PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Comprehensive Researcher Achievement Model (CRAM): a framework for measuring researcher achievement, impact and influence derived from a systematic literature review of metrics and models
AUTHORS	Braithwaite, Jeffrey; Herkes, Jessica; Churruca, Kate; Long, Janet; Pomare, Chiara; Boyling, Claire; Bierbaum, Mia; Clay-Williams, Robyn; Rapport, Frances; Shih, Patti; Hogden, Anne; Ellis, Louise A.; Ludlow, Kristiana; Austin, Elizabeth; Seah, Rebecca; McPherson, Elise; Hibbert, Peter; Westbrook, Johanna

VERSION 1 - REVIEW

REVIEWER	Simon Deeming Hunter Medical Research Institute, Australia
REVIEW RETURNED	23-Aug-2018

GENERAL COMMENTS	Thank you for the opportunity to review the draft manuscript. How to appropriately assess academic performance remains a challenging issue, particularly as the demands upon academics are stretched further across teaching, academic research and wider impacts. The following feedback is provided constructively to hopefully assist the authors. - While the criteria upon which to assess academic performance for promotion represents a general, if unspecified, objective, the text does not specify a specific aim or aims for the research. The aim is probably to 'identify approaches to assessing researcher's achievements published in the academic literature', but it could be more clearly stated as an explicit aim. - The review (not the model and recommendations) focuses upon 'academic' /bibliometric impact. No definition for impact is provided, but the results demonstrate that the focus lies with bibliometrics. The ARC's alternative definition for impact is not addressed. Wider impacts are occasionally raised (L237, L421), but the text does not address this consideration, nor define it as out of scope. - The paper could be improved by defining impact and retaining this clarity throughout. This probably necessitates not calling it 'impact', but 'academic impact' or 'bibliometric impact'. Furthermore, the search strategy addresses grant application, funding body and funding system, but does comprehensively address funding success as a measure of 'academic impact'. Consequently, the value of the review lies more explicitly with 'academic impact' as measured by publication impact. - This general point also applies to the title, which alludes to 'influence' meaning academic influence in this instance, as well as

	'achievement' and 'impact', which could have greatly different interpretations to those focused upon teaching or clinical/policy change/commercialisation. - The review is well-conducted and very thorough. Consequently, the review's insights do contribute to the evidence base. This stated, the review found a further 17 reviews of these issues, which naturally limits the novelty of many of the results. There is an extensive existing literature base in this subject. - The results do pull together some insights that may not have been placed in the wider bibliometric context to date e.g. correlation between the h-index and less transparent bibliometric measures. - L390 - Incomplete sentence - The Comprehensive Researcher Achievement Model (CRAM) - The model is informed by the main conclusion of the review, being that all assessment methods for 'academic publication impact' are imperfect and therefore performance should be considered in a broader light. Fine. However, the aim of the review was to conduct a systematic literature review of 'academic impact' (as defined by this reviewer), as opposed to designing a model for academic promotion. The latter would potentially entail any entirely different search strategy and encompass the purpose of academia, the range of potential objectives, the purpose of assessment models and the range of existing institution-based methods to conduct such assessment. Consequently, in light of the rigour of the review, the basis for the CRAM appears weak, particularly given the inclusion of a broad range of assessors that were neither asked their view, nor to which the review has been targeted. - There is nothing wrong with the focus of 'academic publication impact' for the sake of academic promotion, but the CRAM fails to consider the other potential impacts with the same rigour. Consequently, the model reads as an over reach, given the evidence base compiled from the method. For example, patents are identified in the text as a simplistic measure (which they are), but they are then presented as an 'Assessment component' in the CRAM. - Unfortunately, I think the apparent lack of thought underpinning the CRAM undermines the thorough work undertaken in the review. Research impact and research translation are complicated subject areas for which there is a paucity of 'real' evidence, do not lend themselves to easy technical solutions and merit further research. The authors should be commended for their contribution to this thinking. However, this field is also prone to stretching research toward broader conclusions and relevance, beyond that substantiated by the evidence. Unfortunately, this paper appears to fall into this trap. A more discrete/defined presentation of the aim, method, results and implications for thinking regarding 'academic promotion criteria' would, in the view of this reviewer, improve the quality of the paper. Thank you for the opportunity to review. I hope these comments are taken in the constructive light in which they are presented. Kind regards.

