BMJ Open

Quality Output Checklist and Content Assessment (QuOCCA): a new tool for assessing research quality and reproducibility

Martin E Héroux [1,2] Annie A Butler,[1,2] Aidan G Cashin [1,2] Euan J McCaughey,[1,3] Andrew J Affleck,[1,4] Michael A Green,[1,2] Andrew Cartwright,[1] Matthew Jones,[2] Kim M Kiely,[1,2] Kimberley S van Schooten,[1,2] Jasmine C Menant,[1,2] Michael Wewege,[1,2] Simon C Gandevia[1,2]

ABSTRACT

Research must be well designed, properly conducted and clearly and transparently reported. Our independent medical research institute wanted a simple, generic tool to assess the quality of the research conducted by its researchers, with the goal of identifying areas that could be improved through targeted educational activities. Unfortunately, none was available, thus we devised our own. Here, we report development of the Quality Output Checklist and Content Assessment (QuOCCA), and its application to publications from our institute's scientists. Following consensus meetings and external review by statistical and methodological experts, 11 items were selected for the final version of the QuOCCA: research transparency (items 1–3), research design and analysis (items 4–6) and research reporting practices (items 7–11). Five pairs of raters assessed all 231 articles published in 2017 and 221 in 2018 by researchers at our institute. Overall, the results were similar between years and revealed limited engagement with several recommended practices highlighted in the QuOCCA. These results will be useful to guide educational initiatives and their effectiveness. The QuOCCA is brief and focuses on broadly applicable and relevant concepts to open, high-quality, reproducible and well-reported science. Thus, the QuOCCA could be used by other biomedical institutions and individual researchers to evaluate research publications, assess changes in research practice over time and guide the discussion about high-quality, open science. Given its generic nature, the QuOCCA may also be useful in other research disciplines.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We developed a simple and broadly applicable tool to assess key aspects of research quality in biomedical publications.
⇒ The Quality Output Checklist and Content Assessment (QuOCCA) includes 11 items that focus on research transparency, research design and analysis and research reporting practices.
⇒ The QuOCCA intentionally does not provide an overall score as its primary goal is to promote discussion and education about high-quality, open science.
⇒ The QuOCCA is accompanied by a detailed instructional guide to assist and educate users.
⇒ Given its generic and broad scope, some QuOCCA items may not be relevant to certain fields.

initiatives that aim to support, guide and even coax researchers into conducting more open, rigorous and reproducible science are increasingly common. For example, several funding bodies and governmental agencies now reward, or at least request, more open, reproducible, higher quality research.[26–28] An increasing number of journals have adopted policies, checklists and reporting guidelines that aim to improve the quality, transparency, reproducibility and reporting of the research they publish.[29–33] However, there is little uniformity in what is required, and fundamental aspects of good scientific reporting are often overlooked. There is also a growing library of educational resources produced by funding bodies, research institutions, journals and researchers that address different facets of the problem.[34–36]

Aware of the issues currently affecting science, our independent medical research institute has taken formal steps to assist its researchers to conduct more open, rigorous and reproducible science. One such step

INTRODUCTION

The goal of biomedical research is to generate new knowledge that is reproducible and, where possible, useful to clinicians and policymakers. To achieve these goals, research must be well designed, properly conducted and clearly and transparently reported. For a variety of reasons,[1–11] a substantial portion of the published literature falls short of the mark.[12–25] In response to these trends,

was the formation of a Research Quality Committee in 2018[37] with a mandate to (1) raise awareness about the importance of research reproducibility and quality, (2) educate and train researchers to improve research quality and the use of appropriate statistical methods, (3) foster an environment in which robust science and the validity of all research findings are prioritised, (4) promote open discussion of research reproducibility and quality, (5) promote a culture of open high-quality science by encouraging strategies such as preregistration of plans, making data available and putting unpublished work in a publicly accessible location and (6) seek broad adoption of improved research quality and reproducibility at national and international levels.

The committee required a tool that could assess the quality of research published by the institute's scientists in a consistent way. Areas of concern could then be identified and form the basis for the Committee's educational initiatives to improve research quality and reproducibility. Our institute has a diverse range of research areas, from genomics, cellular physiology and medical imaging to human physiology, biomechanics, psychology, population health and clinical trials. Hence, the tool needed to focus on key concepts related to research transparency, research design and analysis and research reporting, which were applicable across a wide range of studies, rather than focused on individual fields. A search of the published literature and the reporting guidelines available at the Equator Network[38] revealed that no such tool existed. Thus, we developed the Quality Output Checklist and Content Assessment (QuOCCA), named in honour of the quokka, a small marsupial native to Australia.

This *Communication* describes the development of the QuOCCA, its theoretical underpinnings, and its application to publications from our medical research institute spanning 2 years.

