

# BMJ Open Are methodological quality and completeness of reporting associated with citation-based measures of publication impact? A secondary analysis of a systematic review of dementia biomarker studies

Shona Mackinnon,<sup>1</sup> Bogna A Drozdowska,<sup>2</sup> Michael Hamilton,<sup>3</sup> Anna H Noel-Storr,<sup>4</sup> Rupert McShane,<sup>4</sup> Terry Quinn<sup>2</sup>

**To cite:** Mackinnon S, Drozdowska BA, Hamilton M, *et al*. Are methodological quality and completeness of reporting associated with citation-based measures of publication impact? A secondary analysis of a systematic review of dementia biomarker studies. *BMJ Open* 2018;**8**:e020331. doi:10.1136/bmjopen-2017-020331

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020331>).

Received 31 October 2017  
Revised 23 January 2018  
Accepted 23 February 2018



<sup>1</sup>School of Medicine, University of Glasgow, Glasgow, UK  
<sup>2</sup>Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK  
<sup>3</sup>University of Strathclyde, Glasgow, UK  
<sup>4</sup>Cochrane Dementia and Cognitive Improvement Group, Oxford, UK

## Correspondence to

Bogna A Drozdowska;  
b.drozdowska.1@research.gla.ac.uk

## ABSTRACT

**Objective** To determine whether methodological and reporting quality are associated with surrogate measures of publication impact in the field of dementia biomarker studies.

**Methods** We assessed dementia biomarker studies included in a previous systematic review in terms of methodological and reporting quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) and Standards for Reporting of Diagnostic Accuracy (STARD), respectively. We extracted additional study and journal-related data from each publication to account for factors shown to be associated with impact in previous research. We explored associations between potential determinants and measures of publication impact in univariable and stepwise multivariable linear regression analyses.

**Outcome measures** We aimed to collect data on four measures of publication impact: two traditional measures—average number of citations per year and 5-year impact factor of the publishing journal and two alternative measures—the Altmetric Attention Score and counts of electronic downloads.

**Results** The systematic review included 142 studies. Due to limited data, Altmetric Attention Scores and electronic downloads were excluded from the analysis, leaving traditional metrics as the only analysed outcome measures. We found no relationship between QUADAS and traditional metrics. Citation rates were independently associated with 5-year journal impact factor ( $\beta=0.42$ ;  $p<0.001$ ), journal subject area ( $\beta=0.39$ ;  $p<0.001$ ), number of years since publication ( $\beta=-0.29$ ;  $p<0.001$ ) and STARD ( $\beta=0.13$ ;  $p<0.05$ ). Independent determinants of 5-year journal impact factor were citation rates ( $\beta=0.45$ ;  $p<0.001$ ), statement on conflict of interest ( $\beta=0.22$ ;  $p<0.01$ ) and baseline sample size ( $\beta=0.15$ ;  $p<0.05$ ).

**Conclusions** Citation rates and 5-year journal impact factor appear to measure different dimensions of impact. Citation rates were weakly associated with completeness of reporting, while neither traditional metric was related to methodological rigour. Our results suggest that high publication usage and journal outlet is not a guarantee of

## Strengths and limitations of this study

- Studies included were identified through a comprehensive, systematic search.
- A range of different potential determinants of publication impact was considered for analysis, selected based on findings from previous research.
- Due to limited data, alternative metrics could not be used as an outcome measure.
- An analysis of the context in which a paper was cited was not conducted.

quality and readers should critically appraise all papers regardless of presumed impact.

