Author’s response to reviews

Title: Quality of reports of investigations of research integrity by academic institutions

Authors:

Andrew Grey (a.grey@auckland.ac.nz)
Mark Bolland (m.bolland@auckland.ac.nz)
Greg Gamble (gd.gamble@auckland.ac.nz)
Alison Avenell (a.avenell@abdn.ac.uk)

Version: 1 Date: 31 Jan 2019

Author’s response to reviews:

Professor Maria Kowalczuk
Co-Editor-in-Chief
Research Integrity and Peer Review

1 February 2019

Dear Professor Kowalczuk,

Re: RIPR-D-18-00028 Quality of reports of investigations of research integrity by academic institutions

Thank you for the opportunity to submit a revised manuscript. We thank the reviewers for their comments, and have responded to each point below. In the revised manuscript, alterations are in bold type.

We hope the revised manuscript will be acceptable for publication

Yours sincerely,

Andrew Grey MD

Department of Medicine
University of Auckland
Reviewer #1: Grey and colleagues examined the investigation reports from 3 institutions on the integrity of a large number of publications co-authored by researchers from these institutions (integrity concerns on which were raised by the current authors themselves). Ticking against a quality and adequacy checklist published earlier by Gunsalus et al. (2018), the authors find the reports inadequate. Overall, the study could be of potential interest but there are issues that need to be addressed.

Technical issues:

1. It was stated by the authors that "...Overall, each report was considered unacceptable". The basis for this conclusion, although not unintuitive, is still not particularly clear. Meeting all 26 items in Gunsalus et al.'s checklist is (arguably) a tall order, so what degree of compliance would be considered "acceptable"? Furthermore, would all items in that checklist have equal weightage? It would be helpful if the authors have, or could cite a publicly available 'positive control', ie., an exemplary investigation report which would be considered acceptable.

Response: The checklist includes a summary evaluation with 4 categories - 'acceptable as is', 'minor revisions needed', 'major revisions needed' and 'report is not acceptable'. No definition of these categories is included with the checklist. Each assessor, however, independently scored each report as unacceptable, on the basis of only 0-3 of 26 individual checklist items in each report being satisfactorily addressed. To clarify this, we have added the following text to the Methods section:

‘The checklist includes an overall assessment, with 4 categories that include ‘acceptable as is’, ‘minor revisions needed’, ‘major revisions needed’ and ‘report is not acceptable’.

In the Results text we have altered the wording to read:

‘Overall, each report was considered unacceptable by both assessors. In the respective reports, only 3/26, 1/26 and 0/26 individual checklist items were addressed adequately: a further 5/26, 5/26 and 4/26 items could not be assessed. Common deficiencies were widespread, and occurred in each domain of the checklist, including the presentation of concerns/allegations…’

We are not aware of a publicly available scoring of an exemplary institutional report, but one of the authors of the checklist used it to assess an institutional report on the Retraction Watch website https://retractionwatch.com/wp-content/uploads/2018/06/Katiyar-checklist2.pdf. In that assessment, 19 of 26 items were addressed either adequately or in part, and the reviewer considered the report needed minor revision.

In the Discussion, we added the following text, and reference to the assessment on the Retraction Watch website:
‘They contrast markedly with a higher quality report of an institutional investigation of a different case, assessed using the same checklist [7].’

2. The case raise by the authors appears to be of a rather elaborate scale, with >200 paper involving multiple authors spread across 4 institutions. Several high profile cases of misconduct where reports are publically posted and which reports were more adequate in the terms specified by the checklist involved far less number of papers and perpetrators. Thus, in making inferences on "deficiencies in the quality and reporting of institutional investigation of concerns about research integrity", the authors should take the complex and elaborate nature of their case into account.

Response: We acknowledge that this case is complicated, both in scope and multiplicity of institutions and countries. Nonetheless, such cases are important, and deserve, in our opinion, careful, complete and coordinated investigations. Our analysis suggests that this did not happen.

We added the following text to the Discussion section:

‘The considerable scope of the concerns, and the involvement of more than 1 institution and more than 1 country might have influenced the quality of the investigations undertaken.’

3. The authors noted that "Institutions 1, 2 and 4 investigated 38/49, 7/200 and 5/34 potentially affected papers, respectively". Was their conclusion of "Reports failed to address publication-specific concerns and focussed more strongly on determining research misconduct than evaluating the integrity of publications" based on these numbers? This point could perhaps be better clarified in 'Discussion'.

Response: The investigations both failed to assess all potentially affected publications AND failed to address detailed concerns raised about individual publications. The latter finding was pivotal to our observation that determination of (narrowly defined) research misconduct took priority over determining the integrity of individual publications. We have clarified this point by adding the following text to the Methods section:

‘Specific and detailed concerns were raised about potentially affected publications co-authored by employees at institutions 2, 3 and 4. These concerns were wide-ranging, including those about research governance, authorship misconduct, implausible productivity and study conduct, duplicate results reporting, impossible or implausible data, and data errors (Table 1).’

And providing additional detail in the Results section of institution 1’s focus on misconduct:

‘For example, the investigation committee of institution 1 reserved judgment on 24/38 papers because it ‘did not obtain clear evidence of fraud for these publications.’ When we subsequently
communicated specific concerns about a wide range of aspects of the integrity of 18 of these 24 papers (concerns listed in Table 1), institution 1 initiated further investigation.

