Appropriate and inappropriate uses of quantitative metrics in research assessment

A guidance document

UEFISCDI 2/27/26

Funded by the European Union

Publication details

Deliverable 2.2. Guide regarding the appropriate and inappropriate use of quantitative metrics in research assessment

UEFISCDI – The Executive Agency for Higher Education, Research, Development and Innovation Funding

Published: February 2026

DOI: 10.5281/zenodo.19206635

Authors: UEFISCDI, Ioana Spanache, Ioana Trif

*This guide was developed by **UEFISCDI** within the project **ARIA** – Advancing Responsible and Inclusive Assessment at UEFISCDI. The project is funded under the Horizon Europe programme through the **CoARA Boost project** Cascade Funding Call – Second Round. The ARIA project supports the advancement of responsible and inclusive research assessment practices, in alignment with the principles of the Agreement on Reforming Research Assessment.*

Contents

1. Introduction

Deliverable D2.2 Guide regarding the appropriate and inappropriate use of quantitative metrics in research assessment was developed within Activity 2 Updating reviewer guidelines of the ARIA project - Advancing Responsible and Inclusive Assessment at UEFISCDI, funded through the Horizon Europe program - CoARA Boost Cascade Funding Call 2.

The ARIA project, implemented by UEFISCDI from 1 September 2025 to 31 August 2026, has as its main objective the acceleration of CoARA principles adoption in internal research evaluation processes. It represents a concrete step in the implementation of UEFISCDI's Action Plan for Research Assessment Reform, published on 10 February 2025.

2. Scope of the guide

Scope

This guide provides practical guidance on the **responsible, appropriate, and inappropriate uses of quantitative metrics** in research assessment, with a particular focus on **evaluation of research project proposals**, including the assessment of:

- the scientific quality of the proposal;
- the feasibility and credibility of the work plan;
- the prior experience, expertise, and relevant contributions of the applicants and team members, where these are part of the evaluation criteria.

The guide does **not** aim to eliminate the use of quantitative information. Instead, it clarifies **how metrics may be used responsibly**, **when they are misleading**, and **how they should be interpreted alongside expert judgement**.

Who is this guide for?

- External evaluators and reviewers
- Panel members, including rapporteurs and panel chairs
- Staff of research funding and performing organisations
- Applicants and their research teams, to improve transparency and mutual understanding

The guide is relevant across disciplines and funding instruments, and is intended to support fairness, consistency, and quality of expert evaluation across the RDI ecosystem.

3. What quantitative metrics are (and are not)

Metrics are not microscopes for individual quality; they are telescopes for systemic patterns.

What they *are*

- Proxies that **describe certain aspects** of research activity, dissemination or visibility
- Tools that support **pattern detection**, contextualisation, benchmarking at scale, and monitoring over time

What they are *not*

Quantitative metrics are **not** direct measures of:

- research quality *per se*,
- originality or innovativeness,
- societal or policy relevance,
- individual researcher merit,
- future performance or project success.

They cannot capture:

- methodological soundness,
- conceptual ambition,
- ethical robustness,
- appropriateness of the research design,
- relevance to the specific objectives of a call.

Quantitative metrics are proxies, meaning they provide indirect information about certain observable aspects of research activity and visibility, such as publication, dissemination, collaboration, or scholarly attention. They do not measure research quality or merit directly but capture patterns that emerge from aggregated behaviour within research systems. For this reason, metrics are most useful when applied at scale: to detect trends, enable contextual benchmarking, and monitor developments over time. When used at the level of individual researchers or proposals, their meaning becomes fragile and easily distorted.

Example

Citation counts reflect use and attention, not scientific validity or importance. In biomedical research, several retracted papers continue to be cited long after retraction, and most post-retraction citation contexts do not acknowledge that the paper has been retracted, illustrating that citations counts do not equate to scientific validity or quality. In a database-wide study of 7 813 retracted biomedical papers with 169 434 citations, only 5.4 % of post-retraction citation contexts explicitly noted the retraction, and the way retracted papers were cited did not change after retraction (Hsiao & Schneider, 2022).