## INTRODUCTION

Arguably, the greatest recent advances and greatest controversies in dementia management have been around diagnosis. A particular area that has generated excitement in both lay and scientific press is the use of dementia biomarkers. A biomarker has been defined as ‘a characteristic that can be objectively measured and evaluated as an indicator of normal biological or pathogenic processes or pharmacological responses to a therapeutic intervention’.<sup>1</sup>

In theory, biomarker results from midlife can predict later life dementia. Putative dementia biomarkers, including proteins in cerebrospinal fluid (CSF) and neuroimaging techniques are gaining traction in clinical practice and have been recognised in new dementia diagnostic criteria.<sup>2</sup> However, there is a concern that enthusiasm and uptake of these technologies is premature

and the supporting evidence may not be sufficiently robust.<sup>3</sup>

The scientific assessment of diagnostic test accuracy (DTA) is a field that has historically been prone to methodological limitations and biases.<sup>4</sup> This has been particularly true in the area of dementia research. The DTA landscape is evolving. With an aim to improve design, conduct, reporting, assessment and comparison of DTA studies, best practice guidance such as the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) and Standards for Reporting of Diagnostic Accuracy (STARD) have been developed.<sup>5 6</sup>

QUADAS consists of a checklist of 14 items to consider when assessing potential bias arising from methodological limitations in DTA studies. The tool aims to capture aspects relating to internal and external validity. It was derived through a Delphi procedure,<sup>7</sup> informed by two systematic reviews—the first looking at potential sources of bias and variation in DTA studies<sup>8</sup> and the second focusing on existing quality assessment tools.<sup>9</sup> STARD is a 25-item checklist, developed following an extensive literature search. Its purpose is to support evaluation of the completeness and accuracy of study reporting in DTA.

As well as methodological and reporting quality, another way to quantify the ‘success’ of a scientific paper is to measure impact. For scientific data to effect change, the results need to reach the appropriate audience and then inform subsequent research, policy or practice. The importance of the construct of broader impact is increasingly recognised by government, universities and funding bodies. However, methods to measure or quantify impact remain open to interpretation with no consensus definition. Impact can operate at many levels and there is a difficulty in assigning a single definition to this multidimensional construct.<sup>10</sup> As it is challenging, or in some cases impossible, to directly estimate how a single publication has influenced current thought and inspired advances in research and policy change, impact is typically assessed using surrogate measures.

Traditional surrogate measures of impact have involved the use of citation-based metrics, particularly citation counts for the individual paper and the impact factor of the journal it is published in. Citation counts are assumed to reflect the received attention and usage of a paper within a scientific community, while publication in a journal with a high impact factor is often assumed to indicate adherence to strict requirements and high journal standards, as well as recognition by peer reviewers.<sup>11</sup>

In more recent years, with the advent of electronic publication and usage, academic and public social media, there has been a growing interest in ‘Altmetrics’—an alternative method of assessing publication impact.<sup>12</sup> Altmetrics aim to quantify the digital reads, online mentions and usage of research papers, some beyond the boundaries of scientific publishers and communities. Gathering information from social media, blogs and mainstream news outlets, Altmetrics are considered to have two important advantages—allowing insight into how much attention

scientific outputs receive from the general public and almost immediately providing means of evaluating impact without the delays associated with acquiring citations.

One would assume that scientific papers indicated to have the highest impact in terms of traditional and alternative metrics would also be the papers with the highest quality of reporting and methodological rigour. However, reports from certain research areas suggest that this is not necessarily the case.<sup>13–16</sup> This is an important issue, as papers with high visibility and usage are likely to influence the direction of future research, clinical practice and healthcare policy. Given the volume of recent original research concerning dementia biomarkers and the substantial interest some of these publications have generated, we felt that dementia biomarkers would be a useful ‘substrate’ to expand the description of the association between measures of study quality and quantitative measures of study impact.

### Study objective

The overall aim of this study was to assess whether methodological quality and completeness of reporting in dementia biomarker papers, assessed using the QUADAS and STARD checklists, respectively, was associated with traditional and alternative metrics of impact. Recognising the complexity and variability of influences on publication impact, we additionally explored other plausible factors that could influence impact and described whether reporting and study quality had independent associations with the various traditional and contemporary measures of impact.