MATERIALS AND METHODS
Developing the QuOCCA

The QuOCCA was developed after several formal and informal meetings of members of the Research Quality Committee. A total of 13 committee members took part: three final-year PhD students, seven senior postdoctoral researchers, two group leaders and the IT manager who also has a PhD in neurophysiology. The key research areas of the members spanned clinical medicine, population health, physiology, histology, meta-research, cybersecurity, research reproducibility, evidence synthesis, biomechanics, biomedical engineering, MRI and physics (see online supplemental appendix 1).

Key guiding principles were first identified. The QuOCCA should target peer-reviewed publications of original findings, not editorials, letters, book chapters and reviews. It should include relatively few items, and these should focus on research transparency, research design and analysis and research reporting. Regardless of the topic area of the publication, anyone with a working

knowledge of biomedical research should be able to administer the tool; no field-specific or statistical expertise should be required.

An initial version of the tool was drafted in 2019 and piloted on 10 papers published by institute researchers that spanned a range of disciplines. Feedback was also sought from several institute researchers who were not members of the committee and who had a variety of research backgrounds. These steps helped improve the wording of certain items. They also helped eliminate items that were too difficult or time-consuming to answer (eg, 'Are the methods described in sufficient detail (to allow replication)?'; 'Are reporting guidelines (eg, ARRIVE,[39] CONSORT,[40] PRISMA[41] adhered to?') or dwelt on statistical practices that were limited in scope or could be considered contentious (eg, 'Are any one-sided tests used and, if so, is this justified in the text?'). On this latter point, the committee felt that, while issues of statistical practices and statistical reporting were relevant, expertise is often required to determine their appropriateness. Moreover, issues of statistical reporting are covered in detail by the SAMPL (Statistical Analyses and Methods in the Published Literature) Guidelines.[42]

Final QuOCCA items

Following consensus meetings and external review by statistical and methodological experts, 11 items were selected for the final version of the QuOCCA (figure 1, online supplemental appendix 2). The QuOCCA was designed to guide discussion, not rate and rank researchers or their publications. Thus, a decision was made to not generate a total score as this can lead to superficial changes in research practice aimed at obtaining higher scores, rather than fundamental changes that generate more open, rigorous and reproducible research.

Research transparency
Item 1a, b

Registration of the experimental design, methods and analyses is one of the simplest ways to improve the rigour of research.[8 43–45] It is mandatory for clinical trials[46] and now encouraged across all research types.[47–52] It reduces the influence of hindsight bias and prevents further analyses to be carried out or hypotheses to be formulated after the data have been collected and analysed.[53] It also protects against cherry picking—the selective publication of findings.[19 54 55] Importantly, registration does not preclude exploratory (unregistered) analyses and the report of serendipitous findings. Also, purely exploratory research has a critical place in science. However, it should be reported as such.

Item 2

Transparency and open science dictate that data underpinning a scientific publication should be made publicly available for others to reanalyse and reuse, including in meta-analyses.[27 56–58] This is also essential to enable

QuOCCA Quality Output Checklist and Content Assessment			
This checklist is intended for peer-reviewed research papers. It should not be used for reviews, chapters, editorials etc.			

Manuscript Title:	
Manuscript Authors:	
Person submitting this form:	
Date:	

TRANSPARENCY:	N/A	YES	NO
1a. Were the study's hypotheses and analyses plans **registered** prior to the conduct of the study (i.e. pre-registered)?		☐	☐
b. If so, was the main conclusion reported in the abstract (or summary) based on the primary hypothesis/outcome?	☐	☐	☐
2. Are the **primary data accessible** to independent researchers on a public website?		☐	☐
3. Is **code used for the study available** on a public website to allow for reproduction or analysis of data?		☐	☐
DESIGN AND ANALYSIS:	N/A	YES	NO
4. Was **ethics** approval obtained?	☐	☐	☐
5a. Was the **sample size** based on a formal sample size calculation done prior to starting the study?	☐	☐	☐
b. If so, was the planned sample size adhered to?	☐	☐	☐
6. Was data analysis **blinded**?	☐	☐	☐
REPORTING PRACTICES:	N/A	YES	NO
7. Are any **reporting guidelines** specified (such as those found at www.equator-network.org)?		☐	☐
8a. Are all **measures of variability** defined in figures, tables and text?	☐	☐	☐
b. Are any data summarised using standard error of the mean (SEM)?	☐	☐	☐
c. If the SEM is used, are sample sizes specified for all reported SEM?	☐	☐	☐
9a. Were any data excluded?	☐	☐	☐
b. If so, was a criterion given?	☐	☐	☐
10a. If null-hypothesis testing of significance was used, is a **probability threshold** specified for all statistical tests?	☐	☐	☐
b. If used, are **exact probability values** used throughout the report, excluding figure legends?	☐	☐	☐
11. Are claims made for the importance or significance of results associated with a P-value greater than or equal to 0.05 (or other threshold) i.e. misleading spin of reported results?	☐	☐	☐