This phenomenon is well documented in the broader scientometric literature. Studies of retracted publications show that articles continue to be cited after retraction, and most post-retraction citations do not mention that the paper has been retracted (Budd, Sievert, & Schultz, 1998; Bar-Ilan & Halevi, 2017; Shuai, Rollins, & Bollen, 2017).

Examples of types of frequently encountered quantitative metrics

The following list is not exhaustive, but reflects some of the most common metrics encountered:

- Number of publications
- Total citation counts
- h-index
- Journal Impact Factor (JIF) of publication venues
- Field-normalised citation indicators (e.g. MNCS)
- Share of publications in indexed databases (e.g. Web of Science, Scopus)
- Grant income or number of funded projects
- Altmetrics (mentions in social media, policy documents, news outlets)
- Collaboration indicators (co-authorship, international co-authorship)

Each of these metrics has legitimate but limited uses, and well-documented risks of misuse.

4. General principles for responsible use of metrics

1. Metrics can support qualitative assessment, not replace it

Quantitative metrics may inform expert judgement, but must never substitute for it. Metrics can only provide contextual background (Hicks et al., 2015; Wilsdon et al., 2015).

Example of appropriate use:

In the UK **Research Excellence Framework (REF)**, quantitative indicators may be included in the impact case studies submitted by Units of Assessment (UoAs) as part of the evidence base used by expert panels during the peer review process. However, the evaluation of impact-defined through the two main criteria of **reach** and **significance**-remains fundamentally grounded in expert judgement rather than metric-based scoring.

REF 2021 guidelines explicitly state that institutions should select indicators and supporting evidence that are most appropriate for demonstrating the claimed impact. A broad range of evidence types may therefore be used, including qualitative testimonies, quantitative indicators, and tangible or material evidence. The framework explicitly encourages the use of **diverse forms of evidence**, reflecting the varied nature of societal impact. Policy discussions and evaluations of the REF consistently emphasise that metrics should **inform but not replace qualitative expert assessment**, reinforcing the principle that quantitative indicators serve as contextual support for peer review rather than as a substitute for it[1].

2. Metrics must be interpreted in relation to the evaluation purpose

Metrics should be used only if they are relevant to the specific evaluation question. In project evaluations, the purpose is not to rank researchers, but to **assess whether the research team has the relevant expertise and capacity to carry out the proposed work**, and whether the proposal demonstrates things such as clear scientific motivation, novelty, and methodological soundness; realistic objectives and timelines; appropriate use of resources; and credible potential for scientific and, where relevant, broader impact, while representing a well-justified use of public funding.

Example:

In evaluations of funding applications by the Research Council of Finland, review panels focus primarily on the quality of the research plan and the applicant's competence,

[1] REF 2021. Index of revisions to the 'Panel criteria and working methods' (2019/02). https://2021.ref.ac.uk/media/1450/ref-2019_02-panel-criteria-and-working-methods.pdf

rather than on quantitative indicators such as citation or journal-based metrics. This reflects the purpose of project evaluation, which is to assess the proposed research and the capability of the applicant to carry it out, rather than to rank researchers based on bibliometric indicators.

In line with this approach, the use of journal-based metrics (e.g., Journal Impact Factors (JIF) or JUFO classifications) is explicitly prohibited in the review process. Citation metrics may be considered only as supporting information, and their use must be responsible and mindful of their limitations, particularly because citation practices vary across disciplines and are not reliably comparable in multidisciplinary review panels (Research Council of Finland, 2023)[2].

3. Field and career-stage differences must be explicitly accounted for

Publication, citation, and dissemination practices can vary substantially across disciplines, sub-disciplines, publication cultures, and career stages, and these differences must be explicitly considered in research assessment. What constitutes a typical publication output, citation trajectory, or authorship pattern in one field may be unusual or even impossible in another, and expectations also differ significantly between early-career and senior researchers.

Example:

For example, a senior humanities researcher and an early-career biomedical researcher cannot be meaningfully compared using raw publication or citation counts, as such comparisons ignore structural differences in publishing rhythms, output formats, and citation dynamics. When these contextual factors are overlooked, the use of quantitative metrics introduces systematic bias rather than supporting fair and informed evaluation.

