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.

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, 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
was the formation of a Research Quality Committee in 2018\textsuperscript{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\textsuperscript{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 Communications 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,\textsuperscript{39} CONSORT,\textsuperscript{40} PRISMA\textsuperscript{41} adhered to?’) or dwell 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.\textsuperscript{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.\textsuperscript{6} 43–45 It is mandatory for clinical trials\textsuperscript{46} and now encouraged across all research types.\textsuperscript{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.\textsuperscript{53} It also protects against cherry picking—the selective publication of findings.\textsuperscript{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.\textsuperscript{27,56–58} This is also essential to enable
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 (e.g., 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.\textsuperscript{65,64} 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.\textsuperscript{56} 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.\textsuperscript{67} There are a growing number and range of reporting guidelines, many of which are collected under the EQUATOR network.\textsuperscript{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,\textsuperscript{21,69} the SEM is rarely the appropriate summary statistic\textsuperscript{70,71} and is often misinterpreted.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{71} Moreover, there are calls for more conservative thresholds to be used to reduce the chance of false-positive findings.\textsuperscript{76} To properly interpret statistical results, exact p values should be reported.\textsuperscript{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.\textsuperscript{13,17,21,24} Researchers are increasingly encouraged to focus on CIs (or credible intervals) when they report and interpret their statistical results\textsuperscript{74,64,78} with some moving towards Bayesian statistics. While these newer reporting practices and statistical approaches may also be improperly reported,\textsuperscript{29} 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.\textsuperscript{23,80} This questionable research practice occurs frequently across the biomedical sciences.\textsuperscript{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.\textsuperscript{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
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 (e.g., 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 for more efficient data entry because the output from these tools allows for more efficient and relevant reporting.

The QuOCCA is an assessment tool. It was not designed to provide a replacement for reporting guidelines that assess statistical methods or particular study designs, 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 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. But enforcement tends to work only as long as it is in place. 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. 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

Twitter Aidan G Cashin @AidanCashin, Euan J McCaughey @EuanMcCaughey and Matthew Jones @MattJones0203

Acknowledgements We would like to acknowledge Bronwyn Chapman for her contribution to the Research Quality Committee, and for managing the assessment of institute papers. We would also like to acknowledge Professor Rob Herbert for thoughtful discussions as the QuOCCA was being developed, and Professor Stephen Lord and Prof. Kim Delbaere for their early contributions to the QuOCCA project.

Contributors MEH: conceptualisation, design of checklist, auditing manuscripts, interpretation of results, drafting and revising the manuscript. AJA: conceptualisation, design of checklist, auditing of manuscripts, interpretation of results, revising the manuscript. AC: conceptualisation, design of checklist, auditing of manuscripts, online implementation and data management, interpretation of results, revising the manuscript. SGD: conceptualisation, design of checklist, auditing of manuscripts, interpretation of results, revising the manuscript. MAJ: auditing of manuscripts, interpretation of results, revising the manuscript. All authors approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

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

REFERENCES


