



Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities

Danielle B Rice,^{1,2} Hana Raffoul,^{2,3} John P A Ioannidis,^{4,5,6,7} David Moher^{8,9}

For numbered affiliations see end of the article.

Correspondence to: D Moher
dmoher@ohri.ca
(or @dmoher on Twitter
ORCID 0000-0003-2434-4206)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;369:m2081

<http://dx.doi.org/10.1136/bmj.m2081>

Accepted: 6 April 2020

ABSTRACT

OBJECTIVE

To determine the presence of a set of pre-specified traditional and non-traditional criteria used to assess scientists for promotion and tenure in faculties of biomedical sciences among universities worldwide.

DESIGN

Cross sectional study.

SETTING

International sample of universities.

PARTICIPANTS

170 randomly selected universities from the Leiden ranking of world universities list.

MAIN OUTCOME MEASURE

Presence of five traditional (for example, number of publications) and seven non-traditional (for example, data sharing) criteria in guidelines for assessing assistant professors, associate professors, and professors and the granting of tenure in institutions with biomedical faculties.

RESULTS

A total of 146 institutions had faculties of biomedical sciences, and 92 had eligible guidelines available for review. Traditional criteria of peer reviewed publications, authorship order, journal impact factor, grant funding, and national or international reputation were mentioned in 95% (n=87), 37% (34), 28% (26), 67% (62), and 48% (44) of the guidelines, respectively. Conversely, among non-traditional

criteria, only citations (any mention in 26%; n=24) and accommodations for employment leave (37%; 34) were relatively commonly mentioned. Mention of alternative metrics for sharing research (3%; n=3) and data sharing (1%; 1) was rare, and three criteria (publishing in open access mediums, registering research, and adhering to reporting guidelines) were not found in any guidelines reviewed. Among guidelines for assessing promotion to full professor, traditional criteria were more commonly reported than non-traditional criteria (traditional criteria 54.2%, non-traditional items 9.5%; mean difference 44.8%, 95% confidence interval 39.6% to 50.0%; P=0.001). Notable differences were observed across continents in whether guidelines were accessible (Australia 100% (6/6), North America 97% (28/29), Europe 50% (27/54), Asia 58% (29/50), South America 17% (1/6)), with more subtle differences in the use of specific criteria.

CONCLUSIONS

This study shows that the evaluation of scientists emphasises traditional criteria as opposed to non-traditional criteria. This may reinforce research practices that are known to be problematic while insufficiently supporting the conduct of better quality research and open science. Institutions should consider incentivising non-traditional criteria.

STUDY REGISTRATION

Open Science Framework (https://osf.io/26ucp/?view_only=b80d2bc7416543639f577c1b8f756e44).

Introduction

Important deficiencies exist in the quality and transparency of research conducted across disciplines.^{1 2} Many efforts have been made to combat these inadequacies by developing, for example, reporting guidelines (for example, the Consolidated Standards of Reporting Trials (CONSORT) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements), registration of studies before data collection (for example, clinicaltrials.gov), and data sharing practices.^{3 4} Despite these strategies, poorly conducted and inadequately reported research remains highly prevalent.⁵ This has important consequences, especially in the field of medicine, as research is heavily relied on to inform clinical decision making.

Institutions have the ability to influence large scale improvements among researchers, as universities hire new faculty and promote and tenure existing faculty. Universities can provide incentives and rewards (for example, promotions) for scholarly work that is conducted appropriately, reported transparently, and

WHAT IS ALREADY KNOWN ON THIS TOPIC

Academics tailor their research practices according to the evaluation criteria applied within their academic institution

Ensuring that biomedical researchers are incentivised by adhering to best practice guidelines for research is essential given the clinical implications of this work

Changes to the criteria used to assess professors and confer tenure have been recommended, but no systematic assessment of promotion and tenure criteria being applied worldwide has been done

WHAT THIS STUDY ADDS

Across countries, university guidelines focus on rewarding traditional research criteria (peer reviewed publications, authorship order, journal impact, grant funding, and national or international reputation)

The minimum written requirements for promotion and tenure criteria are predominantly objective in nature, although several are inadequate measures to assess the impact of researchers

Developing and evaluating more appropriate, non-traditional indicators of research may facilitate changes in the evaluation practices for rewarding researchers

adheres to best publication practices. A recent survey conducted in the UK found that academics tailor their publication practices to align with their institutional evaluation criteria.⁶ These criteria, however, may include metrics that are known to be problematic for assessing researchers.⁷ Current incentives and rewards may also be misaligned with the needs of society. Reward systems within universities typically include criteria within promotion and tenure documents such as the quantity of publications and novelty of findings rather than the reliability, accuracy, reproducibility, and transparent reporting of findings.⁸ Inappropriate criteria being applied for career advancement can inadvertently contribute to research waste,⁹ with billions of dollars invested in non-usable research.¹⁰ For example, universities that emphasise the quantity of published papers can increase undeserved authorship, “salami slicing,” and publication in very low quality journals (for example, predatory journals) without peer review and contribute to the problems of reproducibility.

Institutional criteria for decisions about promotion and tenure can vary and may not be evidence based.¹¹ Some institutions set minimum quantitative “thresholds” for promotion, whereas others provide qualitative phrasing of criteria that scientists must meet. Recent articles identifying the limitations of the current criteria used to assess scientists for promotion and tenure have been largely conceptual in nature and have limited empirical evidence.^{11–14} In a recent study, evaluation documents were reviewed and potential limitations were identified.¹⁴ A group of 22 people, including academic leaders, representatives of health policy organisations and funders, and scientists, participated in a panel workshop about the perceived strengths and weaknesses of approaches used for assessing career advancement. Strategies to encourage implementation and uptake of more responsible indicators for assessing scientists were discussed, including six general principles for assessing scientists. These principles included aspects such as rewarding researchers for open science practices and the transparent and complete reporting of research.¹⁴ Before implementing changes to existing criteria, however, we must better understand the standards currently being applied.