## METHODS

### Study search and selection

Our data collection and analysis followed a preregistered protocol (reviewregistry89 in Registry of Systematic Reviews/Meta-Analyses; <http://www.researchregistry.com>). We performed a Medline search in collaboration with an information scientist from Cochrane Dementia to identify papers on dementia biomarkers published between January 2000 and August 2011. We chose this time horizon to allow sufficient time for all included studies to acquire citations to evaluate potential impact. The full search strategy is provided in the supplementary materials (see online supplementary appendix 1) or alternatively can be found at: <http://www.medicine.ox.ac.uk/alouis/content/diagnostic-test-accuracy-search-strategy>.

Biomarkers of interest included  $\beta$ -amyloid and tau levels in CSF and results from positron emission tomography and MRI. We included studies with a longitudinal design, involving participants who at baseline had objective cognitive impairment with no dementia. Two assessors independently rated the quality of the included primary studies, using the primary versions of both the QUADAS and STARD tools. Both checklists having undergone revision and newer versions are available, used the earlier iterations as these would have been the benchmark measures

of methodological quality and reporting at the time the papers were published. We have described the search strategy, study selection process and quality assessment in detail previously.<sup>3</sup>

### Variables

With the aim to investigate the relationship between study quality and measures of impact, we considered total QUADAS and STARD scores, calculated as the sum of individual item scores, the main determinants of interest. However, we recognised additional variables as being related to quality, reliability and generalisability of research evidence, including: total number of participants at baseline, number of participants with dementia at the end of a study, reporting of quantitative accuracy measures (eg, sensitivity, specificity) and providing a statement on COI.

We included two further variables in view of the specific research area of analysed studies—type of biomarker used (CSF, neuroimaging) and use of data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).<sup>17 18</sup> Although the latter does not directly imply differences in study content or quality, ADNI has received substantial financial support from both public and private sectors, is considered an influential project with a high number of associated publications, and a model for similar programmes subsequently developed worldwide.<sup>19</sup> Thus it appeared plausible that reported use of ADNI data may be associated with increased publication impact.

We also considered a group of potentially confounding variables in view of findings from other areas of research, reporting on a number of factors influencing publication impact, despite their lack of association with study quality. This involved: study location (North America vs Europe, Africa, Asia, Australia and South America),<sup>20</sup> funding source (industry vs academia),<sup>13</sup> journal subject area (neuroscientific vs multidisciplinary or other research area),<sup>21</sup> and authors' conclusions,<sup>22</sup> here specifically relating the utility of the investigated biomarker (positive vs neutral and negative). Finally, as citation rates are considered to be a time-dependent measure,<sup>23</sup> we also included the number of years since publication as a potential determinant.

We were initially interested in analysing four outcome measures. Two were traditional impact metrics – average number of citations per year since the paper's year of publication and 5 year impact factor of the publishing journal. We collected these data using Web of Science citation reports (Clarivate Analytics). As a third measure, we chose an alternative metric, reflecting the attention a publication receives in social media<sup>24 25</sup> – the Altmetric Attention Score, determined using the Bookmarklet tool.<sup>26</sup> The final outcome measure we considered for the analysis related to counts of electronic reads/downloads reported by journal websites. A full list of the variables is provided with definitions in the supplementary materials (see online supplementary appendix 2).

A single researcher (SM) collected study-level data and impact data, with a random selection (20%) cross-checked by an independent researcher (TQ). An exception was made for categorising authors' conclusions on biomarker utility and journal subject area, having recognised that these tasks may be prone to subjectivity. Therefore, both researchers assessed each included study on these variables independently, with discrepancies discussed and resolved through consensus. The publishing journals were categorised based on their name in all cases where it provided a clear indication of the journal's subject area (eg, *Radiology*, *Neurology*). If the subject area could not be directly inferred from the name (eg, *Brain*, *JAMA*), categorisation was based on the journal's own description of its scope, as presented on its website's home page.