Another concrete example of accounting for career-stage differences is the FRIPRO funding scheme administered by the Research Council of Norway. The scheme includes separate open-ended calls for early-career and experienced researchers, with differentiated track record requirements. Early-career applicants submit a CV including an “early achievements track record” with up to five publications, while experienced researchers submit a CV with a track record of up to ten publications. By setting different expectations for publication output according to career stage, the scheme avoids direct comparison based on identical quantitative thresholds and supports proportionate evaluation[3].

[2] Research Council of Finland. Responsible researcher evaluation. https://www.aka.fi/en/from-research-to-society/responsible-science/responsible-researcher-evaluation/
[3] FRIPRO financial scheme. https://www.forskningsradet.no/en/call-for-proposals/2023/researcher-project-early-career-fripro/

4. Avoid misplaced concreteness and false precision

Numerical values can create a **false sense of accuracy or precision,** giving the impression that small numerical differences reflect meaningful distinctions in research performance or quality. At the level of individual researchers or proposals, minor variations in indicators such as h-index values, citation counts, or publication totals can be frequently driven by chance, database coverage, or disciplinary conventions rather than by substantive differences in merit.

Example:

Distinguishing between applicants on the basis of an h-index of 11 versus 13 is not scientifically justified and should not influence funding decisions. When numerical values are treated as exact or decisive, rather than as rough contextual signals, they risk distorting expert judgement instead of supporting it.

5. Appropriate uses of quantitative metrics (with real examples)

4.1. Portfolio-level and aggregate analysis

Appropriate use

- Monitoring trends at:
  - Institutional level
  - Faculty or department level
- Used *exploratorily*, not mechanically

Quantitative metrics are appropriately used at **aggregate levels**, such as institutions, departments, or large research portfolios, to identify **general trends and patterns** rather than to assess individual researchers or proposals. At this level, metrics help highlight areas of strength, emerging fields, or gaps that may require further qualitative analysis. Typical examples include publication volumes by field, field-normalised citation indicators, open access shares, or collaboration rates. Universities in the Netherlands and Nordic countries, for instance, use such aggregated, field-normalised data to support internal strategic reflection and planning. Importantly, these analyses are used to **inform discussion**, not to mechanically determine funding decisions or replace expert judgement in project evaluation.

4.2. Field-Normalised Comparisons

Appropriate use

- Comparing like with like:
  - Same discipline
  - Same career stage
  - comparable publication cultures.

When quantitative metrics are used for comparison, they should be applied **only to comparable cases**, taking into account differences between disciplines, career stages, and publication cultures. Field-normalised indicators adjust for some of these structural differences and can therefore support **high-level, contextual comparisons**, for example within the same discipline or among researchers at similar career stages.

The Mean Normalised Citation Score (MNCS) is defined as an indicator of the average number of citations of a university's publications, normalised for differences in scientific field, publication year, and document type (Waltman et al., 2012).MNCS is one of the indicators used in the CWTS Leiden Ranking (Waltman et al., 2012). However, even when normalised, such metrics are **not suitable for ranking individual applicants** or for making fine-grained distinctions between proposals in competitive project evaluations, where expert judgement remains essential.

4.3. Metrics as input to Peer Review

Appropriate use

- Metrics as **contextual background information** for panels
- Expert reviewers remain fully responsible for final judgement.

Quantitative metrics may be used as **contextual background information** to support peer review, for example to help situate a proposal or a track record within a broader research landscape. In this role, metrics can inform discussion, especially where expert opinions differ or additional context is useful, but they must **never function as automatic scoring tools**. A well-documented example is the **UK Research Excellence Framework (REF)**, where bibliometric indicators are used selectively and only in some fields (e.g., Clinical Medicine, Public Health, Biology, Chemistry, Physics, Computer Science, Economics) to provide contextual information for expert panels. The REF 2021 Guidance on Submissions states that some sub-panels may consider citation counts as additional information about the academic significance of outputs. However, outputs are primarily assessed through expert review against the criteria of originality, significance, and rigor, and reviewers remain responsible for the final judgement.