Understanding the variability of criteria and thresholds for promotion and tenure applied across institutions requires a systematic empirical assessment. Therefore, we aimed to identify and document a set of pre-specified traditional and non-traditional criteria used to assess scientists for promotion and tenure within faculties of biomedical sciences among a large number of universities around the world.

Methods

The protocol for this study was registered within the Open Science Framework database (https://osf.io/26ucp/?view_only=b80d2bc7416543639f577c1b8f756e44) before the study's data collection. We used the STROBE checklist for cross sectional

studies to ensure that methods and findings are clearly reported (see appendix 1).¹⁵

Approach to selecting university institutions

We used the Centre for Science and Technology Studies (CSTS) Leiden ranking of world universities from 2017 (<https://www.leidenranking.com/ranking/2017/list>) to select institutions for inclusion in the study. We selected a random sample of 20% (170/854) of institutions from the Leiden ranking list by using online random sampling software (<https://www.randomizer.org/>). We selected the CSTS ranking list for the field of “Biomedical and Health Sciences,” which represents the field to which publications from universities are assigned. We planned to include all randomly selected institutions on this list, regardless of the faculties present at each university. We used the default settings on the CSTS website. The default indicator settings include type of indicator (‘impact’) and indicators (“P, P(top 10%), PP(top 10%)”). This indicator represents the number and proportion of a university's publications that, compared with other publications in the same field and in the same year, are among the top 10% of most frequently cited publications. We ordered the list by publications and selected the calculation of impact indicators by using fractional counting option. A minimum publication output was set at the default value of 100.

Searching of institution criteria

Two reviewers searched for institutional criteria by using an iterative process. They searched each institution's (that is, university's) website for the criteria and policies used for evaluation, promotion, and tenure in the faculty of medicine or biomedical sciences. In instances where medical schools were their own entity, they searched the medical school's website for evaluation, promotion, and tenure documents. The reviewers first determined whether each institution had a relevant biomedical sciences department or faculty (for example, faculty of medicine, department of science). If a faculty of biomedical sciences existed, they searched for keywords on the faculty's website including “academic performance”, “career mobility”, “criteria”, “evaluation”, “guidelines”, “policy”, “tenure”, and “promotion” to find documents related to promotion and tenure. If no faculty related to biomedical sciences existed, or if promotion and tenure guidelines were not publicly available at the faculty level, the reviewers referred to the available institution level guidelines. If publicly available criteria could not be located after searching with these methods, we contacted human resource personnel, professors, and academic affairs administrators for the institution directly on up to two occasions to request access to faculty or institution level criteria. In some countries, promotions first need to meet criteria set at a state or national level. After searching for faculty level and institution level guidelines if these were not available, we also searched for state or national guidelines by applying the same search techniques used for

universities. For institutions' websites that were published in languages other than English or French, a person who was fluent in the relevant language searched the promotion and tenure information available on the website to facilitate data collection, and we sent emails in the language of the institution's website. Twelve translators searched university, regional, and national websites for documents to facilitate data extraction. These people also translated an email to send to institution representatives when guidelines could not be found online.

Approach to selecting list of criteria

We purposively selected 12 criteria related to promotion and tenure to enable a comparison between traditional (for example, quantity of research) and non-traditional (for example, reproducibility of research) criteria used to assess scientists for promotion and tenure (see box 1). Although the characterisation of traditional and non-traditional criteria was ultimately subjective, we based our decisions on evidence and policy initiatives from multiple jurisdictions.¹²⁻¹⁴ We used an iterative process to select the characteristics. An early version of the criteria included 10 items; however, after pilot testing a set of five institutions, we added two additional items and made revisions to the wording of some items. The final set of criteria included five

traditional criteria (peer reviewed publications, authorship order, journal impact, grant funding, national or international reputation) and seven non-traditional criteria (citations, data sharing, publishing in open access mediums, registration of research, adherence to reporting guidelines, alternative ways for sharing research, accommodations for employment leave).

Data collection

For each eligible institution, we extracted the following information: university name, faculty name, country, and human development index rating of country. We reviewed guidelines used by faculties of biomedical sciences or institutions for the evaluation of professors, where available, to determine whether each of the 12 items from our list of criteria for promotion and tenure of faculty were present. We recorded the relevant mentions for each criterion, regardless of exactly how the criterion was considered or operationalised. We did not intend to arbitrate whether the proposed version of the criterion was appropriate and technically sophisticated; however, we collected information about whether guidelines applied thresholds for each criterion.

When promotion and tenure guidelines were available, we first reviewed the table of contents and located the section on criteria for promotion and tenure and reviewed this section of the document, including any sections that the document referred to for context. If a table of contents was not provided, we reviewed the document in its entirety. We then reviewed and extracted the presence of criteria and the relevant statement for each level of promotion criteria published for universities, including promotions to assistant professor, associate professor, and full professor and the granting of tenure, as well as whether a criterion was mentioned for at least one of these levels. We considered these levels of promotion on the basis of a North American framework of career advancement. Where institutions applied different labels to ranks/positions (for example, researcher level A), we sought documentation for the appropriate equivalent categorisation of the promotion levels. If documentation describing the position was not available, we consulted with professors from the institution's country to equate positions with those being applied in our study. If no equivalent position was available, we did not include the institution in our sample ($n=3$ institutions). We extracted this information for tenure track positions rather than non-tenure track positions. We did not extract promotion and tenure criteria for aspects of career advancement related to teaching or clinical duties or for positions that comprised more educational or clinical activities than research activities. We also extracted the level of the promotion criteria available (that is, faculty level criteria, departmental level criteria), the year that the promotion guideline was published, the associated URL of the criteria, and the date that the website was searched. Two reviewers (DBR, HR) independently