### Statistical analysis

We used Spearman correlation coefficients and Mann-Whitney U tests to explore univariable associations between the outcome measures and continuous and dichotomous determinants, respectively. As this entailed performing multiple tests (15 per outcome measure), we adjusted the critical alpha level using the Holm-Bonferroni technique,<sup>27</sup> which allows to reduce the possibility of a Type I error, while offering increased power as compared with the Bonferroni method. For interpreting observed effect sizes, we used the Rule of Thumb, presented in *Applied Statistics for Behavioural Sciences (fifth edition)*.<sup>28</sup>

Following these procedures, we conducted multivariable linear regression analyses for each measure of publication impact. When assessed, we found that assumptions on linear relationships between variables and lack of multicollinearity were met. However, an inspection of scatterplots and normal probability plots of residuals revealed violations of the assumptions on homoscedasticity and normality of error distribution. Therefore, we applied a natural log transformation to the dependent variables. A subsequent inspection of plots indicated that both assumptions were satisfied.

With mean citation rates and 5 year journal impact factor likely being related, when we entered one impact measure into the analysis as an outcome, we included the other as one of the explanatory variables to account for its contribution to the variance of the former. We chose a stepwise regression method, which allowed us to compute the most parsimonious linear model by adding only variables that significantly increase its explanatory value, while removing variables that become non-significant as others are added. We conducted all analyses using SPSS (Version 22, IBM).

### RESULTS

The search strategy identified 19104 research publications, of which we found 142 eligible for inclusion. References for these studies are provided in online supplementary appendix 3 in the supplementary materials. We were able to extract complete study data on variables



**Table 1** Descriptive statistics of study level and impact data

QUADAS score (range 0–14), mean ( $\pm$ SD)	8.9 ( $\pm$ 1.5)
STARD score (range 0–25), mean ( $\pm$ SD)	16.1 ( $\pm$ 4.0)
Baseline sample size, mean ( $\pm$ SD)	171.3 ( $\pm$ 288.2)
No. of participants with dementia at end of study, mean ( $\pm$ SD)*	47.3 ( $\pm$ 50.4)
Report on biomarker accuracy, n (%)	52 (36.6)
Statement on COI, n (%)	52 (36.6)
CSF biomarker, n (%)	56 (39.4)
Imaging biomarker, n (%)	102 (71.8)
ADNI study population, n (%)	11 (7.7)
North American study location, n (%)	60 (42.3)
Pharmaceutical funding source, n (%)	12 (8.5)
Neuroscientific journal, n (%)	85 (59.9)
Positive authors' conclusions, n (%)	122 (85.9)
No. of years since publication, mean ( $\pm$ SD)	9.1 ( $\pm$ 3.0)
Average no. of citations per year, mean ( $\pm$ SD)	12.8 ( $\pm$ 15.5)
5-year journal impact factor, mean ( $\pm$ SD)*	6.1 ( $\pm$ 3.9)

\*Data from 141 studies.

ADNI, Alzheimer's Disease Neuroimaging Initiative; COI, conflict of interest; CSF, cerebrospinal fluid; QUADAS, Quality Assessment of Diagnostic Accuracy Studies checklist; STARD, Standards for Reporting of Diagnostic Accuracy checklist.

of interest for all but two included papers – one, where we could not determine the number of patients with dementia at the end of follow-up, and one, where there was no report on the impact factor for the publishing journal. We excluded these two studies only from analyses involving variables with the missing values. Independent validation of extracted data revealed no errors in results recorded. Extracted study data, as well as QUADAS and STARD scores, are presented for each individual publication in the supplementary materials (see online supplementary appendix 4 and 5, respectively).

We found that counts of electronic reads/downloads were available from only a few platforms, and therefore we did not pursue this analysis. Similarly, we did not include Altmetric Attention Scores as an outcome measure, due to only 15% of the studies having a score greater than 0 ( $M=0.67$ ;  $SD=2.16$ ; range: 0–14).

