentives for institutions to strengthen their efforts in research improvement in general is desirable, more prescriptive requirements or measures of performance need to be supported by evidence that they influence, or measure, fundamental aspects of research quality rather than epiphenomena susceptible to gaming. Again, working across multiple institutions will provide a platform for more robust evaluation as well as benchmarking. Quality is not zero sum—a rising tide lifts all boats.

Training in foundational skills—and in particular, data skills—remains at the heart of this. Many researchers lack basic formal training in file-naming conventions, version control, and data management. Where training exists, it is often ad hoc and specific to a particular research group, leading to friction and inefficiency when researchers move between groups and institutions. Although institutions may offer relevant training, attendance is often not incentivized and is typically focused on early-career researchers when in fact it is mid-career and senior researchers whose skills are most in need of updating. A systems perspective is necessary, to ensure that training exists and is incentivized and rewarded.

Focusing on data skills also represents an opportunity to foster multidisciplinarity. While the implicit focus of the examples we have discussed is on the life and biomedical sciences, there are many quantitative disciplines where similar issues apply, and many other disciplines—for example, in qualitative research and in the arts and humanities—where quite different issues relating
to data exist. These include interesting conversations about what constitutes data across a range of disciplines and what common solutions might be found—as well as which cases require specific solutions. Primary resource texts, archives, and representations of art are all forms of (often digital) data.

Effective change will require a coordinated approach, involving researchers themselves, institutions, funders, publishers, learned societies, and other sectoral organizations. By developing a consensus across this community regarding what constitutes effective practice, we can launch a cycle of continuous improvement (Figure 1), with training, incentives, and interventions to improve standards, complemented and reinforced by reviews of performance as well as the celebration of success and a general culture of raised ambitions. Ongoing evaluation—research on research, the development and refinement of reporting guidelines, and a culture of engagement across the community—and in particular with research users themselves, will lead to real change.

ACKNOWLEDGMENTS

The UK Reproducibility Network is supported by a number of research funders and other sectoral organizations. See https://www.ukrn.org/ for more details.

REFERENCES

1. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. 37, 658–665.
2. Thelwall, M., Munafo, M., Mas-Bleda, A., Stuart, E., Makita, M., Weigert, V., Keene, C., Khan, N., Drax, K., and Kousha, K. (2020). Is useful research data usually shared? An investigation of genome-wide association study summary statistics. PLoS ONE 15, e0229578.
3. Wootten, R.E., and Sallis, H.M. (2020). Let’s call it the effect allele: a suggestion for GWAS naming conventions. Int. J. Epidemiol. Published online September 2, 2020. https://doi.org/10.1093/ije/dyaa149.
4. Hartwig, F.P., Davies, N.M., Hemani, G., and Davey Smith, G. (2016). Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int. J. Epidemiol. 45, 1717–1726.
5. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018.

About the Organization

The UK Reproducibility Network (https://www.ukrn.org/) is a peer-led consortium that aims to ensure the UK retains its place as a center for world-leading research. Its steering group comprises Marcus Munafo (Bristol), Chris Chambers (Cardiff), Alexandra Collins (Imperial), Laura Fortunato (Oxford), Nicole Janz (Nottingham), and Malcolm Macleod (Edinburgh).