Box 1: Criteria of interest for promotion and tenure

Traditional criteria

1. Is any quantitative or qualitative mention made about publications required? If quantitative, please specify the requirement
2. Is any quantitative or qualitative mention made about the specific authorship order in publications? If so, please specify order (eg, first, senior, single) required
3. Is any mention made of journal impact factors? If quantitative, what are the minimum thresholds?
4. Is any mention made of grant funding? If quantitative, what are the minimum thresholds (ie, amount of funding and/or number of grants as principal investigator)?
5. Is any mention made of requiring that research is recognised at a national or international level? If so, please specify the requirement

Non-traditional criteria

1. Is any mention made of citations? If quantitative, what are the thresholds of minimum requirement? Are specific citation databases mentioned?
2. Is any mention made of data sharing? If quantitative, what are the minimum thresholds (eg, percentage of data that is to be made available)?
3. Is any mention made of publishing in open access mediums? If quantitative, what are the minimum thresholds (eg, percentage of studies to be published in open access journals)?
4. Is any mention made of registration (including preregistration challenge) of studies? If yes, are there thresholds of minimum requirement (eg, percentage of studies that are to be registered)?
5. Is any mention made of adherence to reporting guidelines for publications? If so, are specific guidelines mentioned?
6. Is any mention made of alternative metrics for sharing research (eg, social media and print media)? If so, are specific metrics mentioned?
7. Is any mention made of accommodations or adjustments to expectations due to employment leave? If so, please specify the description of accommodations (eg, an extra year to defer tenure consideration) and the type of eligible circumstances (eg, parental leave, medical leave)

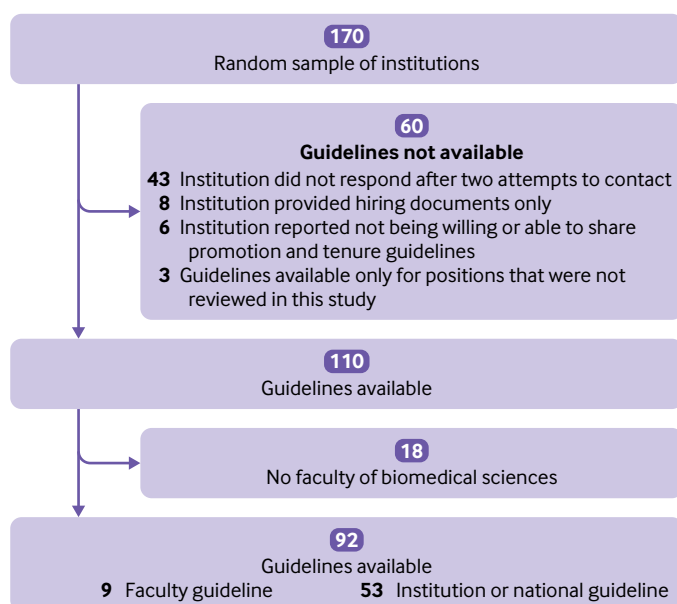


Fig 1 | Included universities

extracted all data, and results were compared for consistency. Where consensus was not achieved between reviewers after discussion, a third team member (DM) was consulted to resolve discrepancies in the interpretation of criteria. For guidelines published in languages other than English or French, translation of the relevant documents was conducted by one person and verified by a second reviewer (DBR) using Google Translate. A Hungarian translator was not available for one guideline. For this guideline, one reviewer (DBR) used Google Translate to extract data. We used a standardised electronic data collection form in Distiller Systematic Reviewer (Evidence Partners, Ottawa, Canada) for data collection.

Approach to synthesis

We summarised institutions' characteristics and promotion and tenure criteria in table format to facilitate inspection and discussion of findings.

We compared the percentage of criteria that were included in promotion and tenure guidelines with a paired sample *t* test. We present categorical variables as percentages and counts and continuous variables as means and standard deviations or medians and interquartile ranges. We compared institutions that had criteria available with institutions that did not have criteria available by using independent samples *t* test, χ^2 tests, or non-parametric tests.

We did exploratory analyses for the full professor position, as this had the most data available. We conducted two multiple linear regressions to assess the associations between institutions' characteristics (independent variables: level of criteria, continent, human development index, and Leiden ranking) and the number of criteria present for traditional and non-traditional items for guidelines for professors (dependent variable), as most institutions had guidelines for this promotion level. We did logistic regressions for each criterion present among 10% to 90% of institutions at the rank of full professor to assess the associations between institutional characteristics (independent variables) and the presence of each criterion (dependent variable). We selected institutional characteristics as covariates given their availability and the potential relevance to the type of criteria applied (for example, regional differences in career advancement procedures). If an independent variable did not have at least one institution with and one without the item criteria, we removed that variable from the logistic regression. Before doing regression analyses, we did preliminary tests to confirm that no violations of multiple regression assumptions existed. We used Microsoft Excel for summing study results and SPSS version 21.0 for statistical tests. All statistical analyses were two tailed with $P < 0.005$ significance level to adhere to recent recommendations for a lowered threshold of statistical significance.^{14 16}

Patient and public involvement

This research did not involve consultation with patients or the public.

Results

Deviations from protocol

We refined our inclusion criteria to exclude institutions that did not have a faculty of medicine, biomedical sciences, life sciences, health sciences, or medical sciences in order to focus on institutions that had a department directly related to studying and subsequently disseminating biomedical research. In the regression analyses, we excluded data from continents that had fewer than two institutions with guidelines available. This resulted in exclusion of one institution from South America and one institution from Africa from regression analyses.