Table 1 presents descriptive data for the included studies. In relation to key variables of interest, studies on average obtained a QUADAS score of 8.9, with a lowest score of 3 and a highest of 12, meaning that none of the studies satisfied all 14 checklist items. The average STARD score was 16.1, with a minimum of 7 and a maximum of 25 – the highest possible score, which was obtained by four studies. We observed relatively large variability in terms of traditional metrics. The mean for average number of citations per year was 12.8, with a minimum of 0.4 and maximum of 117.7, while on average the 5 year

journal impact factor was 6.1, with a lowest value of 1.2 and highest of 33.6.

Results from univariable analyses are presented in table 2. Having applied the Holm-Bonferroni correction, we found no significant association between measures of methodological and reporting quality and either studied measure of impact. We observed that a higher average number of citations per year was weakly associated with papers being more recent, as well as published in a neuroscientific journal. Neuroscientific journal subject area was also very weakly associated with 5 year journal impact factor, together with papers including a statement on COI and a North American study location. However, overall, the strongest positive correlation we observed, at a moderate level, was between the two measures of impact.

Results of the linear regression only partially reflected the observed univariable associations. The most parsimonious model for explaining variance in the average number of citations per year involved four of the considered variables: 5 year journal impact factor ( $\beta=0.42$ ;  $P<0.001$ ), journal subject area ( $\beta=0.34$ ;  $P<0.001$ ), number of years since publication ( $\beta=-0.29$ ;  $P<0.001$ ) and total STARD score ( $\beta=0.13$ ;  $P<0.05$ ); ( $F(4, 135)=32.81$ ,  $P<0.001$ ,  $R^2=0.49$ ). While a model incorporating three variables explained the greatest proportion of variance in 5 year journal impact factor ( $F(3, 136)=20.47$ ,  $P<0.001$ ,  $R^2=0.31$ ), although here it included: average number of citations per year ( $\beta=0.45$ ;  $P<0.001$ ), statement addressing COI ( $\beta=0.22$ ;  $P<0.01$ ), and total number of study participants at baseline ( $\beta=0.15$ ;  $P<0.05$ ).

## DISCUSSION

Across a substantial literature describing the test accuracy of dementia biomarkers, we found that citation rates varied considerably, with some studies showing relatively high impact in terms of traditional metrics. At the same time, the methodological and reporting quality of studies, measured using the QUADAS and STARD tools respectively, was on average quite low. We found little evidence of an association between traditional measures of paper impact and its reporting or methodological quality.

Although we found that STARD scores were an independent determinant of the average number of citations per year, the association was very weak, with journal characteristics explaining considerably more of the variance. These findings are consistent with results from previous studies in other fields of clinical research, where either limited or no association was found between citation rates and indicators of quality.<sup>29–31</sup>

In addition, we did not observe any significant relationship between the quality checklist scores and 5 year journal impact factor. The latter was however very weakly associated with the total number of study participants, a factor impacting on the precision of diagnostic accuracy estimates.<sup>32</sup> There was also a weak association with the inclusion of a statement on conflict of interest (COI),

**Table 2** Univariable associations between publication and journal characteristics and measures of impact

	Average citations per year		5-year journal impact factor	
	Effect size (r)	P value	Effect size (r)	P value
QUADAS score	0.12	0.144	0.04	0.670
STARD score	0.18	0.035	0.02	0.797
Baseline sample size	0.05	0.522	0.17	0.040
No. of participants with dementia at end of study	0.11	0.190	0.13	0.120
Report on biomarker accuracy	0.11	0.187	0.07	0.384
Statement on COI	0.11	0.204	0.28	0.001*
CSF biomarker	0.07	0.410	0.04	0.646
Imaging biomarker	0.09	0.301	0.10	0.251
ADNI population	0.17	0.046	0.21	0.012
Continent	0.19	0.024	0.29	0.001*
Funding	0.15	0.084	0.18	0.031
Journal subject area	0.46	<0.001*	0.28	0.001*
Authors' conclusions	0.03	0.707	0.06	0.500
No. of years since publication	-0.32	<0.001*	-0.05	0.562
Average no. of citations per year	-	-	0.63	<0.001*
5-year journal impact factor	0.63	<0.001*	-	-

\*P value indicating a significant association after applying a Holm-Bonferroni correction.