The REF also emphasises that citation data are not used automatically or uniformly across panels: each panel specifies in its Panel Criteria whether and how such data are considered. Where used, citation indicators must be interpreted cautiously, recognising limitations such as disciplinary differences in citation practices, the lower relevance of citations for recently published work, and language-related biases. This illustrates how metrics may **inform peer review without replacing expert judgement**.

4.4. Monitoring System-Level Effects

Appropriate use

- Evaluating the effects of policies or funding instruments over time.

Quantitative metrics are particularly well suited for **monitoring the effects of policies or funding instruments over time**, where the focus is on understanding changes at the level of the research system rather than assessing individual performance. In this context, metrics are used to track trends across portfolios or programmes and to support evidence-informed policy learning.

For example, funding organisations often analyse **collaboration indicators**, such as rates of international co-authorship or participation of multiple institutions in funded projects, to assess whether specific funding schemes are effectively encouraging collaboration or interdisciplinarity. Importantly, these indicators are used to evaluate **system-level effects**, without being linked directly to individual rewards, sanctions, or proposal-level decisions.

6. Inappropriate uses of quantitative metrics

5.1. Journal-Based Metrics for Individual Assessment

Inappropriate

- Using Journal Impact Factor (JIF) to:
  - evaluate individual articles
  - judge research proposal quality
  - assess researcher merit
  - make hiring or promotion decisions

The JIF reflects the **average citation rate of a journal over a limited time window**, not the quality, originality, or relevance of any specific article published within it. Using

journal prestige as a proxy for scientific merit **risks confusing venue reputation with research quality** and may distort evaluation by privileging publication strategy over substantive contribution. This practice is explicitly criticised by DORA, which emphasises that journal-level metrics should not be used to evaluate individual researchers or research outputs. Overreliance on such indicators can incentivise prestige-driven behaviour rather than rigorous, innovative, or socially relevant research, and is therefore incompatible with fair and responsible project evaluation.

Illustrative example – misuse

During panel discussions, reviewers note that one applicant has published several papers in high-impact-factor journals, while another has published in more specialised venues. Although the proposal from the second applicant is judged to be methodologically stronger and more innovative, comments such as "they haven't published in top journals" begin to influence the tone of the discussion. The journal reputation becomes an implicit proxy for quality, subtly affecting the final scoring despite not being part of the formal evaluation criteria.

Good practice

In contrast, a review panel focuses on the **content and contribution** of previous work rather than the prestige of the publication venue. When journal information is available, it is not treated as evidence of quality but as background context. Reviewers assess the scientific merit of the proposal independently, considering the applicant's expertise in relation to the project's objectives, field norms, and career stage. The emphasis remains on substance rather than on journal branding.

5.2. Mechanical thresholds and cut-offs

Inappropriate

- Fixed numerical thresholds:
    - "minimum X publications"
    - "at least Y citations"

The use of fixed numerical thresholds—such as requiring a "minimum number of publications" or "at least a certain number of citations"—as eligibility criteria or decisive evaluation filters is inappropriate in project proposal assessment. Such thresholds reduce complex academic trajectories to single quantitative cut-offs and risk excluding applicants before their ideas, methodology, or potential contribution are properly examined (e.g. eligibility standards). They also fail to account for differences in discipline, career stage, non-linear career paths, or the nature of interdisciplinary research. Funding

systems that have relied on strict publication-count thresholds have been criticised in research assessment reform literature for disproportionately disadvantaging early-career researchers and those working across fields. Mechanical cut-offs may create an appearance of objectivity and efficiency, but they undermine fair and context-sensitive evaluation. **Initiatives such as the** *San Francisco Declaration on Research Assessment (DORA)* **explicitly recommend that funding agencies highlight, especially for early-stage investigators, that the scientific content of a paper is more important than publication metrics or the identity of the journal in which it was published, discouraging the use of simplistic quantitative measures such as publication counts or journal-based metrics in funding decisions.**

Illustrative example - misuse

In some funding schemes, eligibility rules require applicants to demonstrate a minimum number of publications within a defined time window (e.g. the last five years). An applicant returning from parental leave or transitioning from industry into academia may not meet the numerical threshold, despite proposing a highly innovative and feasible project aligned with the call objectives. The application is declared ineligible before peer review. The rule is applied consistently, yet its impact is structurally exclusionary.