Characteristics of institutions

Of the 170 institutions reviewed, 146 had faculties of biomedical sciences. We were able to obtain a total of 92 (63%) institutions' guidelines for promotion and

Table 1 | Summary characteristics of included institutions (n=146). Values are numbers (percentages) unless stated otherwise

Variables	Institutions with guidelines available (n=92)	Institutions with guidelines not available (n=54)	P value
Criteria level available:			
Institution	53 (58)	NA	NA
Faculty	39 (42)		
Median (IQR) Leiden ranking	346 (162-553)	462 (199-616)	0.23
Continent:			0.001
South America (n=6)	1 (17)	5 (83)	
Australia (n=6)	6 (100)	0 (0)	
Europe (n=54)	27 (50)	27 (50)	
North America (n=29)	28 (97)	1 (3)	
Asia (n=50)	29 (58)	21 (42)	
Africa (n=1)	1 (100)	0 (0)	
Human development index:			0.92
Very high (n=107)	68 (64)	39 (36)	
High (n=37)	23 (62)	14 (38)	
Medium (n=2)	1 (50)	1 (50)	

IQR=interquartile range; NA=not applicable.

Table 2 | Criteria of interest for promotion and tenure. Values are numbers (percentages)

Criteria	Presence of criteria for:				Example of quantitative information if present	Example of relevant quote from university website
	Assistant professor (n=49)	Associate professor (n=79)	Full professor (n=83)	Tenure (n=26)		
Traditional criteria						
1. Is any quantitative or qualitative mention made about publications required? If quantitative, please specify the requirement	39 (80)	76 (96)	79 (95)	22 (85)	Minimum two research papers or one paper and one book	"Publication in refereed journals or series, or by publishers recognized as leaders in the field"
2. Is any quantitative or qualitative mention made about the specific authorship order in publications? If so, please specify order (eg, first, senior, single) required	11 (22)	28 (35)	28 (34)	9 (35)	Three papers as first or corresponding author	"In case of multi-authored work, at least one of the peer reviewed publications must be sole authored"
3. Is any mention made of journal impact factors? If quantitative, what are the minimum thresholds?	12 (24)	24 (30)	23 (28)	2 (8)	At least one impact factor >3.0; at least one impact factor >5.0 or accumulated impact factor >11.0	"At least 3 publications in international journals with reasonable impact factor are required"
4. Is any mention made of grant funding? If quantitative, what are the minimum thresholds (ie, amount of funding and/or number of grants as principal investigator)?	26 (53)	50 (63)	56 (67)	15 (58)	Principal investigator of one provincial project; principal investigator or main member (top three) of project with more than 500 000 RMB	"The candidate has engaged in research grants/contracts as Principal or Co-Investigator at a funding level appropriate to the discipline, possibly in collaboration with other Universities or organizations"
5. Is any mention made requiring that research is recognised at a national or international level? If so, please specify the requirement	11 (22)	26 (33)	39 (47)	11 (42)	NA	"Performance of exceptional distinction and achievements that are recognized as distinguished internationally or nationally (meeting the benchmarks)"
Non-traditional criteria						
6. Is any mention made of citations? If quantitative, what are the thresholds of minimum requirement? Are specific citation databases mentioned?	12 (24)	23 (29)	23 (28)	6 (23)	One paper cited more than 20 times	"Achieves a citation rate or proportion of research outputs in most prestigious outlets [...] in line with discipline and leading universities"
7. Is any mention made of data sharing? If quantitative, what are the minimum thresholds (eg, percentage of data that is to be made available)?	1 (2)	1 (1)	1 (1)	0	NA	"Sound data management is a basic requirement for this (academic analysis) and provides additional guarantees for a flawless methodology, for sharing and reusing data by other researchers in an Open Science context and for the accountability of a researchers own academic integrity"
8. Is any mention made of publishing in open access mediums? If quantitative, what are the minimum thresholds (eg, percentage of studies to be published in open access journals)?	0	0	0	0	NA	NA
9. Is any mention made of registration (including preregistration challenge) of studies? If yes, are there thresholds of minimum requirement (eg, percentage of studies that are to be registered)?	0	0	0	0	NA	NA
10. Is any mention made of adherence to reporting guidelines for publications? If so, are specific guidelines mentioned?	0	0	0	0	NA	NA
11. Is any mention made of alternative metrics for sharing research (eg, social media and print media)? If so, are specific metrics mentioned?	3 (6)	3 (4)	2 (2)	1 (4)	NA	"Ghent University has invested in a proper information system of research output (biblio, IWETO/FRIS) for many years and is currently extending this to other research-related areas (Gismo). As soon as publications and activities are properly registered in the information system, the administrative burden on researchers in the context of an evaluation (promotion, project applications) should be reduced to a minimum"
12. Is any mention made of accommodations or adjustments to expectations due to employment leave? If so, please specify the description of accommodations (eg, an extra year to defer tenure consideration) and the type of eligible circumstances (eg, parental leave, medical leave)?	22 (45)	28 (35)	29 (35)	13 (50)	Female faculty member may extend her contract up to two years in association with pregnancy and delivery and up to one year in case of adopting a child 6 years old or younger	"Where staff have had a career break, long term absence or other extenuating circumstances which impact on their output/performance, they are encouraged to provide this information including what impact such breaks/absences have had on them undertaking their role"
NA=not applicable						

NA=not applicable.

