Scientific rigor and the art of motorcycle maintenance

Marcus Munafò, Simon Noble, William J Browne, Dani Brunner, Katherine Button, Joaquim Ferreira, Peter Holmans, Douglas Langbehn, Glyn Lewis, Martin Lindquist, Kate Tilling, Eric-Jan Wagenmakers & Robi Blumenstein

The reliability of scientific research is under scrutiny. A recently convened working group proposes cultural adjustments to incentivize better research practices.

"The real purpose of the scientific method is to make sure Nature hasn't misled you into thinking you know something you actually don't know.... If you get careless or go romanticizing scientific information, giving it a flourish here and there, Nature will soon make a complete fool out of you."

—Robert M. Pirsig,
Zen and the Art of Motorcycle Maintenance:
An Inquiry Into Values

How to ensure quality? In the 1970s the US automobile industry used production methods that relied upon cars being entirely assembled before checking for obvious defects, which resulted in many faulty cars—'lemons'—rolling off the production line and into showrooms. Cars were built to be repaired later rather than to be reliable from sale. Responding to this weakness, Japanese automobile manufacturers implemented practices learned from the statistician W. Edwards Deming, central to which was the rigorous application of quality-control procedures at every step of the manufacturing process. Japanese automobile manufacturers were able to produce cars more reliably and more efficiently, leading to a prolonged period of global dominance.

Like auto manufacturing in the 1970s, scientific research is producing too many lemons.

It is a fitting analogy for the current state of biomedical research, where the low reproducibility of key findings is now being widely discussed[1]. Problems such as publication bias[2], low statistical power[3], data fabrication[4] and questionable research practices[5] are not new, but there is increasing concern that their scale has grown as competition for resources has intensified[6] and, consequently, incentive structures have become distorted. Researchers are susceptible to systemic influences, such as the 'publish or perish' culture and the propensity for journals to prioritize 'significant' novel results, which encourage smaller, quicker, cheaper studies measuring multiple outcomes. The rewards are high—research funding and career advancement—but the consequences serious; poor reproducibility has hindered translation of academic research by the pharmaceutical industry and discouraged investment in drug development[7]. The current scientific career structure works against good scientific practice.

CHDI Foundation (New York, NY)—a sizeable funder of research into Huntington's disease (HD) that seeks to expedite the translation of basic research into therapeutics for HD patients—convened a working group in London in September 2013 to identify practical and implementable policies that could foster a culture that further incentivizes best scientific practice. The HD research community has advantages in that it is a moderately sized pool of committed researchers who encompass diverse disciplines; these characteristics,

Marcus Munafò, William J. Browne, Katherine Button and Kate Tilling are at the University of Bristol, Bristol, UK; Simon Noble & Robi Blumenstein are at CHDI Management/CHDI Foundation, New York, New York, USA; Dani Brunner is at PsychoGenics, Inc., Tarrytown, New York, USA and Columbia University, New York, New York, USA; Joaquim Ferreira is at the University of Lisbon, Portugal; Peter Holmans is at Cardiff University, Cardiff, UK; Douglas Langbehn is at the University of Iowa, Iowa City, Iowa, USA; Glyn Lewis is at University College London, London, UK; Martin Lindquist is at Johns Hopkins University, Baltimore, USA; and Eric-Jan Wagenmakers is at the University of Amsterdam, Amsterdam, Holland.
e-mail: simon.noble@chdifoundation.org

allied with the incentive of an urgent need for effective therapeutics for patients, offers a good testing ground for the introduction of new practices.

Here, we briefly summarize the recent debate surrounding the reliability and reproducibility of biomedical research and then outline tangible steps that CHDI is now taking to further ensure the rigor of the research that it manages and/or funds. We believe these changes will be instructive for other patient foundations and funding organizations seeking to bolster the quality of work in their fields, and they may also suggest ways in which the broader scientific community could re-evaluate the planning, design and registration of biomedical research studies before results are ultimately published in the literature.