REVIEWER	Ali Azeez Al-Jumaili, PhD The University of Iowa College of Pharmacy,IA, USA and University of Baghdad College of Pharmacy, Baghdad, Iraq
REVIEW RETURNED	12-Sep-2018

GENERAL COMMENTS	Thanks for submitting this comprehensive review. Please answer/address the following comments: 1. Page 8, line 147 (inclusion criteria): Did you include the measures of researcher publications and achievements in medical fields only or in all science disciplines? 2. Page 8, line 152: Did you conduct meta-analysis or just followed the meta-analysis protocols? If yes, where is the table containing the meta-analysis results? 3. Page 8, line 166: In general, non-empirical articles can be bias since they do not rely on quantitative findings. So, what is the pros from including non-empirical studies in your review? Given that the review included 166 non-empirical studies and only 19 of them declared conflict of interest, from my perspective including such study will compromise the findings of your review. 4. Page 7, it is unclear whether you included qualitative study in your review? If yes, which methodology you used to extract the findings? If no, include this in your exclusion criteria. 5. Page 10, line 285, the authors chose h-index (google scholar) as one parameter. Why did not consider Researchgate scores (which also based on number of publications and citations) for the same reason? 6. Page 12, line 263: CiteScore may be more popular than JIF to measure the quality of journals. Why did you choose JIF instead? 7. Page 14, line 332, altmetric: how can you track downloading of articles through social media given that there are so many of them such as Facebook, twitter, LinkedIn ------etc? 8. Page 32, histogram figure: Could you add caption about what do you mean by positive and negative discussion? 9. Did you run any statistical analysis including the reviewed study findings/data? If yes, where are your analysis findings? 10. Page 15, line 441: Can you suggest a certain percentage for each domain (contribution of each component) of the CRAM model to facilitate adopting by users? Thanks

REVIEWER	Chengzhi Zhang Nanjing University of Science and Technology, China
REVIEW RETURNED	26-Sep-2018

GENERAL COMMENTS	What is the practical value of this study? The model lacks evidence here.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Simon Deeming	
Thank you for the opportunity to review the draft manuscript. How to appropriately assess academic performance remains a challenging issue, particularly as the demands upon academics are stretched further across teaching, academic research and wider impacts. The following feedback is provided constructively to hopefully assist the authors.	Much appreciated.
- While the criteria upon which to assess academic performance for promotion represents a general, if unspecified, objective, the text does not specify a specific aim or aims for the research. The aim is probably to 'identify approaches to assessing researcher's achievements published in the academic literature', but it could be more clearly stated as an explicit aim.	We have expressed a clear aim ("We aimed to identify what is known about methods for assessing researcher achievements, drawing on this to propose a new composite assessment model"), and have revamped the abstract in accordance with the Editor's request. We have also made our aim more explicit at the end of the Introduction.
- The review (not the model and recommendations) focuses upon 'academic' /bibliometric impact. No definition for impact is provided, but the results demonstrate that the focus lies with bibliometrics. The ARC's alternative definition for impact is not addressed. Wider impacts are occasionally raised (L237, L421), but the text does not address this consideration, nor define it as out of scope.	This is a useful point. We have defined impact broadly in our introduction (as the "outcomes of research"). Overwhelmingly, this is academic or bibliometric impact, rather than real-world or practical impact, because that's what the literature we synthesised discusses to a greater extent; we did a systematic review and went where the literature went. Where possible in the Discussion we have teased out where we think our review leads in terms of broader impact and distinguished academic or citation-based impact from real-world or practical impact: lines 414 onwards. The ARC's definition is not really relevant to an international paper in BMJ Open. Many countries are pursuing ideas about impact. (Incidentally, we are doing a literature review here, so there was no space