Figure 1 Items of the Quality Output Checklist and Content Assessment (QuOCCA).

external parties to reproduce published research, ensuring its integrity.[59]

Item 3
This item focuses on the availability of analytic code used to generate the results.[8 60] As with data availability, the availability of code is essential for reproducibility.[8 60 61] This should not be inferred based on whether or not primary data are accessible.[62]

Research design and analysis
Item 4
Research conducted on, with or about participants (whether human or not), or their tissue or data, should have ethical approval from an appropriate institutional ethics committee or institutional review board (eg, university, hospital, regional or other) before the research begins. This is crucial to ensure the safety and well-being of participants and offers an opportunity for external review of the proposed research.

Item 5a, b
Prior to collecting data, researchers should determine the sample size required for their study to have sufficient statistical power or to generate precise estimates of the investigated effect.[13 63 64] This is essential from a participant burden and safety perspective as it ensures that the study can answer the specified research question.[65] It is also essential from a research quality perspective as it

increases the likelihood that genuine effects are precisely estimated.[63 64] Such calculations are not always warranted or possible. However, in these cases, care must be taken when reporting and interpreting results from statistical analyses. Moreover, if the study is exploratory, it should be reported as such.

Item 6

Blinded or masked analysis of data, like blinding of participants and researchers, minimise the impact of cognitive biases on research results.[66] Blinded analysis ensures that data are analysed impartially. This step may not be required in exploratory research.

Reporting practices
Item 7

Reporting guidelines have become increasingly used as a way to standardise what is reported in publications. For the most part, they specify minimal standards for the reporting of study methods and results. They cover issues which, if not addressed, can produce bias.[67] There are a growing number and range of reporting guidelines, many of which are collected under the EQUATOR network.[38 40 68]

Item 8a–c

Science must be accurately reported. This includes specifying what measures of variability and confidence were used, as these details are required to properly understand the data and results. Examples of such measures are the SD, IQR, 95% CIs and SE of the mean (SEM). However, despite high use in certain biomedical disciplines,[21 69] the SEM is rarely the appropriate summary statistic[70 71] and is often misinterpreted.[72] At a minimum, if the SEM is used, it should be accompanied by the sample size as this allows for the SD to be computed.

Item 9a, b

To exclude data, post hoc without a clear justification is a questionable research practice.[19 55 73 74] Thus, it is important for researchers to justify the exclusion of data and to specify what criterion was used.

Item 10a, b

A probability threshold, usually denoted as a p value or alpha level, is used to determine whether or not the null hypothesis can be rejected.[75] In biomedical research, the probability threshold is usually set at 0.05. However, it should be set based on the aims of the study and the study design.[71] Moreover, there are calls for more conservative thresholds to be used to reduce the chance of false-positive findings.[76] To properly interpret statistical results, exact p values should be reported.[71 77] The focus on null hypothesis statistical testing and p values for this item and the next item was intentional. Statistical analyses that focus on the null hypothesis and p values are often poorly reported and misinterpreted, and they underpin the reproducibility crisis.[13 17 21 24] Researchers are increasingly encouraged to focus on CIs (or credible intervals) when

they report and interpret their statistical results[24 64 78] with some moving towards Bayesian statistics. While these newer reporting practices and statistical approaches may also be improperly reported,[79] the QuOCCA was designed to focus on the key source of problems in biomedical sciences: null hypothesis statistical testing.

Item 11

Spin includes the reporting practice in which results are presented in a more favourable light than can be justified by the data.[23 80] This questionable research practice occurs frequently across the biomedical sciences.[21 23 69] This item used a strict definition of spin based on the reported threshold for statistical significance.

QuOCCA instructional guide

We developed an Instructional Guide to help people who intend to administer the QuOCCA (online supplemental appendix 2). The guide provides explicit details for each item, including its scope, definitions of key terms, exact interpretation of 'yes', 'no' and 'N/A' and several examples.

Institutional implementation of the QuOCCA

A committee member identified and collated the full text for all research articles, which included any member of institutional staff as a named author between January 2017 and December 2018. These papers were distributed to pairs of committee members who independently screened and assessed eligible full-text articles. Data were collected using a standardised form hosted on RedCap.[81 82] Conflicts in QuOCCA ratings were identified and resolved by discussion between the pairs of raters. Finally, the QuOCCA should be quick to administer and should not require the assessor to read and understand the entire paper. Once familiar with the tool, an assessor should be able to search the paper for relevant information.