The replication problem

Research culture can change, and specific fields have already adopted new practices to increase scientific rigor. In recent decades, the literature of complex genetic traits identified numerous candidate genes but few studies were replicated reliably, leading to years of wasted research[8]. The development of genome-wide association methods necessitated greater statistical stringency and much larger sample sizes, which transformed the field's reliability and identified many robust genetic signals[9]. This required international collaboration, data sharing across large consortia and widespread adoption of practices, such as a clear distinction between discovery and replication samples, independent replication and meta-analysis of studies.

More broadly, statistical and methodological problems in clinical trial reporting in the 1980s and 1990s[10] prompted the development of the CONSORT guidelines[11], and major funding bodies now typically support only clinical trials with specialist statistician input and an independent steering committee to oversee data analysis and trial conduct. This prompted adoption of reporting guidelines in other fields—animal studies (ARRIVE)[12], systematic reviews and meta-analyses (PRISMA)[13], and observational epidemiology (STROBE)[14]. *Nature* and affiliated journals have also recently introduced more stringent checklists to improve the reporting of methodological and statistical information[15], as have others[12].

Greater emphasis on replicating initial findings is clearly important, and essential if science is to be self-correcting, but conducting resource-intensive direct replication studies currently offers scant reward, including few publishing opportunities. Journals have an important role to play here; *Nature Biotechnology* recently published a replication study and discussed its importance[16]. Journals could openly solicit replication of key published findings, with publication guaranteed[17], but these efforts would require funding. CHDI is now considering introducing an option into its research funding agreements to pause publication of selected studies, solicit (and fund) replication through a mutually agreed upon independent academic laboratory or contract research organization, and then have the original and replication researchers publish jointly with appropriate recognition. This approach clearly entails substantial cultural change and may be perceived as counter to the strongly held notion of academic freedom, but ultimately one hopes this would be seen as beneficial to all interested parties, including the wider scientific community.

Beyond simple replication

More prosaically, CHDI recently worked with collaborators and members of the HD research community to compile a freely available manual[18] defining experimental best practices for HD mouse models to aid reproducibility (and clinical translatability); further community input will be built into future iterations.

> Data sharing and deposition in public repositories can bring sunlight to the scientific process, increasing transparency and reproducibility

Even so, replication alone will always be an inefficient, retrospective fix; unless we strive to ensure quality throughout the research process, too many lemons will still be produced. A better appreciation of and adherence to appropriate statistical plans would be a significant improvement (pun intended) to help ensure adequate sample sizes and appropriate power. A large, well-powered study that gives conclusive results may be a more efficient and ethical (when humans or other animals are involved) use of resources than a series of inconclusive studies[19]. When sample size is constrained by other factors (e.g., cost, scarcity or ethical considerations) then type I and type II error rates can be set at non-conventional levels to optimize true discovery rates[20]. In addition, most disciplines place too much emphasis on *P*-value thresholds—'finding things' or 'not finding things'—and not enough on the extent to which we make comparisons with varying degrees of accuracy. We should think quantitatively about the confidence we have in our findings, and the precision of our effect size estimates, and explore alternative statistical approaches (e.g., Bayesian methods)[21].

CHDI is now looking into ways to provide this statistical and methodological training—such as developing online courses in conjunction with Coursera (http://www.coursera.org/) that postdoctoral researchers in funded laboratories will have to complete—to develop expertise in future research leaders. Importantly, this training will be augmented by an independent standing committee (including some of the authors of this Commentary) comprising experts from outside the HD research field (to prevent conflicts of interest) to review and offer advice on the suitability of the statistics and methods proposed in research protocols. This standing committee will be able to call on external expertise as required, and perhaps evolve to become a more supportive, proactive resource for research groups.

CHDI will also create a repository for protocols reviewed by the independent standing committee: upon study completion these will be made publicly available so that research findings can be judged against a priori hypotheses and planned statistical analyses. Pre-registration of clinical trials protocols is increasingly required by journals, promoting transparent reporting and preventing 'HARKing' (hypothesizing after results are known)[22].