	to do a policy analysis or to factor in the ARC's position on impact. It's a great idea for future work and we might well look at this down the track.)
- The paper could be improved by defining impact and retaining this clarity throughout. This probably necessitates not calling it 'impact', but 'academic impact' or 'bibliometric impact'. Furthermore, the search strategy addresses grant application, funding body and funding system, but does comprehensively address funding success as a measure of 'academic impact'. Consequently, the value of the review lies more explicitly with 'academic impact' as measured by publication impact.	The systematic review aimed to identify the methods applied to examine researcher achievements. Thus the definition of impact was emergent and reflects the tools being used. Our search strategy provides the broad parameters regarding the elements considered within the definition of achievement/impact. We have increased the use of the term academic impact acknowledging this was a key focus for much of the literature. In line 95 we call what you are seeing as real-world impact "influence".
- This general point also applies to the title, which alludes to 'influence' meaning academic influence in this instance, as well as 'achievement' and 'impact', which could have greatly different interpretations to those focused upon teaching or clinical/policy change/commercialisation.	This is a tough area to study and review. See our response above. Our results show the ways in which these aspects of impact are reflected in the literature searched.
- The review is well-conducted and very thorough. Consequently, the review's insights do contribute to the evidence base. This stated, the review found a further 17 reviews of these issues, which naturally limits the novelty of many of the results. There is an extensive existing literature base in this subject.	There is indeed an extensive literature. The value here, however, is that we have, through our comprehensive (and painstaking) review brought a lot of things together in a new model. This is the benefit of having a very large research team do a wide-ranging review. We do believe we have added value. Only a few of the reviews (i.e., Caminiti C, Iezzi E, Ghetti C, De'Angelis G, Ferrari C. A method for measuring individual research productivity in hospitals: development and feasibility. BMC Health Services Research. 2015;15; Wildgaard L, Schneider JW, Larsen B. A review of the characteristics of 108 author-level bibliometric indicators. Scientometrics. 2014;101(1):125-158; and Patel VM, Ashrafian H, Ahmed K, et al. How has healthcare research performance been assessed? A systematic review.

	Journal of the Royal Society of Medicine. 2011;104(6):251-261.) did something a little similar to us—and they only focused on bibliometrics or health care research (respectively) whereas we reviewed diverse ways of assessing researcher achievement (including altmetrics, past funding) across fields. Other reviews were non-systematic or looked at only one bibliometric (e.g., reviewed the h-index).
- The results do pull together some insights that may not have been placed in the wider bibliometric context to date e.g. correlation between the h-index and less transparent bibliometric measures.	We appreciate this note.
- L390 - Incomplete sentence	Thanks indeed for your close reading of the manuscript. We have attended to this.
- The Comprehensive Researcher Achievement Model (CRAM) - The model is informed by the main conclusion of the review, being that all assessment methods for 'academic publication impact' are imperfect and therefore performance should be considered in a broader light. Fine. However, the aim of the review was to conduct a systematic literature review of 'academic impact' (as defined by this reviewer), as opposed to designing a model for academic promotion. The latter would potentially entail any entirely different search strategy and encompass the purpose of academia, the range of potential objectives, the purpose of assessment models and the range of existing institution-based methods to conduct such assessment. Consequently, in light of the rigour of the review, the basis for the CRAM appears weak, particularly given the inclusion of a broad range of assessors that were neither asked their view, nor to which the review has been targeted.	We hear the point, but respectfully demur from it. The model arises from work reviewing a very large body of literature involving 18 researchers on the team. We looked widely and had broad inclusion criteria. As we state in our aim, the goal was to take existing literature and draw on this to present a model. In that search we identified some good metrics, but general cautions in using a single one to evaluate a researcher… hence our model. We did not set out to undertake primary data collection to investigate the issues you suggest. That is in fact a different research task. We agree that there may be value in this as a future piece of work.