Good practice

In contrast, a funding programme may define expectations regarding research experience but allow flexibility in how these are demonstrated. Instead of rigid publication thresholds, evaluators are instructed to assess applicants' expertise in relation to their **career stage, time available for research, field-specific publication practices, and any documented career interruptions**. Where quantitative indicators are considered, they are interpreted contextually rather than used as automatic eligibility filters. The focus remains on the scientific merit and feasibility of the proposal, supported by a proportionate and context-aware assessment of the applicant's experience.

5.3. Cross-disciplinary comparisons without normalisation

Inappropriate

- Comparing:
    - Humanities vs biomedical sciences
    - Applied engineering vs theoretical physics

Comparing researchers or proposals from different disciplines using raw quantitative metrics is inappropriate, as publication and citation practices vary significantly across fields. Direct comparisons between, for example, humanities and biomedical sciences, or between applied and theoretical domains, ignore structural differences in output formats, citation dynamics, and dissemination cultures. Humanities researchers are systematically disadvantaged by raw citation counts due to monograph-based publishing traditions, slower citation cycles, and the frequent use of local or national languages. Without proper contextualisation or normalisation, cross-disciplinary metric comparisons introduce bias rather than fairness and risk distorting evaluation outcomes.

Illustrative example – misuse

In a multidisciplinary panel, reviewers notice that applicants from one field tend to have significantly higher citation counts and publication numbers than those from another. Although no formal ranking is imposed, these differences gradually influence the panel discussion, with higher numerical indicators implicitly interpreted as signals of stronger scientific standing. Over time, proposals from fields with lower publication and citation densities receive comparatively less enthusiasm, not because of lower proposal quality, but because disciplinary differences in metric patterns are not explicitly addressed. The resulting bias is subtle and unintentional, yet structurally disadvantageous.

Good practice

In a similar multidisciplinary call, evaluators first assess the quality, originality, and feasibility of each proposal within its disciplinary context. When reviewing applicants' experience, they consider field norms, publication formats, and career stage before interpreting any quantitative indicators. Metrics are used only to provide background context within comparable domains, and cross-field numerical comparisons are avoided. Expert judgement, grounded in disciplinary understanding, guides the final decision.

5.4. Metrics as Proxies for Societal Impact

Inappropriate

- Assuming:
  - citations equal societal relevance,
  - altmetrics equal impact.

Using quantitative metrics as direct evidence of societal or policy impact is inappropriate. **Citation counts reflect attention within the scholarly literature** and

cannot be equated with societal relevance, while altmetrics-such as social media mentions, media coverage, or online downloads-primarily capture visibility, topicality, or public interest. These indicators do not demonstrate that research has produced meaningful, sustained, or verifiable impact beyond academia. For example, high levels of social media attention may reflect visibility or *"unknown attention"* rather than substantive influence on policy, practice, or societal outcomes. In an empirical test using UK Research Excellence Framework data, altmetric indicators (including social media mentions) showed weak or near-zero correlation with expert peer assessments of societal impact, suggesting they capture a different dimension of attention not directly tied to meaningful societal influence (Bornmann, Haunschild & Adams, 2019).

Illustrative example – misuse

In a competitive call that includes an "expected societal impact" criterion, reviewers note that one applicant's previous publications received substantial media attention and high Altmetric scores. During discussion, these visibility indicators are implicitly treated as evidence of strong societal impact. However, there is no documented uptake of the research in policy decisions, professional guidelines, or practical implementation. The high attention reflects topical interest rather than demonstrable societal change, yet it subtly influences the panel's perception of impact potential.

Good practice

In a similar evaluation, reviewers distinguish clearly between visibility and impact. They examine qualitative evidence of societal engagement-such as documented collaboration with stakeholders, policy briefings, adoption of results in practice, or letters of support-alongside a realistic and well-articulated impact pathway in the proposal. Quantitative indicators of attention are considered contextual background information, but conclusions about impact are based on credible mechanisms of change and substantiated evidence rather than on visibility metrics alone.