RESULTS

Five pairs of raters assessed 221 articles published in 2017 and 231 articles published in 2018. The mean time to assess each paper was 10 min (SD 3). QuOCCA results for 2017 and 2018 were similar (figure 2, online supplemental appendix 3). Some QuOCCA items include more than one question. In total, there are 14 primary questions and 4 follow-up questions (ie, 1b, 5b, 8c, 9b). In our audit of articles from 2017 to 2018, items 1a, 2, 3 and 7 were relevant to 100% of the articles, whereas items 4, 5a, 6, 8a, 8b, 9a, 10a, 10b, 11 were relevant to 89–99% of the articles.

The items related to research transparency revealed that less than 10% of studies were registered or made their data or code available. As for items related to research design and analysis, ~95% of studies indicated they obtained ethical approval. However, <20% of papers based their sample size on formal sample size calculations, although, when this was
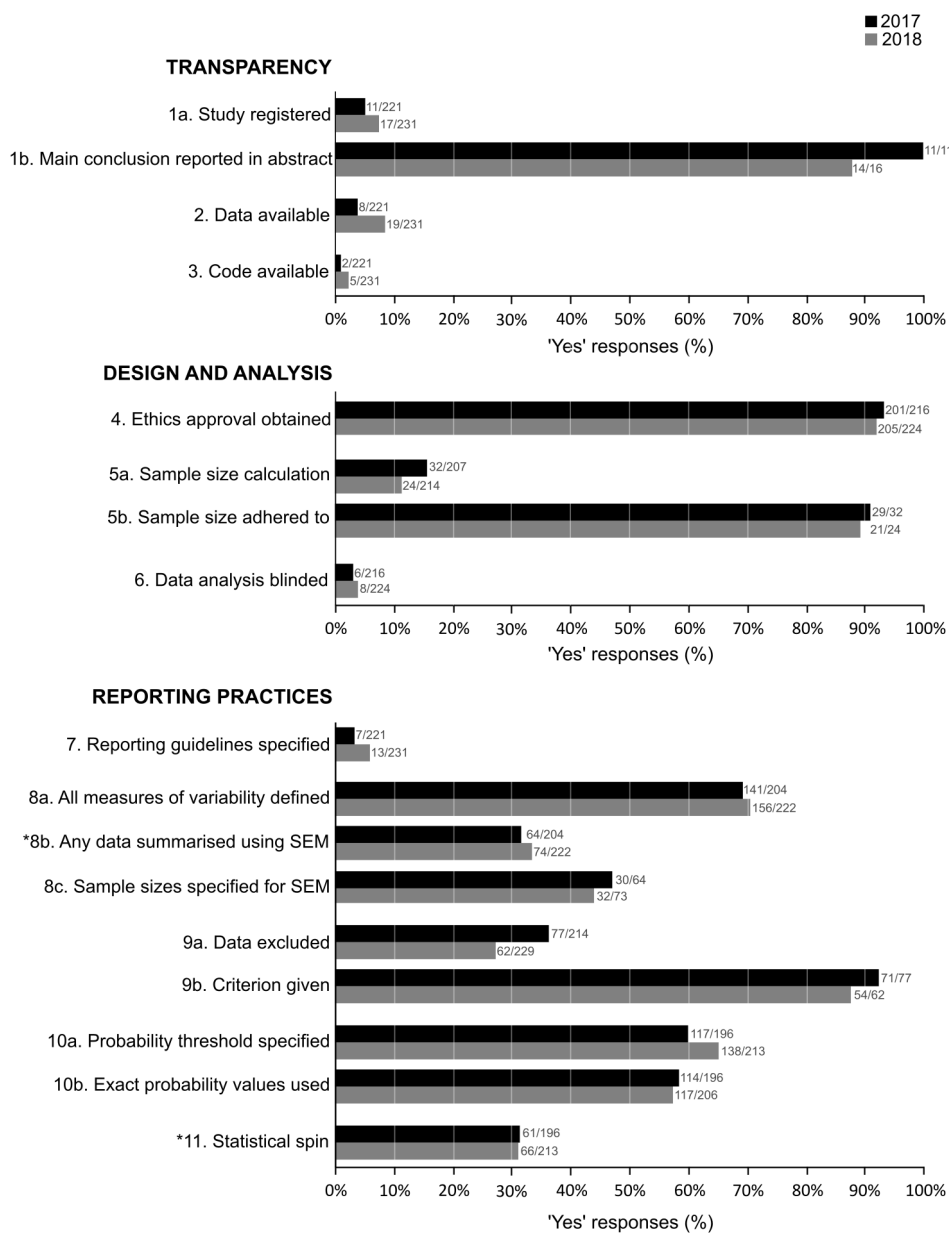
Figure 2 Descriptive results of responses to QuOCCA items. The count and percentage of 'Yes' responses (2017=black; 2018=grey) for each item of the QuOCCA. A total of 221 articles were audited for 2017 and 231 for 2018. For primary questions, the number of 'not applicable' (ie, N/A) responses can be determined by comparing their denominators to the total number of articles considered per year. For follow-up questions (ie, 1b, 5b, 8c, 9b), the number of N/A responses can be determined by comparing their denominators to the numerators of the questions that precedes them. QuOCCA, Quality Output Checklist and Content Assessment; SEM, standard error of the mean. *Items where 'Yes' response indicates a reporting practice that should be avoided.