ADNI, Alzheimer's Disease Neuroimaging Initiative; COI, conflict of interest; CSF, cerebrospinal fluid; QUADAS, Quality Assessment of Diagnostic Accuracy Studies checklist; STARD, Standards for Reporting of Diagnostic Accuracy checklist.

which plays an important role in supporting research integrity.<sup>33</sup>

The finding that, unlike in the case of STARD, QUADAS scores were not related with either of the analysed impact measures, may partially be due to the assessment of methodological quality being dependent on completeness of reporting. In cases of Inadequate reporting, the quality of study methods may easily be misjudged, and in turn, the estimated association between methodological quality and study impact distorted.

The study results may appear concerning, yet one could argue that finding any relationship whatsoever between citation rates and indicators of quality is a positive sign. We recognise that a proportion of papers included in our study were published at a similar period, or even previously, to the QUADAS and STARD checklists, and some time is required before new guidelines become widely and fully incorporated into research practice. These quality tools should inform all aspects of a research project from study design, through manuscript publication, to external assessment. We hope that the current generation of DTA studies is making greater use of these tools and that we will see the published outputs soon.

There is no 'gold standard' measure of impact. For reasons of practicality, impact is typically assessed indirectly, through surrogate measures, which was also the case in the present study. However, the appropriateness of using traditional metrics for the evaluation of scientific output has been strongly questioned.<sup>10 34 35</sup>

These measures are criticised for susceptibility to bias (eg, related to language of publication) and risk of overlooking important differences in context, relating both to norms within specific research fields, as well as motives for citation. Journal impact factor in particular is argued to offer a poor representation of usage for an individual paper, being based on averaging across multiple publications, often with very skewed citation distributions and citation rates.<sup>10 21 36</sup> Our results appear to confirm that individual paper citation rates and journal impact factor measure different dimensions of impact, with one explaining little variance within the other and different variables contributing to the two final models.

Despite the limitations of traditional metrics, they are still often used to evaluate the achievements of research groups and institutions and are considered for the purposes such as promotion and funding allocation.<sup>11 34</sup> An increasing number of researchers are voicing their concerns regarding this practice and suggest that quantitative measures be used only to support qualitative expert assessment,<sup>34 35</sup> allowing a more accurate evaluation of a study's novelty and its contribution to a specific field of research and practice. If impact were to be assessed in this advocated way, it is possible that a stronger association would be found between publication impact and its methodological and reporting quality. However, as long as traditional metric remain the primary method of assessing publication impact, it is important that even

papers with presumed high impact are scrutinised in terms of their quality.

Strengths of this study involve following best practice in conducting our analysis of associations and including a variety of potential determinants of publication impact, selected in view of previous research findings.<sup>37 38</sup> We further performed a comprehensive and explicit publication search, aided by the Cochrane group. One of the limitations, however, is that this search focused on one database only (Medline), potentially increasing the risk of missing relevant papers. As our purpose was to assess a range of sources with a spread of impact, it is not such a concern if the search was not completely comprehensive. The titles returned were however similar to the search results from a more comprehensive Cochrane review of biomarkers.<sup>39 40</sup>

Another important limitation of this and similar studies is that due to restricted time and resources, we did not conduct an analysis of citation context. Although papers are typically referenced in view of the evidence they provide in support of a particular hypothesis, they may also be mentioned in a perfunctory way, summarising methods or even in a negative way. This variability in citation context may partially explain why the relationship between measures of impact and study quality appears so modest, and it is possible that a stronger association would be found if only references made in a positive context were to be considered.