Recently, *Cortex* and *Perspectives on Psychological Science* have introduced pre-registration for human studies that commit the journal to publication based on the quality of the protocol, regardless of eventual outcome (for further details, see https://osf.io/8mpji/wiki/home/). The Declaration of Helsinki also now recommends pre-registration of all studies involving human participants, not just clinical trials as previously[23]. Such practices increase transparency while addressing the distorting effects of the current fixation on only publishing 'positive' results and not null findings. In that regard, in 2010 CHDI funded the establishment of *PLOS Currents: Huntington Disease*, a free, archived, indexed and citable, open-access online journal to encourage unconventional scientific reports, including brief observations and 'negative' data.

Data sharing and deposition in public repositories can bring sunlight to the scientific process, increasing transparency and reproducibility, and CHDI is exploring ways to foster this data sharing; for example, de-identified clinical data from the ongoing worldwide observational study Enroll-HD will be made available to any interested researcher, and a website now in development will make accessible (as soon as curated and before publication whenever possible) large data sets that CHDI funding has helped generate (data sets currently hosted at http://www.chdifoundation.org/datasets/).

Some journals are also revisiting adherence to their own policies requiring the data in published studies to be made publicly available.

Conclusions

Science is conducted on the principle that it is self-correcting, but the extent to which this is true is an empirical question[24]. The more that quality control becomes integrated into the scientific process itself, the more the whole process becomes one of continual improvement. Implementing this at the level of production implies a culture of incentivizing, educating and empowering those responsible for production, rather than policing quality after the fact with 'quality inspectors' (i.e., peer reviewers) or, even more distally, requiring attempts at replication. We think this insight, applied successfully to automobile manufacturing in the 1970s, can also be profitably applied to the practice of scientific research to build a more solid foundation of knowledge and accelerate the research endeavor.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.3004).

1. Johnson, G. New truths that only one can see. *New York Times* 21 January (2014).
2. Greenwald, A.G. *Psychol. Bull.* **82**, 1–20 (1975).
3. Sedlmeier, P. & Gigerenzer, G. *Psychol. Bull.* **105**, 309–316 (1989).
4. Edwards, A.W. *Biol. Rev. Camb. Philos. Soc.* **61**, 295–312 (1986).
5. Martinson, B.C., Anderson, M.S. & de Vries, R. *Nature* **435**, 737–738 (2005).
6. Timmer, J. *Ars Technica* http://arstechnica.com/science/2007/04/how-doubling-the-nihs-budget-created-a-funding-crisis/ (20 April 2007).
7. Brunner, D., Balci, F. & Ludvig, E.A. *Behav. Processes* **89**, 187–195 (2012).
8. Ioannidis, J.P. *et al. Nat. Genet.* **29**, 306–309 (2001).
9. Hirschhorn, J.N. *N. Engl. J. Med.* **360**, 1699–1701 (2009).
10. Pocock, S.J., Hughes, M.D. & Lee, R.J. *N. Engl. J. Med.* **317**, 426–432 (1987).
11. Schulz, K.F. *et al. Br. Med. J.* **340**, c332 (2010).
12. Kilkenny, C. *et al. PLoS Biol.* **8**, e1000412 (2010).
13. Welch, V. *et al. PLoS Med.* **9**, e1001333 (2012).
14. von Elm, E. *et al. PLoS Med.* **4**, e296 (2007).
15. Editorial. *Nat. Methods* **10**, 367 (2013).
16. Editorial. *Nat. Biotechnol.* **31**, 943 (2013).
17. Wagenmakers, E.J. & Forstmann, B.U. *Cortex* **51**, 105–106 (2014).
18. Menalled, L. et al. *A Field Guide to Working with Mouse Models of Huntington's Disease* (The Jackson Laboratory, 2014).
19. Button, K.S. *et al. Nat. Rev. Neurosci.* **14**, 365–376 (2013).
20. Ioannidis, J.P., Hozo, I. & Djulbegovic, B. *J. Clin. Epidemiol.* **66**, 903–910 (2013).
21. Liang, F. *et al. J. Am. Stat. Assoc.* **103**, 410–423 (2008).
22. Kerr, N.L. *Pers. Soc. Psychol. Rev.* **2**, 196–217 (1998).
23. The World Medical Association. *J. Am. Med. Assoc.* **310**, 2191–2194 (2013).
24. Ioannidis, J.P. *Perspect. Psychol. Sci.* **7**, 645–654 (2012).