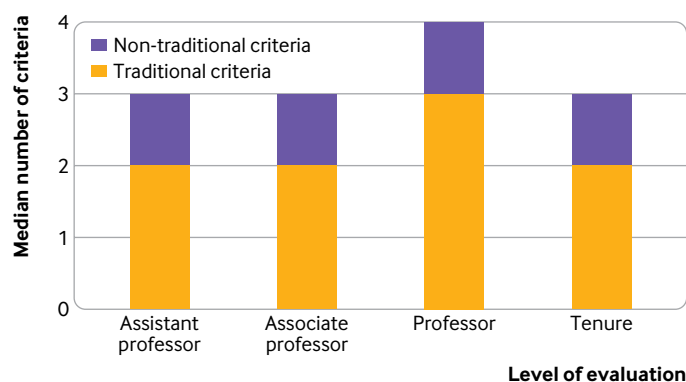


Fig 2 | Median number of criteria present by level of promotion

tenure (appendix 2; fig 1). For the other institutions, we could not find or access such guidelines either online or after communication (appendix 3). Of institutions with available guidelines, 39 (42%) were specific to faculties of biomedical sciences.

The universities that we could evaluate mostly had a very high development index rating (68/92; 74%), and they were almost equally split between Europe (n=27; 29%), Asia (29; 32%), and North America (28; 30%). Guidelines referred to were last updated between 1993 and 2018 (median 2016, interquartile range 2011-2017). On the basis of Leiden ranking of world universities, institution rankings ranged from 6 to 842 (median 345, interquartile range 157-549) (table 1). Of the 92 guidelines reviewed, evaluations for promotion to positions equivalent to assistant professor (n=49; 53%), associate professor (79; 86%), full professor (83; 90%), and tenure (26; 28%) were present, with most (82; 89%) institutions having guidelines available for more than one level of promotion. We found no statistically significant differences between institutions that did versus did not have criteria available for institution rankings ($P=0.14$) or human development index (χ^2 , df=2; n=93; $P=0.92$). We observed notable differences across continents on whether guidelines were accessible (Australia, 6/6 (100%); North America, 28/20 (97%); Europe, 27/54 (50%); Asia, 29/50 (58%); South America, 1/6 (17%)) (χ^2 , df=5; n=93; $P=0.001$) (table 1).

Presence of traditional and non-traditional criteria

The traditional criteria that were present most often were peer reviewed publications (any mention, 87/92 (95%); assistant professor, 39/49 (80%); associate professor, 76/79 (96%); professor, 79/83 (95%); tenure, 22/26 (85%)) and grant funding (any mention, 62/92 (67%); assistant professor, 26/49 (53%); associate professor, 50/79 (63%); professor, 56/83 (67%); tenure, 15/26 (58%)) (table 2). Authorship order (any mention, 34/92 (37%)), journal impact factor (any mention, 26/92 (28%)), and national or international reputation (any mention, 44/92 (48%)) were used less frequently. The exact mentions and how they were supposed to be operationalised varied across guidelines, and only a minority were quantitative. Thirty two (35%) institutions had at least one mention of a specific number of peer reviewed publications. The requirements for publications differed between institutions with some, for example, requiring a specific number per year or over the previous 10 years (from as few as one publication required in total to as many as 53 publications required in the previous 10 years). Institutions that required fewer publications often reported that publishing in journals with higher impact factors was necessary (for example, one publication in a journal with an impact factor of at least 10 or two publications in journals with impact factors of at least 5). Fourteen (15%) institutions had at least one mention of a specific amount of money for funding (range 300 000 RMB (41 766 USD) to 3 000 000 RMB). For authorship order, 24 (26%) of 92 institutions encouraged first author publications, 20 (22%) encouraged last or corresponding author publications (although many of these institutions also promoted first author publications), and four (4%) encouraged sole authored publications. No institutions mentioned that middle author publications or multi-authored papers were favourable. For journal impact, 11/26 (42%) institutions mentioned specific numbers for desirable impact factor metrics, but the desirable impact factor thresholds varied enormously across institutions (≥ 3 , 4, 5, 9, 10, 11, 20, 30; see appendix 4). No institutions had any numerical recommendations on assessment of national or international reputation.

Non-traditional criteria that were present included adjustments to expectations when professors go on

Table 3 | Multiple linear regression analyses of institutional characteristics and presence of traditional and non-traditional criteria among full professors (n=81)

Variables	Traditional criteria		Non-traditional criteria	
	Unstandardised β (95% CI)	P value	Unstandardised β (95% CI)	P value
Criteria level (reference=faculty level)	-0.60 (-1.21 to 3.63)	0.06	-0.08 (-0.48 to 0.31)	0.68
Leiden ranking*	-0.01 (-0.06 to 0.032)	0.88	0.01 (-0.476 to 0.026)	0.66
Very high human development index (reference=high) (n=61)	-0.63 (-1.45 to 1.9)	0.13	0.29 (-0.22 to 0.81)	0.26
Continent (reference=Asia):				
Australia (n=6)	1.78 (0.58 to 2.99)	0.004	1.07 (0.32 to 1.82)	0.006
Europe (n=22)	0.48 (-0.41 to 1.36)	0.29	0.32 (-0.23 to 0.86)	0.26
North America (n=27)	0.99 (0.13 to 1.84)	0.03	0.43 (-0.11 to 0.96)	0.12

*Increase of 50 points.

Table 4 | Percentage of institutional characteristics and presence of traditional and non-traditional criteria among full professors by region (n=83)

Continent	Mean percentage of traditional criteria	Mean percentage of non-traditional criteria
South America (n=1)	20	0
Australia (n=6)	70	21
Europe (n=22)	45	10
North America (n=27)	64	13
Asia (n=26)	49	3
Africa (n=1)	80	0

leave (any mention, 34/92 (37%); assistant professor, 22/49 (45%); associate professor, 28/79 (35%); professor, 29/83 (35%); tenure, 13/26 (50%)), citations of research (any mention 24/92 (26%); assistant professor, 12/49 (24%); associate professor, 23/79 (29%); professor, 23/83 (28%); tenure, 6/26 (23%)), and, rarely, alternative metrics for sharing research (any mention, 3/92 (3%); assistant professor, 3/49 (6%); associate professor, 3/79 (4%); professor, 2/83 (2%); tenure, 1/26 (4%)). Data sharing was mentioned only in one (1%) institution. Mentions of publishing in open access outlets, registering research, and adhering to reporting guidelines were absent from all institutions (table 2; fig 2). Non-traditional criteria were mostly qualitative. For citations, however, 25% (6/24) of institutions that included this item proposed specific numbers (see appendix 4).