Finally, we included studies published up until August 2011. Although the time elapsed from publication was sufficient for acquiring citations, providing a good opportunity for measuring citation-based impact, the associations we studied may have been subject to gradual change, for example, due to increasing endorsement of best practice guidelines. This may limit the applicability of our findings in view of more recent studies. A related issue is that, due to lack of sufficient data, we were not able to investigate the relationship between measures of quality and contemporary, alternative metrics. As altmetrics are relatively new, it has been suggested that indeed this measure has little usefulness when assessing papers published before 2011,<sup>41</sup> which was the case in the present study. With research evidence indicating at most a moderate relationship between citation rates and altmetrics,<sup>41–43</sup> it seems that these two metrics reflect somewhat different kinds of impact and therefore may be affected by different variables.

Future research should investigate whether the association with indicators of study quality is different for alternative metrics, reflecting online mentions and digital reads than for traditional measures of impact, focusing on citation counts and the impact factor of publishing journals. This would necessitate analysing more recent publications, where it is recommended to use the revised versions of QUADAS and STARD checklists for quality assessment—QUADAS-2, STARD 2015 and STARDem, the latter having dementia-specific guidance.<sup>44–46</sup> An additional advantage of an updated analysis would be

the possibility of assessing whether the introduction and promotion of quality checklists has resulted over time in strengthening associations between measures of study quality and any relevant measures of publication impact—traditional and alternative, quantitative and qualitative.

## CONCLUSION

Findings from this study indicate that publication in a presumed high-impact journal or frequent citation is no guarantee that a paper is methodologically robust. Clinicians, researchers and policy-makers should critically appraise all published evidence regardless of impact. An assessment of study quality should further accompany the use of traditional metrics when evaluating the scientific impact of research work within academic institutions. To improve methodological and reporting quality of future studies, it is important that researchers and authors adhere to current best practice guidelines, while journal editors mandate submitting manuscripts accompanied by completed quality checklists.

**Acknowledgements** Jenny McCleery, Coordinating Editor of the Cochrane Dementia and Cognitive Improvement Group, supported this work.

**Contributors** SM, AHN-S, RM and TQ contributed to the conception and design of the study and acquisition of data; SM, BAD and MH contributed to the analysis and interpretation of the data; SM, BAD and TQ drafted the final manuscript. All authors critically revised the manuscript, approved its final version and agreed to be accountable for its content.

**Funding** SM was supported by an Alzheimer's Society Research Bursary, grant reference: 302 (AS-URB-16-004); BAD was supported by a Stroke Association Priority Program Award, grant reference: PPA 2015/01\_CS0; TQ was supported by a Chief Scientist Office and Stroke Association Senior Clinical Lectureship, grant reference: TSA LECT 2015/05.

**Competing interests** AHN-S, RM, TQ have all published methodological and reporting guidance around raising standards in dementia test accuracy research.

**Patient consent** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
2. Dubois B, Feldman HH, Jacova C, *et al*. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* 2010;9:1118–27.
3. Noel-Storr AH, Flicker L, Ritchie CW, *et al*. Systematic review of the body of evidence for the use of biomarkers in the diagnosis of dementia. *Alzheimers Dement* 2013;9:e96–e105.
4. Rutjes AWS. Evidence of bias and variation in diagnostic accuracy studies. *Can Med Assoc J* 2006;174:469–76.
5. Whiting P, Rutjes AW, Reitsma JB, *et al*. The development of QUADAS: a tool for the quality assessment of studies of diagnostic

- accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
6. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–4.
  7. Murry JW, Hammons JO. Delphi: a versatile methodology for conducting qualitative research. *Rev High Ed* 1995;18:423–36.
  8. Whiting P, Rutjes AW, Reitsma JB, *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.
  9. Whiting P, Rutjes AW, Dinnes J, *et al.* A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol* 2005;58:1–12.
  10. Bollen J, Van de Sompel H, Hagberg A, *et al.* A principal component analysis of 39 scientific impact measures. *PLoS One* 2009;4:e6022.
  11. Wang J, Shapira P. Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *PLoS One* 2015;10:e0117727.
  12. Priem J, Taraborelli D, Groth P, *et al.* Altmetrics: a manifesto. 2010 <http://altmetrics.org/manifesto>.
  13. Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. *PLoS One* 2007;2:e403.
  14. Leimu R, Koricheva J. What determines the citation frequency of ecological papers? *Trends Ecol Evol* 2005;20:28–32.
  15. Berghmans T, Meert AP, Mascaux C, *et al.* Citation indexes do not reflect methodological quality in lung cancer randomised trials. *Ann Oncol* 2003;14:715–21.
  16. Seglen PO. Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthop Scand* 1998;69:224–9.
  17. Mueller SG, Weiner MW, Thal LJ, *et al.* The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am* 2005;15:869–77.
  18. Mueller SG, Weiner MW, Thal LJ, *et al.* Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* 2005;1:55–66.
  19. Weiner MW, Veitch DP, Aisen PS, *et al.* The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement* 2013;9:e111–e94.
  20. Braun T, Glänzel W, Grupp H. The scientometric weight of 50 nations in 27 science areas, 1989–1993. Part I. All fields combined, mathematics, engineering, chemistry and physics. *Scientometrics* 1995;33:263–93.
  21. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314:497–502.
  22. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218–28.
  23. Todorov R, Glänzel W. Journal citation measures: a concise review. *Journal of Information Science* 1988;14:47–56.
  24. Wouters P, Costas R. *Users, narcissism and control: tracking the impact of scholarly publications in the 21st century*. SURF foundation, 2012.
  25. Sugimoto C. "Attention is not impact" and other challenges for altmetrics: Discover the future of research: Wiley Exchanges, 2015.
  26. Altmetric. Bookmarklet for researchers. <https://www.altmetric.com/products/free-tools/bookmarklet/> (accessed Jan 2017).
  27. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979;6:65–70.
  28. Hinkle DE, Wiersma W, Jurs SG. *Applied statistics for the behavioral sciences*. 5th edn, 2003.
  29. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 2002;287:2847–50.
  30. West R, McIlwaine A. What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction* 2002;97:501–4.
  31. Nieminen P, Carpenter J, Rucker G, *et al.* The relationship between quality of research and citation frequency. *BMC Med Res Methodol* 2006;6:42.
  32. Bachmann LM, Puhan MA, ter Riet G, *et al.* Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006;332:1127–9.
  33. Institute of Medicine. *Conflict of interest in medical research, education, and practice*. Washington, DC: The National Academies Press, 2009.
  34. Hicks D, Wouters P, Waltman L, *et al.* Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 2015;520:429–31.
  35. DORA. *San Francisco declaration on research assessment*. San Francisco, 2014.
  36. Lozano GA, Larivière V, Gingras Y. The weakening relationship between the impact factor and papers' citations in the digital age. *J Assoc Inf Sci Tech* 2012;63:2140–5.
  37. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br J Math Stat Psychol* 1992;45:265–82.
  38. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996;49:907–16.
  39. Ritchie C, Smailagic N, Noel-Storr AH, *et al.* Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2014;10:Cd008782.
  40. Smailagic N, Vacante M, Hyde C, *et al.* <sup>18</sup>F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2015;1:Cd010632.
  41. Costas R, Zahedi Z, Wouters P. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *J Assoc Inf Sci Technol* 2015;66:2003–19.
  42. Zahedi Z, Costas R, Wouters P. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics* 2014;101:1491–513.
  43. Thelwall M, Haustein S, Larivière V, *et al.* Do altmetrics work? Twitter and ten other social web services. *PLoS One* 2013;8:e64841.
  44. Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
  45. Noel-Storr AH, McCleery JM, Richard E, *et al.* Reporting standards for studies of diagnostic test accuracy in dementia: the STARDdem Initiative. *Neurology* 2014;83:364–73.
  46. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 2015;277:826–32.