- There is nothing wrong with the focus of 'academic publication impact' for the sake of academic promotion, but the CRAM fails to consider the other potential impacts with the same rigour. Consequently, the model reads as an over reach, given the evidence base compiled from the method. For example, patents are identified in the text as a simplistic measure (which they are), but they are then presented as an 'Assessment component' in the CRAM.	We've discussed this amongst the team, and understand the point, but just because we didn't review every aspect of the literature, e.g., on patents, doesn't invalidate its inclusion in the CRAM. We reviewed 478 articles – bigger than many other reviews we know. Multiple articles we reviewed mentioned patents – so we felt quite firm in including this as an assessment component without the need to do another review on patents. Individually, the literature suggests all measures of achievement are somewhat simplistic. CRAM can be used quantitatively to calculate academic impact and qualitatively to enrich a peer-review examination of broader impact. It provides a basis for calculations of achievement based on bibliometric and altmetric data. It also makes explicit the components of achievement to be assessed, thereby informing and enabling a more transparent and rigorous peer-review process that considers the less easily quantified and measured components of impact such as knowledge translation and change in practice (see para beginning line 414).
- Unfortunately, I think the apparent lack of thought underpinning the CRAM undermines the thorough work undertaken in the review.	The model arises from the review we did of the literature which was our stated aim, and we have discussed the strengths and limitation of various measures. In the conclusion we acknowledge that this is a generic model and not the ultimate end point. It reflects our key finding that single metrics are not sufficient, but designs need to be composite models and we draw together those which have been found to have some value in the literature. In this way we believe the work presents a useful contribution.

Research impact and research translation are complicated subject areas for which there is a paucity of 'real' evidence, do not lend themselves to easy technical solutions and merit further research. The authors should be commended for their contribution to this thinking. However, this field is also prone to stretching research toward broader conclusions and relevance, beyond that substantiated by the evidence. Unfortunately, this paper appears to fall into this trap. A more discrete/defined presentation of the aim, method, results and implications for thinking regarding 'academic promotion criteria' would, in the view of this reviewer, improve the quality of the paper.	We agree that these are complicated areas and are pleased with your recognition of our efforts. We don't see however how the outcome—the model—is a stretch from the literature reviewed. As to our point above, the model reflects our key finding, ie designs need to be composite models. We take the point about the defined aim and purpose of the study and the value of the model, and have defined the scope more accurately to make this clearer.
Reviewer 2: Ali Azeez Al-Jumaili	
Thanks for submitting this comprehensive review. Please answer/address the following comments:	Thanks indeed for your review of our paper.
1. Page 8, line 147 (inclusion criteria): Did you include the measures of researcher publications and achievements in medical fields only or in all science disciplines?	Many of the articles we reviewed were broadly in the area of health and medical research, and our discussion is concerned with the implications for health and medical research because that is where our interest, as well as the readership of BMJ Open, lie. However, we set no inclusion criteria with regard to scientific discipline, because novel and useful approaches to assessing research achievement might come from diverse fields. Indeed, the papers we reviewed also came from social science disciplines, as well as more specific areas like conservation, astrophysics, engineering, and business. Overwhelmingly, these papers discussed metrics and methods that have general applicability for assessing researcher achievement, indicating our inclusion of them was well-justified.
2. Page 8, line 152: Did you conduct meta-analysis or just followed the meta-analysis protocols? If yes, where is the table containing the meta-analysis results?	We did not conduct a meta-analysis, as the results of our systematic review were too heterogenous for meta-analysis; but the protocol document for systematic

	reviews and meta-analyses are grouped in the one document (for teams that are conducting both simultaneously). Hence, we followed the systematic review protocol and disregarded the components suited only to meta-analysis.
3. Page 8, line 166: In general, non-empirical articles can be bias since they do not rely on quantitative findings. So, what is the pros from including non-empirical studies in your review? Given that the review included 166 non-empirical studies and only 19 of them declared conflict of interest, from my perspective including such study will compromise the findings of your review.	It is true that non-empirical studies are not free from bias, but we did not see this as a reason to exclude the articles. Arguably those proposing a new metric in a study are more invested in positioning it in a positive light, than editorials and commentaries, which are often more critical. Furthermore, the editorials and commentaries we reviewed better captured trends in the broader thinking around research achievement in the academic community (often comparing and contrasting, such as the move away from JIF) than an empirical study evaluating a single metric or model.
4. Page 7, it is unclear whether you included qualitative study in your review? If yes, which methodology you used to extract the findings? If no, include this in your exclusion criteria.	Yes, qualitative studies were included in this review. As described in our Method, we developed a custom data extraction sheet to encompass the diverse and heterogeneous articles included. It focused on documenting the model(s)/metric(s) described in the paper, and their reported strengths and limitations.
5. Page 10, line 285, the authors chose h-index (google scholar) as one parameter. Why did not consider Researchgate scores (which also based on number of publications and citations) for the same reason?	ResearchGate Score is a form of altmetrics based on how outputs are received by the peer network. We have added in a section in the Results mentioning ResearchGate.

