Assessing Value in Biomedical Research
The PQRST of Appraisal and Reward

John P. A. Ioannidis, MD, DSc
Departments of Medicine and Health Research and Policy, Stanford University School of Medicine, Palo Alto, California, Department of Statistics, Stanford University School of Humanities and Sciences, Palo Alto, California, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Palo Alto, California.

Muin J. Khoury, MD, PhD
Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia, and National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

Production of scientific work is regulated by reward systems. Scientists are typically rewarded for publishing articles, obtaining grants, and claiming novel, significant results. However, emphasis on publication can lead to least publishable units, authorship inflation, and potentially irreproducible results. Emphasis on claiming significant results leads to lack of publication of nonsignificant high-quality studies or to massaging data to obtain "positive" results. Emphasis on novelty leaves no incentives to spend resources on replicating prior findings to probe their correctness. Data owners have a publishing advantage without incentives to share with competitor scientists.

In the past, grapevine knowledge among the few knowledgeable experts allowed discerning the good work from the waste. But currently the noise-to-signal ratio is tremendous, with the proliferation of technologies (such as genomics) and journals. Thousands of new journals publish work for a fee, regardless of the quality of the work. To change the tide, the criteria by which scientists and their teams are rewarded for their efforts by agencies that fund them and institutions that host them should be revisited, aligning criteria with the desired outcomes: research that is productive, high-quality, reproducible, shareable, and translatable—or PQRST for short.

Productivity metrics should reward high-influence science rather than least publishable units and decrease publication bias against negative results. Instead of counting each and every publishable unit, even now several major universities ask only for the top papers from each candidate for appointment or promotion. However, the process can be standardized. Citation databases such as ISI Web of Knowledge/Essential Science Indicators automatically identify scientific fields and the x% top-cited articles in each scientific field and each year (x can be set at 1%, 10%, or other desirable percentage). Authorship contributions should also be considered when allocating credit for multi-authored papers using standard formulas such as harmonic adjustments.

Another useful metric is the proportion of published scientific work emanating from a research project. Funders can keep records of the publication of funded projects. Even for clinical trials funded by federal resources, a substantial proportion remain unpublished several years after their completion. For other types of research, nonpublication is likely to be even more frequent. Should investigators receive more funds if they used previous funds without publishing anything? In fields such as clinical research of interventions, public study registration has become widely accepted (and even enforced by regulatory agencies) and the concept of registration can be extended to other fields when appropriate. Registration allows documenting whether research studies result in published reports within a reasonable time of completion. Registration of protocols and analysis plans can also help evaluate selective reporting, ie, whether only some outcomes or analyses have been published and whether analyses deviate from promised plans.

For quality assessments, focusing on top-cited articles in assessing productivity already captures some aspects of quality, but citations and quality are not perfectly correlated. Funders can ask that protocols and studies fulfill specific quality standards in their design, implementation, and analysis. The National Institutes of Health is already moving in this direction with implementation of checklists. There are also numerous reporting standards for diverse research fields, as summarized by the EQUATOR initiative. Each field should agree on what quality features are essential. Nevertheless, difficulties in rating quality objectively should not be underestimated. Under pressure to comply, investigators may simply check off requested items while ignoring other fundamental issues relevant to the specific study. Quality assessments may focus on very few, uncontroversial, and easily verifiable study aspects. These assessments also may be used to promote improvement of that particular aspect in the whole field, eg, routine use of randomization and blinded assessments in preclinical animal studies.

Reproducibility can range from repeating analyses with raw data to independent replication using different materials or study participants or even study designs different (more rigorous) than the original study. Some types of reproducibility checks are easy. Others are prohibitively difficult, eg, performing another similar trial with new participants and 10 years of follow-up to independently replicate the results of a clinical study. This should be taken into account in deciding whether replication and reproducibility checks should be requested routinely (eg, when easy and inexpensive to do) or under select circumstances (eg, only for the most influential papers, if difficult and expensive to perform).