done it was generally adhered to (85%–90%). Also, <5% of studies performed blinded analysis. Finally, items related to reporting practices revealed that less than 10% of papers used a reporting guideline. Also, ~30% of papers included measures of variability or confidence that were not defined. Also, ~30% of papers included SEM, with the accompanying sample size specified less than half the time. Similarly, ~30% of papers indicated they excluded data, with most of these papers (87%–92%), specifying the criterion used to make these decisions. Only ~60% of papers specified a probability threshold for all statistical tests; a similar percentage consistently reported exact p values. Finally, 30% of all papers included spin (eg, interpreting p values greater than 0.05 as statistical trends).

DISCUSSION
The research diversity at our institute and our consensus-based development approach ensured that the retained

items in our research checklist were broadly applicable and relevant to open, high-quality, reproducible and well-reported science. Thus, the QuOCCA can be used by other institutions, individual researchers, publishers or funding bodies to evaluate research publications, assess changes in research practice over time and guide the discussion on high-quality, open science.

Like other tools, the QuOCCA focuses on published papers. This means that, for some items, it is not possible to determine whether non-adherence to an item is due to genuine non-adherence or incomplete reporting. Also, certain QuOCCA items may focus on a practice that is not problematic in certain fields, or that may not be relevant depending on the type of paper being assessed. Having said this, our audit of over 450 papers spanning a wide variety of research areas found that QuOCCA items were relevant in 89%–100% of cases.

Like other tools, the quality of the assessment depends on how easy it is to administer and how familiar assessors are with it, which is why we developed the QuOCCA Instructional Guide. Despite all assessors being familiar with the Guide, those who had previously conducted similar assessments tended to make fewer mistakes; this was anecdotally reported by assessors after their consensus meetings. Also, items that required searching (eg, data exclusion) or verification (eg, exact p values and spin) had higher rates of initial disagreement between raters (20%–30%). Anecdotally, some assessors also found it more difficult to assess papers that were not in their field; for example, a population health researcher assessing a molecular physiology paper. Hence, we believe that assessment will always be more accurate if papers are assessed by pairs of assessors. Thus, we recommend that junior assessors be paired with more experienced assessors, and that assessors evaluate papers from somewhat related fields. When the number of assessed papers is small, a paper copy of the QuOCCA can be used and conflicts can be identified and resolved by the pair(s) of assessors. However, when the number of papers is substantial, it may be more efficient to use software tools like RedCap for data entry because the output from these tools allows the detection of conflicts to be automated and the data to be analysed, summarised and visualised.

The assessment of 2 years of publications from our institute revealed limited engagement with several recommended practices. These results will be useful to guide future educational initiatives, the effectiveness of which can be assessed with the QuOCCA. Our audit included all papers where one or more institute researchers appeared as authors. Thus, in some cases, the research was conducted and the paper drafted by colleagues from other institutions. It can be difficult to raise issues of research quality with colleagues, especially when the work is being led by others. Thus, to assess the effectiveness of our educational initiative, we may choose to focus on papers where the senior or corresponding author is from our institution. Since our audit was performed on papers published by researchers in various disciplines,

the results may not reflect the reporting practice of any one discipline. For example, papers in epidemiology and population health rarely use the SEM to summarise variability. When the QuOCCA is used by a multidisciplinary institute, it may be informative to provide a breakdown of results by discipline, which would allow for more targeted educational examples and activities.

The QuOCCA is an assessment tool. It was not designed to be a replacement for reporting guidelines that assess statistical methods or particular study designs,[38 83–86] nor was it designed to be used in tandem with specific reporting guidelines. It is a tool that can be used to ascertain the openness and quality of published research, methods and reporting. Quality, in the context of the QuOCCA, refers to the inclusion of research practices that are linked to open, transparent, well-reported science. A paper that addresses all the items in the QuOCCA will be more reproducible, more transparent and likely of higher quality. It remains the responsibility of researchers and journals to refer to the Equator Network[38] to identify relevant reporting guidelines for specific study types.