6. Page 12, line 263: CiteScore may be more popular than JIF to measure the quality of journals. Why did you choose JIF instead?	The reviewer makes an excellent point about the value of CiteScore; however, our aim for this review was to examine models and metrics for assessing an individual researcher's achievement. While JIF is a journal-level measure, we focused on it in our review because, in the literature we examined, it was discussed as a metric used—rightly or wrongly—to assess individual researchers.
7. Page 14, line 332, altmetric: how can you track downloading of articles through social media given that there are so many of them such as Facebook, twitter, LinkedIn ------etc?	This is an interesting and complicated question. Unfortunately, the mechanics of the algorithms behind these metrics are beyond the scope of the article. Furthermore, one of the criticisms of altmetrics we identified in our review was the lack of standardisation for their calculation. The tracking of for example downloads or webpage views (i.e., the data used as the input for these algorithms) can be through academic databases (e.g., Scopus displays altmetrics) or devoted websites (e.g., https://www.altmetric.com/).
8. Page 32, histogram figure: Could you add caption about what do you mean by positive and negative discussion?	This is a good suggestion, and we have added a caption accordingly to the document. For information in this study, "Positive discussion" refers to articles that discuss the metric in a favourable light, or focus on the strengths of the metric. "Negative discussion" refers to articles that focus on the limitations or shortcomings of the metric.
9. Did you run any statistical analysis including the reviewed study findings/data? If yes, where are your analysis findings?	As noted above, the range of articles included was diverse and heterogenous (e.g., empirical and non-empirical), which precluded a conventional meta-analysis. In the article we report on some basic descriptive statistics to describe the body of included papers (see beginning of Results). We also examined

	the sentiment with which the most commonly reported on metrics were discussed in the included literature. With regard to the assessment of individual researcher achievement, we found that the JIF was discussed most negatively, and altmetrics were discussed most favourably (see Figure 2).
10. Page 15, line 441: Can you suggest a certain percentage for each domain (contribution of each component) of the CRAM model to facilitate adopting by users?	The CRAM is intended to provide a comprehensive picture of researcher achievement and make explicit the types of qualities and components being assessed (e.g., impact, productivity). The percentages assigned to components though are something likely to change with funding bodies and grant schemes, or assessors and the purposes for which they are making an assessment. For example, some grants may require more translation focus, in which case a greater apportioning of percentage to impact an influence might be warranted. We expect that CRAM will improve the transparency of what is being assessed and how, and provide the basis for peer-assessment, supplemented by consideration of less easily measured achievements (e.g., real-world impacts rather than academic). This is mentioned in lines 427-428.
Reviewer 3: Chengzhi Zhang	
What is the practical value of this study? The model lacks evidence here.	Thank you for your review and this question. The practical value lies in us deriving a model from a very extensive systematic review of the literature to make available for future assessment of researchers for consideration of tenure, promotion and research funding.

10.1177/20542704251325320 on 30 March 2025

VERSION 2 – REVIEW

REVIEWER	Ali Al-Jumaili, PhD The University of Iowa, Iowa, USA
REVIEW RETURNED	24-Nov-2018

GENERAL COMMENTS	The authors responded well to my comments; however, I recommend adding (including) the author responses to my 1st and 4th comments to the discussion or method sections. I accept this manuscript for publication after the revisions.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2- Ali Al-Jumaili, PhD	
The authors responded well to my comments; however, I recommend adding (including) the author responses to my 1st and 4th comments to the discussion or method sections.	We have dealt with these. The first comment relates to limitations, so we have addressed this (lines 464-475). The second comment relates to method and is addressed there (line 157). Thank you for the suggestions.
I accept this manuscript for publication after the revisions.	Thank you for your continued support.