Sharing can also be measured. For each scientist, it is possible to assess how many papers are accompanied by shareable data, materials, or protocols. Indexing databases such as PubMed and ClinicalTrials.gov could note in each new publication record whether such shared resources are available. Many funders and journals have already made sharing practices mandatory for particular research types.

As for translation, much excellent research has no recognizable translational application or benefit. Scientific influence, if any, often becomes manifest many years after the initial discovery. However, translational performance is relevant in research with direct applied aspir-
Table. PQRS Index for Appraising and Rewarding Research

<table>
<thead>
<tr>
<th>Item in PQRS Index</th>
<th>Example</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (productivity)</td>
<td>Number of publications in the top tier % of citations for the scientific field and year</td>
<td>ISI Essential Science Indicators (automated)</td>
</tr>
<tr>
<td>R (reproducibility of scientific work)</td>
<td>Proportion of funded proposals that have resulted in ≥1 published reports of the main results</td>
<td>Funding agency records and automated recording of acknowledged grants (e.g., PubMed)</td>
</tr>
<tr>
<td>S (sharing of data and other resources)</td>
<td>Proportion of registered protocols that have been published 2 y after the completion of the studies</td>
<td>Study registries such as ClinicalTrials.gov for trials</td>
</tr>
<tr>
<td>Q (quality of scientific work)</td>
<td>Proportion of publications that fulfill ≥1 quality standards</td>
<td>Need to select standards (different per field/design) and may then automate to some extent; may limit to top-cited articles, if cumbersome</td>
</tr>
<tr>
<td>T (translational influence of research)</td>
<td>Proportion of publications that have resulted in successful accomplishment of a distal translational milestone, eg, getting promising results in human trials for intervention tested in animals or cell cultures, or licensing of intervention for clinical trials</td>
<td>No wide-coverage automated database currently, would need to be curated by appraiser (e.g., funding agency) and may need to be limited to top-cited papers, if cumbersome</td>
</tr>
</tbody>
</table>

The Table illustrates some suggestions for how to potentially operationalize these principles. The suggestions are not prescriptive but may offer ideas to funders and other stakeholders for next practical steps. The exact combination or weighting of indices will require discussion and consensus among stakeholders. Selected reward system choices should also factor the potential for gaming any appraisal and reward system. Potential untoward consequences of gaming should be anticipated, minimized, and monitored. For example, scientists may acknowledge funding from specific grants for entirely unrelated published work if their career depends on demonstrating that funding resulted in publications. If so, funders should verify that the funded work has indeed been published. Or, if reward is given only for top-cited articles, networking between investigators and journals may create a citation factory of mediocre articles that mutually propel themselves toward the top-cited range. Some of the other indices will correct this; eg, observational nutritional epidemiology has some of the most-cited papers across all science, but much of this work has failed replication.

In addition, funding agencies, universities, research institutions, academies, professional societies, and prestigious award organizations may also have PQRS indices based on the research work they sponsor or perform and the scientists behind this work. Further discussion is needed among stakeholders to refine these indices and to evaluate them within each scientific field. Special consideration will also be needed in rewarding research based on interdisciplinary team science that crosses the boundaries of multiple scientific disciplines.

**ARTICLE INFORMATION**


**Conflict of Interest Disclosures:** The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.

**Funding/Support:** The Meta-Research Innovation Center at Stanford is supported by a grant from the Laura and John Arnold Foundation. The work of Dr Ioannidis has been supported by an unrestricted gift from Sue and Robert O’Donnell.

**Role of the Sponsors:** The funding organizations had no role in the preparation, review, or approval of the manuscript or the decision to submit the manuscript for publication.

**Disclaimer:** The views expressed in this Viewpoint are those of the authors and do not reflect the official opinion of the Centers for Disease Control and Prevention or the National Institutes of Health.

**REFERENCES**