5.5. Individual-Level Ranking and League Tables

Inappropriate

- Ranking individual researchers by:
  - h-index
  - total citations
  - composite scores

Ranking individual applicants using quantitative indicators—such as h-index values, total citation counts, or composite scores (for example, weighted formulas combining publication counts, citation impact, journal metrics, grant income, or other indicators into a single numerical score)—is inappropriate in project proposal evaluation. Such rankings create a misleading appearance of objectivity and precision while reducing complex research careers to simplified numerical hierarchies. Indicators like the h-index are particularly problematic, as they systematically disadvantage early-career researchers, researchers with career breaks, and those working across disciplines where publication and citation patterns differ.

Illustrative example - misuse

A funding scheme automatically ranks applicants using a composite score that combines publication counts, total citations, and previous grant income. As a result, an early-career researcher with an excellent proposal is placed far below senior applicants simply because they have had fewer years to accumulate metrics. The ranking appears objective but masks structural bias and sidelines expert judgement.

Good practice

A review panel receives metric information but only to use it at contextual background Reviewers first assess the proposal's ambition, methodology, feasibility, and expected impact, and afterwards, they look at the applicant's experience in relation to their career stage and field norms. Metrics inform the discussion, but expert judgement drives the decision.

7. Typical misuse scenarios and why they persist

Problematic uses of quantitative metrics often persist not because evaluators intend to be unfair, but because structural and organisational pressures make numerical indicators appear attractive.

Common drivers include:

Pressure for efficiency and scalability

Large funding calls and high application volumes create strong pressure for rapid, standardised screening. In such contexts, numerical indicators can appear to offer a practical and time-efficient way to differentiate applicants and manage workload. Metrics provide an impression of consistency and comparability across cases. However, when used as shortcuts, they risk replacing careful qualitative judgement with simplified numerical filters, especially under time constraints.

Perceived objectivity of numbers

Numbers are often viewed as neutral and impartial. In practice, however, they embed disciplinary norms, database coverage biases, and cumulative advantage effects. The apparent precision of metrics can obscure their contextual limitations.

Legacy systems and incentive structures

In several national performance-based funding models, heavy reliance on publication counts or journal metrics initially shaped institutional incentives and behaviour. Over time, concerns about gaming, salami-slicing of outputs, homogenisation of research strategies, and reduced diversity of outputs led to reforms and renewed emphasis on qualitative evaluation. These experiences demonstrate how metric-driven systems can unintentionally reshape research practices.

Path dependency and cultural habits

Even when formal rules change, established evaluation practices and habits often persist. If evaluators have long relied on publication counts or journal metrics as shorthand indicators of quality, these references may continue to shape discussions informally, even after official guidance discourages their use. Templates, reporting formats, and shared expectations within panels can reinforce this continuity. As a result, reform on paper does not automatically translate into change in practice

8. Decision Matrix: When a Metric Is Appropriate

Quantitative metrics can support evaluation, but only under certain conditions. The following **decision matrix is designed to help evaluators quickly determine whether referring to a metric is appropriate in a given situation**.

Before using a metric in your assessment, consider the following questions:

Step 1: What is the unit of assessment?

Question	If YES	If NO
Am I assessing an aggregated portfolio (e.g. institution, programme, department)?	Metrics are more appropriate as descriptive tools.	Greater caution required.
Am I assessing an individual applicant or proposal?	Use only as contextual background, if at all.	Do not use for ranking or decisive filtering.

Key principle: The smaller the unit (individual article, applicant, proposal), the weaker the reliability of metrics.

Step 2: Is the comparison contextually valid?

Question	If YES	If NO
Are the cases from the same discipline or sub-discipline?	Comparisons may be more meaningful.	Avoid raw comparisons.
Are career stages comparable?	Contextual interpretation possible.	Adjust expectations.
Are publication cultures similar?	Metrics may provide limited context.	Do not rely on raw numbers.

Key principle: Only consider like in relation with like.

Step 3: What role is the metric playing?