And the following text to the Discussion section:

‘Our assessment of 2 of the 3 available investigation reports suggested a much stronger focus on the determination of misconduct, restricted in definition to fabrication, falsification or plagiarism, than the evaluation of research integrity; the third report was too vague to assess in this regard. This observation is supported by the failure of the investigations to address specific concerns about the integrity of individual publications which included those about research governance, authorship misconduct, implausible productivity and study conduct, duplicate results reporting, impossible or implausible data, and data errors. We think…’

4. Further to the above, is it known as to why the other papers highlighted were not investigated or ignored (or are these simply not included in the reports for brevity)? This is an important point because from either an institutional or journal/publisher perspective, ensuring the integrity of the published literature would take priority. Logically, most if not all investigations would start off by a careful scrutiny of the papers concerned.

Response: We do not know why the other potentially affected papers were not addressed in these reports. We agree that the integrity of the literature should be the main priority of any investigation, but, as indicated above, do not think that such was the case in these investigations. We think paragraphs 3 and 4 in the revised Discussion cover this uncertainty and the subjugation of research integrity at the expense of narrowly defined misconduct. We amended paragraph 3 to further clarify this:

‘We found that institutional investigations failed to address all potentially affected papers and to address detailed concerns about individual publications. The latter observation is similar to that in an analysis of journal and publisher responses to such concerns over research published by the same group of researchers (data in submission). In that analysis, none of the 16 decision emails from affected journals responded to all of the specific concerns that had been raised about the relevant publication. We cannot determine the reason for these findings. It could be that the institutional investigations decided to focus on a subset of potentially affected papers, did not…’

5. In concluding, the authors suggested that "A possible solution is the establishment of independent organisations with authority to undertake and report investigations..." This is already in place in some countries and the authors should refer to these and verify if these do indeed produce more 'adequate' investigation reports on record.

Response: We acknowledge the existence in some countries of organisations that at least nominally investigate concerns about research integrity. We are not aware of any data that address whether such organisations produce or coordinate better quality investigations and reports than academic institutions alone. In the current situation, neither of the national
organisations, ORI and MEXT, was helpful in overseeing or conducting investigations. In addition, neither has jurisdiction outside of its own country. We have added the following text to the Results section;

‘The ORI did not acknowledge receipt of emails sent in October and November 2017 outlining concerns about research conducted at the US institution. Our emails to the MEXT in November and December 2017 reporting the concerns about research conducted at the Japanese institutions, including ones written by a Japanese colleague, either failed to elicit a response or generated brief unhelpful replies, promising a response that has not yet materialised.’

And expanded the final paragraph of the Discussion as follows:

‘A possible solution is the establishment of more efficient and adequately resourced independent organisations with the authority and expertise to undertake and report investigations, and implement recommendations, including those which span multiple institutions and countries.’

6. Other issues:

The process of research reported in this manuscript is rather unusual in that the authors had raised concerns on research integrity with the institutions in the first place, then obtaining and analysing their (presumably confidential) reports for adequacy. The authors have indicated that ethics approval for the work is 'Not applicable'. However, in using these other academics entities as means to a research end, I wonder whether a more neutral third party (such as the authors' IRB) should have been consulted on its ethical implications.

Response: We do not think ethical committee review is required for this work because it is an audit of the established practice of reporting and investigating concerns about research integrity. Ethics committee review is not required for audit in New Zealand.

Reviewer #2: A strong contribution to the literature. There are places where it is a bit confusing which institutional response is being discussed, and it took me time to disentangle the saga. Important findings include the lack of correction to the literature and the failure to assess the majority of publications about which concerns were raised. This pattern has been seen elsewhere and is critical to address. That two of the institutions started additional investigations after you raised questions is both positive and disheartening that this step was necessary.

Response: Thank you for the positive comments

It would be helpful for you to address a bit more clearly what elements of integrity of the literature you believe were not addressed because of the perceived focus on research misconduct. It can be partially inferred from the tables, and clarifying that point would make this a stronger paper. The conclusion seems entirely plausible to me, and isn't fully explained.
Response: Please see response to reviewer 1, point 3. We think the additions clarify this aspect of the paper.

I would like to see an expanded discussion of this work and more detail, as well as places that you believe the checklist items are ambiguous and could be clarified. Perhaps a user's guide would be a useful contribution to develop?

Response: We think the revised version provides some important additional discussion and detail. We would be happy to address any further specific suggestions the Reviewer has. We think it beyond the scope of the current work to discuss specific aspects of the checklist, as we did not set out to formally assess its utility. We are aware that the reviewer is one of the authors of the checklist, and offer the following comments:

We think the checklist is a very useful tool for assessing the quality of investigations of research integrity. It was straightforward to apply, and we found interassessor agreement to be good/acceptable. We added the kappa coefficients for interassessor agreement to the Results section. When assessors disagreed, the disagreements were minor, typically one assessor scoring an item either “cannot assess” or “in part” and the other “yes” or “no”.

Ambiguity that we encountered was concentrated mostly in the Evidence section of the checklist – for example, in response to the question “Did the report indicate if evidence was properly sequestered and protected from tampering?” one assessor checked “cannot assess” and the other checked “no”, for a report in which no discussion of this issue was present. Both responses are reasonable, but our consensus was that the best assessment is “no”. For that question, consideration could be given to removing the “cannot assess” option. A similar amendment could be applied to the question “Is there a list of individuals who were interviewed?” - neither “in part” nor “cannot assess” are valid responses.

We would be happy to discuss the checklist further if the developers were interested. There are also potential research projects that could be undertaken evaluating and using the tool.