An individual researcher who uses the QuOCCA is necessarily driven to improve. However, what is less clear is how institutions and journals can best use QuOCCA results. Our own institute has chosen to use these results to devise targeted educational material and activities. However, should feedback be provided to researchers? While this may lead to changes in reporting practices in some researchers, it may also lead to resentment and a justification of current reporting practices in others (eg, 'This is how everyone in my field does it.', 'Registration is mostly relevant to clinical research, not the basic sciences.'). How best to use QuOCCA results is a question our committee plans to tackle in the coming years.

The QuOCCA and the accompanying Instructional Guide are both freely available (figure 1, online supplemental appendix 2), and a dedicated website will include any future versions (https://www.neura.edu.au/about/research-quality/quocca/). Moreover, as part of our Institute's commitment to improving research quality and reproducibility, members of our Research Quality Committee are working to produce video-based educational material for researchers, institutions and journals interested in using the QuOCCA; when completed, this material will be available on the QuOCCA website. To increase its visibility, QuOCCA workshops will be offered at academic and research institutes and relevant conferences. However, we would like to reiterate that the QuOCCA should not be adopted as a tick-box checklist or a reporting guideline to be completed by authors or reviewers at the time of submission or publication. Rather, it should be viewed as an assessment tool to help researchers, research institutions and journals identify areas where improvement and education are needed.

Development of this type of tool is not simple, and we have learnt valuable lessons. Although we tried to be transparent, it would have been better to have kept more records of feedback we received and the rationale

for some decisions. Another uncertainty is our lack of guidance on how to best implement the QuOCCA and apply its results. Should QuOCCA results be transmitted directly to researchers or communicated at an institutional or journal level? Should the focus be on educational activities or on policing and enforcement? Ideally, training researchers would suffice. However, reporting practices may not improve in response to educational initiatives.[21] But enforcement tends to work only as long as it is in place.[87] This suggests that genuine change in reporting practices may not have occurred. These challenges will need to be addressed if scientific reporting is to improve. Finally, we acknowledge that the QuOCCA may not be relevant to researchers in some disciplines. It was not tailored to research areas at our institute but was designed to apply broadly and address some key issues highlighted by us and others in the biomedical literature. Nevertheless, it is probably unavoidable that the QuOCCA (or a subset of items) may not be useful to some researchers, institutions or journals. Also, in future, the QuOCCA may need to be updated, if certain reporting practices highlighted in the QuOCCA are no longer problematic. For example, this would occur if authors were required to register their protocol in order to make strong confirmatory claims.

Conclusion

There is a need to change research reporting practice to ensure that the scientific community and the public have access to accurate and complete records of research.[88 89] Similarly, widespread endorsement and implementation of high-quality and reproducible research helps gain the maximal value from clinical and non-clinical medical research. In line with these objectives, the QuOCCA evaluates published research and serves as a benchmark to guide improvements to the openness, transparency and quality of published research. For maximal impact, the QuOCCA and its results should be paired with a tailored educational programme that increases awareness and engagement with transparent and reproducible research practices.

Author affiliations
¹Neuroscience Research Australia, Sydney, New South Wales, Australia
²University of New South Wales, Sydney, New South Wales, Australia
³Queen Elizabeth National Spinal Injuries Unit, Queen Elizabeth University Hospital Campus, Glasgow, UK
⁴Department of Neuropathology, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia

ORCID iDs
Martin E Héroux http://orcid.org/0000-0002-3354-7104
Aidan G Cashin http://orcid.org/0000-0003-4190-7912

REFERENCES

1 McGauran N, Wieseler B, Kreis J, et al. Reporting bias in medical research - a narrative review. *Trials* 2010;11:37.
2 Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US States data. *PLoS One* 2010;5:e10271.
3 John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 2012;23:524–32.
4 Nosek BA, Spies JR, Motyl M. Scientific utopia: II. restructuring incentives and practices to promote truth over publishability. *Perspect Psychol Sci* 2012;7:615–31.
5 Fanelli D. Redefine misconduct as distorted reporting. *Nature* 2013;494:149.
6 Brembs B, Button K, Munafò M. Deep impact: unintended consequences of journal rank. *Front Hum Neurosci* 2013;7:291.
7 Smaldino PE, McElreath R. The natural selection of bad science. *R Soc Open Sci* 2016;3:160384.
8 Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nat Hum Behav* 2017;1:0021.
9 Grimes DR, Bauch CT, Ioannidis JPA. Modelling science trustworthiness under publish or perish pressure. *R Soc Open Sci* 2018;5:171511.
10 Bruton SV, Medlin M, Brown M, et al. Personal motivations and systemic incentives: scientists on questionable research practices. *Sci Eng Ethics* 2020;26:1531–47.
11 Niles MT, Schimanski LA, McKiernan EC, et al. Why we publish where we do: faculty publishing values and their relationship to review, promotion and tenure expectations. *PLoS One* 2020;15:e0228914.
12 Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017. *PLoS Biol* 2018;16:e2006930.
13 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
14 Ioannidis JPA, Allison DB, Ball CA, et al. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009;41:149–55.