Question	If YES	If NO
Is the metric used as background context?	Acceptable with caution.	Risk of misuse.
Would my judgement remain the same without the metric?	Metric is likely supplementary.	Metric may be driving the decision.
Can I justify my score based on qualitative reasoning alone?	Appropriate use.	Reconsider reliance on metric.

Key principle: Metrics may inform discussion; they must not determine outcomes.

Step 4: Is the metric relevant to the evaluation criterion?

Ask yourself:

- Does this metric directly relate to the criterion being assessed (e.g. feasibility, expertise, impact pathway)?
- Or am I using it as a shortcut to infer quality or merit?

For example:

- Citation counts do not assess methodological soundness.
- Journal Impact Factors do not assess proposal originality.

- Altmetrics do not demonstrate societal impact.

If the metric does not clearly align with the criterion, it should not influence scoring.

Quick Self-Check for Evaluators

Before referencing a metric in discussion or in your assessment, ask:

1. Am I using this number because it is easy, or because it is truly relevant?
2. Have I considered field norms and career stage?
3. Would I defend this reasoning publicly as fair and consistent?
4. Does this metric help explain my judgement, or is it replacing it?

If uncertainty remains, prioritise qualitative expert judgement.

Core Rule

Metrics may provide context. They must not replace expert assessment of quality, originality, feasibility, and impact.

References

- CoARA - Arentoft, M., Berghmans, S., Borrell-Damian, L., Bottaro, S., Faure, J.-E., Gaillard, V., Glinos, K., Albacete, J. L., Morais, R., Morris, J., Schiltz, M., & Stroobants, K. (2022). Agreement on Reforming Research Assessment. Zenodo. https://doi.org/10.5281/zenodo.13480728
- San Francisco Declaration on Research Assessment (DORA). https://sfdora.org/read/
- Hicks, D., Wouters, P., Waltman, L. et al. Bibliometrics: The Leiden Manifesto for research metrics. Nature 520, 429–431 (2015). https://doi.org/10.1038/520429a
- Hsiao TK, Schneider J. Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. Quant Sci Stud. 2022 Feb 4;2(4):1144-1169. doi: 10.1162/qss_a_00155. PMID: 36186715; PMCID: PMC9520488. Budd JM, Sievert M, Schultz TR. Phenomena of Retraction: Reasons for Retraction and Citations to the Publications. *JAMA*. 1998;280(3):296–297. doi:10.1001/jama.280.3.296
- Bar-Ilan, J., Halevi, G. Post retraction citations in context: a case study. Scientometrics 113, 547–565 (2017). https://doi.org/10.1007/s11192-017-2242-0
- Shuai, X. et. al (2017). *A multidimensional investigation of the effects of publication retraction on scholarly impact. Journal of the Association for Information Science and Technology, 68*(9), 2225–2236. https://doi.org/10.1002/asi.23826Wilsdon, James & Allen, Liz & Belfiore, Eleonora & Campbell, Philip & Curry, Stephen & Hill, Steven & Jones, Richard & Kain, Roger & Kerridge, Simon & Thelwall, Mike & Tinkler, Jane & Viney, Ian & Wouters, Paul & Hill, Jude & Johnson, Ben. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. 10.13140/RG.2.1.4929.1363.
- Research Council of Finland. Responsible researcher evaluation. https://www.aka.fi/en/from-research-to-society/responsible-science/responsible-researcher-evaluation/
- FRIPRO financial scheme. https://www.forskningsradet.no/en/call-for-proposals/2023/researcher-project-early-career-fripro/
- REF 2021. Index of revisions to the 'Panel criteria and working methods' (2019/02). https://2021.ref.ac.uk/media/1450/ref-2019_02-panel-criteria-and-working-methods.pdf
- REF 2021. Citation and contextual data guidance. https://2021.ref.ac.uk/guidance/citation-and-contextual-data-guidance/index.html

- REF 2021. Guidance on submissions. https://www.stmarys.ac.uk/research/docs/ref/2019-nov-ref-guidance-on-submissions.pdf
- Lutz Bornmann, Robin Haunschild, Jonathan Adams, Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF), Journal of Informetrics, Volume 13, Issue 1, 2019, Pages 325-340, ISSN 1751-1577, https://doi.org/10.1016/j.joi.2019.01.008.