Characteristics associated with presence of traditional and non-traditional items for professors

In tests of multicollinearity, independent variable tolerance values ranged from 0.4 to 0.9, and the variance inflation factors ranged from 1.1 to 2.8 for both traditional and non-traditional analyses, indicating that multicollinearity was not a major problem.¹⁷ We observed no deviation in the assumption of normality based on the inspection of the normal probability plots of the residuals or evidence of violations of assumptions of outliers, linearity, homoscedasticity, and independence of residuals on the basis of the standardised residual and scatter plot inspections.

Regressions for total scores

Institutions from Australia (unstandardised regression coefficient (β)=1.8 (standard error 0.61); $P=0.004$) and North America ($\beta=0.99$ (SE 0.43); $P=0.03$) tended to have a slightly larger number of traditional criteria present in guidelines compared with other continents (table 3). Australia had an average of 70% (mean 3.5 (SD 1.0) of 5 items) of traditional criteria present, and North America had an average of 64% (mean 3.2 (1.1) of 5 items). A significantly greater mean percentage of traditional items (54.2% (SD 0.24)) than non-traditional items (9.5% (0.11)) were reported among institutions (mean difference 44.8%, 95% confidence interval 39.6% to 50.0%; $P=0.001$). Institutions located in Australia ($\beta=1.07$ (SE 0.38); $P=0.006$) had modestly more non-traditional criteria in their guidelines (table 3). Institutions from Australia had an average of 1.5

(SD 0.5) non-traditional criteria present in guidelines. Table 4 shows mean percentages of traditional and progressive criteria present by continent.

Regressions for individual items

Six of the 12 items were present in 10% to 90% of promotion and tenure guidelines for professors, including grant funding, authorship order, impact factor, national or international reputation, citations, and adjustments to expectations. Encouragement of researchers to have a national or international reputation was significantly more present among institutions from North America than other continents ($\beta=27.44$, 95% confidence interval 3.26 to 231.16; $P=0.002$) (table 5).

Discussion

We found that guidelines for assessing faculty members for promotion and tenure among an international sample of 92 institutions with faculties of biomedicine or health sciences relied on traditional criteria. Almost all institutions' promotion criteria included the presence of peer reviewed publications, many of which also required a minimum number of papers published per year. Conversely, only about a third of institutions discussed citations and none referenced publishing in open access mediums, registering research, or adhering to reporting guidelines for transparently presenting research.

Substantial variability existed across continents as to whether any guidelines were available at all. This may be especially important to consider given that our work was based on a North American framework of career advancement, which affects the terms applied when searching for documents and the interpretation of criteria. Given the substantial rate of non-response from specific regions, we cannot exclude the possibility that such documents exist but could not be retrieved. For some universities, the criteria and related guidelines are not set at the level of the medical faculty or even the whole university but by a higher state authority (for example, the ministry of education in Greece) for all universities or government regulated laws (for example, employment leaves). Although the process for achieving promotion and tenure varies internationally, the concept of career advancement and the need for appropriate criteria are common to all regions. Availability of criteria is probably helpful for transparency; however, the availability of guideline documents does not mean that these are also faithfully adhered to. Criteria and rules may be bent in everyday academic practice. Assessing the adherence to guideline documents for career advancement through surveys or interviews may allow for an improved understanding of how to most meaningfully align the promotion and tenure criteria to best practices in research. This approach to data collection could also shed light on criteria that may be applied but are not stated in promotion and tenure documents.

Table 5 | Logistic regression analyses of institutional characteristics and presence of traditional and non-traditional criteria among full professors (n=81). Values are numbers (percentages) unless stated otherwise