15 Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011;10:712.

16 Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012;483:531–3.

17 Open Science Collaboration. Psychology. estimating the reproducibility of psychological science. *Science* 2015;349:aac4716.

18 Dumas-Mallet E, Button KS, Boraud T, *et al*. Low statistical power in biomedical science: a review of three human research domains. *R Soc Open Sci* 2017;4:160254.

19 Héroux ME, Loo CK, Taylor JL, *et al*. Questionable science and reproducibility in electrical brain stimulation research. *PLoS One* 2017;12:e0175635.

20 Stevanovic A, Schmitz S, Rossaint R, *et al*. CONSORT item reporting quality in the top ten ranked journals of critical care medicine in 2011: a retrospective analysis. *PLoS One* 2015;10:e0128061.

21 Diong J, Butler AA, Gandevia SC, *et al*. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PLoS One* 2018;13:e0202121.

22 Li G, Abbade LPF, Nwosu I, *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.

23 Chiu K, Grundy Q, Bero L. 'Spin' in published biomedical literature: a methodological systematic review. *PLoS Biol* 2017;15:e2002173.

24 Boutron I, Ravaud P. Misrepresentation and distortion of research in biomedical literature. *Proc Natl Acad Sci U S A* 2018;115:2613–9.

25 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4:e5738.

26 Reproducibility and reliability of biomedical research: improving research practice. Available: http://www.acmedsci.ac.uk/policy/ policy-projects/ reproducibility-and-reliability-of-biomedical-research/ [Accessed 5 Oct 2021].

27 McKiernan EC, Bourne PE, Brown CT, *et al*. How open science helps researchers succeed. *Elife* 2016;5:16800. doi:10.7554/eLife.16800

28 Burgelman J-C, Pascu C, Szkuta K, *et al*. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Front Big Data* 2019;2:43.

29 Yosten GLC, Adams JC, Bennett CN, *et al*. Revised guidelines to enhance the rigor and reproducibility of research published in American physiological Society journals. *Am J Physiol Regul Integr Comp Physiol* 2018;315:R1251–3.

30 Prager EM, Chambers KE, Plotkin JL, *et al*. Improving transparency and scientific rigor in academic publishing. *Brain Behav* 2019;9:e01141.

31 Forsythe ID, Howells S, Barrett KE. Reproducibility and data presentation. *J Physiol* 2019;597:5313.

32 Promoting awareness and knowledge to enhance scientific rigor in neuroscience. Available: https://neuronline.sfn.org/collection/ promoting-awareness-and-knowledge-to-enhance-scientific-rigor-in-neuroscience [Accessed 5 Oct 2021].

33 Initial submission. Available: http://www.nature.com/nature/for-authors/initial-submission [Accessed 5 Oct 2021].

34 Reproducibility resources and guidelines by topic. Available: https:// research.columbia.edu/reproducibility-resources-and-guidelines-topic [Accessed 7 Oct 2021].

35 Training and other resources. Available: https://grants.nih.gov/policy/ reproducibility/training.htm [Accessed 5 Oct 2021].

36 Symposium on instruction in reproducible research. Available: https://www.projecttier.org/events/conference-teaching-reproducible-methods/ [Accessed 7 Oct 2021].

37 Neuroscience Research Australia. Research quality. Available: https://www.neura.edu.au/about/research-quality/ [Accessed 3 Oct 2021].

38 Reporting guidelines, 2021. Available: https://www.equator-network. org/ [Accessed 3 Oct 2021].

39 Percie du Sert N, Hurst V, Ahluwalia A, *et al*. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *BMJ Open Sci* 2020;4:e100115.

40 Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

41 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.

42 Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. *Int J Nurs Stud* 2015;52:5–9.

43 Hardwicke TE, Ioannidis JPA. Mapping the universe of registered reports. *Nat Hum Behav* 2018;2:793–6.

44 Nosek BA, Beck ED, Campbell L, *et al*. Preregistration is hard, and worthwhile. *Trends Cogn Sci* 2019;23:815–8.

45 Cashin AG, Richards GC, DeVito NJ. Registration of health and medical research. *BMJ Evid Based Med*;1:111836.

46 Laine C, Horton R, DeAngelis CD, *et al*. Clinical trial registration– looking back and moving ahead. *N Engl J Med* 2007;356:2734–6.

47 Chambers CD, Forstmann B, Pruszynski JA. Registered reports at the European Journal of neuroscience: consolidating and extending peer-reviewed study pre-registration. *Eur J Neurosci* 2017;45:627–8.