Variables	Criteria present	Criteria not present	Adjusted odds ratio (95% CI)
Mention of authorship order	28 (34.6)	53 (65.4)	
Criteria level (reference=faculty level)	17 (50.0)	17 (50.0)	0.35 (0.10 to 1.22)
Institution level	11 (23.4)	36 (76.6)	
Median (IQR) Leiden ranking*	303 (56-613)	354 (182-560)	0.99 (0.89 to 1.10)
HDI (reference=very high)	18 (30.0)	42 (70.0)	
High HDI	10 (47.6)	11 (52.4)	1.53 (0.30 to 7.69)
Continent (reference=Asia)	12 (46.2)	14 (53.8)	
Australia	2 (33.3)	4 (66.7)	1.21 (0.12 to 12.61)
Europe	3 (13.6)	19 (86.4)	0.33 (0.05 to 2.22)
North America	11 (40.7)	16 (59.3)	0.73 (0.14 to 3.87)
Mention of journal impact factor	22 (27.2)	59 (72.8)	
Criteria level (reference=faculty level)	9 (26.5)	25 (73.5)	0.73 (0.18 to 2.86)
Institution level	13 (27.7)	34 (72.3)	
Median (IQR) Leiden ranking*	265 (96-544)	385 (117-567)	0.96 (0.85 to 1.08)
HDI (reference=very high)	11 (18.3)	49 (81.7)	
High HDI	11 (52.4)	10 (47.6)	3.35 (0.65 to 17.41)
Continent (reference=Asia)	12 (46.2)	14 (53.8)	
Australia	2 (33.3)	4 (66.7)	1.50 (0.13 to 16.87)
Europe	5 (22.7)	17 (77.3)	0.84 (0.13 to 5.27)
North America	3 (11.1)	24 (88.9)	0.30 (0.65 to 17.41)
Mention of grant funding	55 (67.9)	26 (32.1)	
Criteria level (reference=faculty level)	27 (79.4)	7 (20.6)	
Institution level	28 (59.6)	19 (40.4)	0.49 (0.13 to 1.83)
Median (IQR) Leiden ranking*	345.0 (87-580)	325 (187-505)	1.01 (0.90 to 1.14)
HDI (reference=very high)	43 (71.7)	17 (28.3)	
High HDI	12 (57.1)	9 (42.9)	1.9 (0.38 to 9.5)
Continent (reference=Asia)	13 (50.0)	13 (50.0)	
Australia	12 (54.5)	10 (45.4)	2.40 (0.44 to 13.10)
Europe	24 (88.9)	3 (11.1)	9.96 (1.52 to 65.31)
Mention of national or international recognition	38 (46.9)	43 (53.1)	
Criteria level (reference=faculty level)	20 (58.8)	14 (41.2)	0.49 (0.11 to 2.26)
Institution level	18 (38.3)	29 (61.7)	
Median (IQR) Leiden ranking*	365 (85-590)	321 (157-565)	1.01 (0.89 to 1.16)
HDI (reference=very high)	36 (60.0)	24 (40.0)	
High HDI	2 (9.5)	19 (90.5)	0.47 (0.05 to 4.14)
Continent (reference=Asia)	2 (7.7)	24 (92.3)	
Australia	5 (83.3)	1 (16.7)	52.36 (2.30 to 1190.95)
Europe	9 (40.9)	13 (59.1)	6.86 (0.71 to 66.85)
North America	22 (81.5)	5 (18.5)	27.44 (3.26 to 231.16)
Mention of citations	23 (28.4)	58 (71.6)	
Criteria level (reference=faculty level)	8 (23.5)	26 (76.5)	0.66 (0.16 to 2.75)
Institution level	15 (31.9)	32 (68.1)	
Median (IQR) Leiden ranking*	252 (94-580)	352 (156-566)	0.99 (0.88 to 1.11)
HDI (reference=very high)	19 (31.7)	41 (68.3)	
High HDI	4 (19.0)	17 (81.0)	1.64 (0.22 to 12.53)
Continent (reference=Asia)	4 (15.4)	22 (84.6)	
Australia	5 (83.3)	1 (16.7)	47.81 (2.25 to 1016.33)
Europe	8 (36.4)	14 (63.6)	5.21 (0.58 to 46.80)
North America	6 (22.2)	21 (77.8)	1.97 (0.24 to 16.33)
Mention of accommodations/adjustments to circumstances	29 (35.8)	52 (64.2)	
Criteria level (reference=faculty level)	15 (44.1)	19 (55.9)	0.93 (0.19 to 4.45)
Institution level	14 (29.8)	33 (70.2)	
Median (IQR) Leiden ranking*	400 (84-574)	325 (162-520)	1.00 (0.89 to 1.12)
HDI (reference=very high)	2 (7.7)	24 (92.3)	
High HDI	4 (66.7)	2 (33.3)	5.07 (0.45 to 57.79)
Continent (reference=Asia)	6 (27.3)	16 (72.7)	1.01 (0.15 to 6.95)
Australia	17 (63.0)	10 (37.0)	4.54 (0.62 to 33.44)

HDI=human development index; IQR=interquartile range.
*Increase of 50 points.

Implications of findings

An important barrier to the implementation of non-traditional criteria relates to the difficulty of selecting and integrating more appropriate measures,¹⁴ which

was described in several promotion documents reviewed. Institutions have noted the imperfections of traditional criteria, such as the impact factor, but reported that few alternatives exist.^{18 19} Integrating

non-traditional criteria to incentivise scientists requires evidence on the accuracy and the validity of non-traditional indicators,¹⁴ and such indicators are starting to emerge. As non-traditional metrics are available, implementing their use more widely—for example, through Declaration on Research Assessment (DORA)'s advisory board—may be one avenue to aid in the dissemination of more appropriate tools for assessing scientists.

Institutions that rely on traditional metrics, such as number of publications and associated journal impact factors, may misinterpret what these metrics mean.¹⁹ Beyond evidence, other reasons exist to consider alternative criteria. They may better align with a university's mission, for example. Similarly, some criteria, such as data sharing, have a high research integrity value; patients support sharing of their data,²⁰ and it facilitates assessments of reproducibility. To facilitate data sharing, the FAIR (Findability, Accessibility, Interoperability and Reusability) principles will probably need to be in place.⁹ An additional barrier to including non-traditional criteria in evaluations is the need for resources to support this change. The institution in our sample that had the greatest number of non-traditional criteria, Ghent University, described having invested in an online system to help in assessing some non-traditional metrics of their researchers.¹⁸ Decreasing the barriers to using non-traditional metrics in evaluations will be necessary for systematic changes to occur.

Limitations of study

Some limitations should be considered when interpreting our study results. Direct involvement of patients and the public was absent from this review. Incorporating the perspectives of patients and the public in future research of promotion and tenure criteria can incentivise research practices that better align with the needs and expectations of society. This could also allow for international differences to be highlighted by speaking with stakeholder groups in various regions. Next, although we searched websites and contacted institutions, not all institutions use pre-specified criteria for assessing promotion, and in some instances we did not find documents. This resulted in only a subset of the intended sample being available for review and included in our analyses. South America and Africa were underrepresented in our sample, so we can draw few conclusions about the criteria of institutions in these regions. An additional limitation is that incentives for professors can occur through other pathways, such as financial bonuses, which may not be publicly available or included in the documents reviewed. Obtaining a more complete understanding of the criteria used for providing financial and reputational incentives in medical faculties may require review of internal documentation on bonuses and awards or recognitions in addition to formal promotions. Furthermore, medical faculties often take into account clinical work and teaching, which we did not include.