48 Baxter MG, Burwell RD. Promoting transparency and reproducibility in behavioral neuroscience: publishing replications, registered reports, and null results. *Behav Neurosci* 2017;131:275–6.

49 Nosek BA, Ebersole CR, DeHaven AC, *et al*. The preregistration revolution. *Proc Natl Acad Sci U S A* 2018;115:2600–6.

50 Mertens G, Krypotos A-M. Preregistration of analyses of preexisting data. *Psychol Belg* 2019;59:338–52.

51 Evans TR, Branney P, Clements A, *et al*. Improving evidence-based practice through preregistration of applied research: barriers and recommendations. *Account Res* 2021:1–21.

52 Paul M, Govaart GH, Schettino A. Making ERP research more transparent: guidelines for preregistration. *Int J Psychophysiol* 2021;164:52–63.

53 Kerr NL. HARKing: hypothesizing after the results are known. *Pers Soc Psychol Rev* 1998;2:196–217.

54 Dwan K, Gamble C, Williamson PR, *et al*. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8:e66844.

55 Héroux ME, Taylor JL, Gandevia SC. The use and abuse of transcranial magnetic stimulation to modulate corticospinal excitability in humans. *PLoS One* 2015;10:e0144151.

56 Begley CG, Ioannidis JPA. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res* 2015;116:116–26.

57 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.

58 Naudet F, Sakarovitch C, Janiaud P, *et al*. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. *BMJ* 2018;360:k400.

59 Packer M. Data sharing in medical research. *BMJ* 2018;360:k510.

60 Baker M. Why scientists must share their research code. *Nature*2016.

61 Irving D, Hertweck K, Johnston L. *Research software engineering with python: building software that makes research possible*. Boca Raton, FL: CRC Press, 2021.

62 McGuinness LA, Sheppard AL. A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts. *PLoS One* 2021;16:e0250887.

63 Button KS, Ioannidis JPA, Mokrysz C, *et al*. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365–76.

64 Calin-Jageman RJ, Cumming G. Estimation for better inference in neuroscience. *eNeuro* 2019;6:ENEURO.0205-19.2019.

65 Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.

66 MacCoun R, Perlmutter S. Blind analysis: hide results to seek the truth. *Nature* 2015;526:187–9.

67 Simera I, Moher D, Hirst A, *et al*. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR network. *BMC Med* 2010;8:24.

68 Percie du Sert N, Ahluwalia A, Alam S, *et al*. Reporting animal research: explanation and elaboration for the ARRIVE guidelines 2.0. *PLoS Biol* 2020;18:e3000411.

69 Héroux M. Inadequate reporting of statistical results. *J Neurophysiol* 2016;116:1536–7.

70 Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. *Adv Physiol Educ* 2004;28:85–7.

71 Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel. *Adv Physiol Educ* 2007;31:295–8.

72 Belia S, Fidler F, Williams J, *et al*. Researchers misunderstand confidence intervals and standard error bars. *Psychol Methods* 2005;10:389–96.

73 Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. *Nature* 2005;435:737–8.

74 Anderson WP. Reproducibility: stamp out shabby research conduct. *Nature* 2015;519:158.

75 Pernet C. Null hypothesis significance testing: a short tutorial. *F1000Res* 2015;4:621.

76 Curran-Everett D. CORP: minimizing the chances of false positives and false negatives. *J Appl Physiol* 2017;122:91–5.

77 Drummond GB, Tom BDM. Presenting data: can you follow a recipe? *J Physiol* 2011;589:5007–11.

78 Kruschke JK, Liddell TM. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev* 2018;25:178–206.

79 Kruschke JK. Bayesian analysis reporting guidelines. *Nat Hum Behav* 2021;5:1282–91.

80 Boutron I, Dutton S, Ravaud P, *et al*. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.

81 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.

82 Harris JK, Combs TB, Johnson KJ, *et al*. Three changes public health scientists can make to help build a culture of reproducible research. *Public Health Rep* 2019;134:109–11.

83 Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666–76.

84 Ma L-L, Wang Y-Y, Yang Z-H, *et al*. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res* 2020;7:7.

85 Mansournia MA, Collins GS, Nielsen RO, *et al*. Checklist for statistical assessment of medical papers: the CHAMP statement. *Br J Sports Med* 2021;55:1002–3.

86 Study quality assessment tools. Available: https: //www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [Accessed 8 Oct 2021].

87 Fidler F, Thomason N, Cumming G. Editors can lead researchers to confidence intervals, but can't make them think. *Psychol Sci* 2004;15:119–26.

88 Altman DG, Moher D. Declaration of transparency for each research article. *BMJ* 2013;347:f4796.

89 Barnett AG, Moher D. Turning the tables: a university league-table based on quality not quantity. *F1000Res* 2019;8:583.