Finally, we should acknowledge that for both traditional and non-traditional criteria, the exact way they are proposed and operationalised can make a difference to whether they might have a positive or negative effect on research quality. With a plethora of metrics being developed for non-traditional criteria, some of them may be much better than others. For example, although citations may be a more accurate representation of one's research impact than journal impact factor, considering the number of citations in isolation from the field of research may not motivate those who work in otherwise important research fields that have low citation density (for example, because few other scientists work in them).

Conclusions and policy implications

Integrating appropriately framed criteria that encourage best practice in research could result in improvements in medical research and evidence based medicine. Systematic changes require collaborative efforts and creativity to overcome barriers to developing and adopting the best metrics. Considering the benefits of creating sustainable changes to the criteria that drive poor medical research internationally, however, would be a turning point in facilitating the transparency, openness, and reproducibility in research practices.

AUTHOR AFFILIATIONS

¹Department of Psychology, McGill University, Montreal, QC, Canada

²Ottawa Hospital Research Institute, Ottawa, ON, Canada

³Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada

⁴Department of Medicine, Stanford University, Stanford, CA, USA

⁵Department of Health Research and Policy, Stanford University, Stanford, CA, USA

⁶Department of Biomedical Data Science, and Statistics, Stanford University, Stanford, CA, USA

⁷Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

⁸Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

⁹School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

We thank Juan Pablo Alperin for providing us with access to the database that included documents for review from North American universities, Becky Skidmore for her recommended search terms, Nimesh Acharya Chander for his help with table formatting, and Tim Ramsay for his guidance with analyses. We also thank all the people who searched websites and translated documents, including Xiaojin Wang, Andrea Carboni-Jiménez, Michal Dedys, Fatemeh Yazdi, Philipp-Clemens Nowotny, Song Xiaoyang, Yan Jin, Kednapa Thavorn, Chen He, Alexander Tsertsvadze, and Francesca Ruggiero. Finally, we acknowledge each of the institutions that responded to our requests for promotion and tenure documents.

Contributors: DM and JPAI had the idea for the study. DBR and DM wrote the study protocol and the initial draft of the article. DM, DBR, and HR agreed on the criteria applied for promotion and tenure after pilot testing a set of university criteria. DBR and HR found institution guidelines and extracted information on promotion criteria. All authors were involved in subsequent protocol revisions. DBR did the statistical analyses. All authors provided critical feedback and approved the final version of the paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. DM is the guarantor.

Funding: No funding was received for this work. DBR is funded through a Canadian Institutes of Health Research Vanier graduate scholarship. DM is funded by a university research chair. METRICS is funded by a grant from the Laura and John Arnold Foundation. Funding sources did not have any role in the design and conduct

of the study, data collection and analysis, interpretation of study findings, or the decision to publish.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: As this study did not involve any human data, ethics approval was not required.

Data sharing: All data associated with this study are posted on the open science framework (https://osf.io/26ucp/?view_only=b80d2bc7416543639f577c1b8f756e44). The study protocol, data extraction forms, and data are also available at this link.

The manuscript's guarantor affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: This work did not involve study participants. This work will be shared through conference presentations and shared with relevant stakeholder groups and initiatives such as the Declaration on Research Assessment (DORA).

A preprint of this manuscript has been deposited at <https://www.biorxiv.org/content/10.1101/802850v1>.

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/>.

- 1 Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;330:1053. doi:10.1136/bmj.38414.515938.8F
- 2 Ioannidis JPA. Acknowledging and overcoming nonreproducibility in basic and preclinical research. *JAMA* 2017;317:1019-20. doi:10.1001/jama.2017.0549
- 3 Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma G, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097. doi:10.1371/journal.pmed.1000097
- 4 Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;8:18. doi:10.1186/1741-7015-8-18
- 5 Moher D, Glasziou P, Chalmers I, et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet* 2016;387:1573-86. doi:10.1016/S0140-6736(15)00307-4
- 6 Wolff C, Rod AB, Schonfeld RC. UK Survey of Academics 2015. 2016. <https://sr.ithaka.org/publications/uk-survey-of-academics-2015/>.
- 7 Hammarfelt B. Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics and history. *Aslib J Inf Manag* 2017;69:607-23. doi:10.1108/AJIM-01-2017-0006
- 8 Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166-75. doi:10.1016/S0140-6736(13)62227-8
- 9 Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18
- 10 Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257-66. doi:10.1016/S0140-6736(13)62296-5
- 11 Moher D, Goodman SN, Ioannidis JP. Academic criteria for appointment, promotion and rewards in medical research: where's the evidence? *Eur J Clin Invest* 2016;46:383-5. doi:10.1111/eci.12612
- 12 McKiernan EC. Imagining the "open" university: Sharing scholarship to improve research and education. *PLoS Biol* 2017;15:e1002614. doi:10.1371/journal.pbio.1002614
- 13 Flier J. Faculty promotion must assess reproducibility. *Nature* 2017;549:133. doi:10.1038/549133a
- 14 Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. *PLoS Biol* 2018;16:e2004089. doi:10.1371/journal.pbio.2004089
- 15 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:e296. doi:10.1371/journal.pmed.0040296
- 16 Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav* 2018;2:6-10. doi:10.1038/s41562-017-0189-z
- 17 Tabachnick BG, Fidell LS. *Using multivariate statistics*. 6th ed. Pearson, 2013.
- 18 Ghent University. Principles for the evaluation of research 2017. <https://www.ugent.be/en/research/research-evaluation.htm>.
- 19 McKiernan EC, Schimanski LA, Muñoz Nieves C, Matthias L, Niles MT, Alperin JP. Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *Elife* 2019;8:e47338. doi:10.7554/eLife.47338
- 20 Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *N Engl J Med* 2018;378:2202-11. doi:10.1056/NEJMsa1713258

Web appendix: Appendix 1-4