

Thought Exercises on Accountability and Performance Measures at the National Heart, Lung, and Blood Institute (NHLBI): An Invited Commentary for *Circulation Research*

Michael S. Lauer, MD

Recently, popular magazines and newspapers broadcast skeptical headlines, such as, “Desperately Seeking Cures”¹; “A Decade Later, Genetic Map Yields Few Cures”²; “Faltering Cancer Trials”³; and “Grant System Leads Cancer Researchers to Play It Safe.”⁴ Some patients, physicians, advocacy groups, journalists, scholars, and policymakers openly wonder about the value of billions of dollars in government-supported research. During the past 10 years, the NIH budget has doubled, yet cardiovascular disease, while decreasing in incidence and severity, is still the leading cause of death and cancer incidence and death rates have declined little. Along with “Where Are the Cures?” critics ask “Who Is Accountable?”⁵

Interest in Measuring the Performance of Science

There is a rapidly growing body of scholarship on the metrics of science, methods by which scientists, scientific organizations, policymakers, and funding agencies can gauge the value and impact of their work and investments.⁶ To assess the impact of the 2009 American Recovery and Reinvestment Act (ARRA), the National Science Foundation, the National Institutes of Health, and the White House Executive Office are engaged in an ambitious “STAR METRICS” project.⁷ Metrics often focus on publications and citations (“bibliometrics”) but also consider commercial products and quantitative impacts on practice or intellectual thought. Although scientists are typically eager to publish their work and see it cited, many worry about seeing their value to society, employers, and funders graded by black-box numbers.⁸ Some professional groups have harshly criticized bibliometrics, noting their inherent flaws and the risk that the act of measuring science will damage innovation, which is at the heart of the scientific enterprise.⁹

At NHLBI, we increasingly recognize that we need to be held accountable for our performance, which means that it is incumbent on us to assess the performance of the scientists and projects we choose to support. It is a genuine challenge,

however, for any nonprofit organization to assess its performance, because we cannot look to readily available metrics, such as revenues, profits, and stock prices. As in any business, it is important to align investments with expected returns. U.S. taxpayers are investing more than \$30 billion a year in the NIH and expect more concrete measures than advancement in biomedical knowledge and a fiscal accounting of funds allocated to research. Many scientists whose career trajectories are largely dependent on securing NIH funding may be frustrated by critics who hold them responsible for high population burdens of disease, along with the inability of a fragmented healthcare system to exploit prior research discoveries. How can we manage expectations and encourage investments appropriate to desired outcomes?

A Thought Exercise About Unemployment and Jobs Training

Imagine that instead of funding research in cardiovascular biology and disease, we at NHLBI were asked to establish and run a jobs-training program in an impoverished city, a city that has seen its manufacturing base disappear and has not yet realized the benefits of the new service, technology-oriented economy. Our ultimate “end-product” would be a city bustling with full-employment, with high levels of personal income and socioeconomic well-being. As a community, we’ll need to bring together numerous partners and stakeholders, including legislators, government executives and agencies, employers, labor unions, and schools and universities. We’ll need to develop a multiple-pronged strategy to deal with short- and long-term factors that contribute to unemployment; one part of that strategy may be to establish a jobs-training program, the program for which we now find ourselves responsible.

Should our jobs-training program be held solely responsible for the city’s unemployment rate? Somehow, that does not seem right, as there are many factors contributing to unemployment that go way beyond what we can possibly address with the resources available to us. However, there are performance measures we could “accept” as reasonable reflections of the quality of our work. These might include the number of unemployed workers we train and the number of courses we offer. But these measures are descriptive only of the resources we have put in place and do not tell us whether we provided the right training to the right people. We could look for other measures of impact, like how many (and what proportion) of our trainees secured employment, how many (and what proportion) secured high-paying jobs, and how many employers choose to come to us to find future

The opinions expressed in this NHLBI Page are not necessarily those of the editors or of the American Heart Association.

From the Office of the Director, Division of Cardiovascular Sciences, NHLBI, Bethesda, MD.

Correspondence to Michael S. Lauer, MD, Office of the Director, Division of Cardiovascular Sciences, NHLBI, 6701 Rockledge Drive, Room 8128, Bethesda, MD 20892. E-mail lauerm@nhlbi.nih.gov (*Circ Res.* 2011;108:405-409.)

© 2011 American Heart Association, Inc.

Circulation Research is available at <http://circres.ahajournals.org>
DOI: 10.1161/RES.0b013e3182125662

Table 1. How Results-Based Accountability Applies to Two Different Types of Government Programs

Case Unit of Accountability/Strategy	City Jobs Program	Research Funding Agency
Population (for communities, states, nation)	City residents and external businesses that may be attracted to relocate	U.S. and global population
Results	Full employment, economic well-being	Nation free of cardiovascular disease
Indicators	Unemployment rate, annual personal income, job satisfaction ratings	Cardiovascular mortality rate Myocardial infarction incidence and case fatality
Strategies	Identify causes of unemployment (manufacturing shifts, tax policies, inadequate education) Identify new opportunities for employment Identify partners (employers, state government, labor unions, schools and universities) Establish programs (eg, job-training program)	Identify causes (biology of cardiovascular disease, population behaviors, economic and social policies) Identify scientific opportunities for investigation Identify partners (researchers, policymakers, physicians, schools and universities) Establish programs (eg, government funding of researchers)
Performance (for programs/agencies)		
Identify customers	Unemployed workers Potential employers	Applicants/potential investigators Universities, along with their applicants and grantees
Identify services	Provide training of new skills	Enable applicants to navigate system by establishing fair competition, answering questions, providing advice Enable grantees to perform research by giving funds, clarifying rules, monitoring progress, allowing flexibility as appropriate
Identify metrics (over time—Figure 1)		
How much?	Courses offered (eg, computer programming, basics of laboratory technique) Number of students enrolled	Applications received and reviewed Grants funded Research funds disbursed
How well?	Faculty hired and retained Students who completed courses and passed exams	Grant reviews processed and completed according to schedule Funds disbursed according to schedule Progress reports received, reviewed, and acted upon according to schedule
What impact?	Number or proportion of students who secured jobs Average income of students who secured jobs	Applicant and grant queries responded to in timely and effective manner Clinical trials that recruit subjects and complete processes on time and within planned budgets Publications (including number, cost per publication, appearance in high-impact journals) Citations (including number, cost per citation, appearance in major systematic reviews or guidelines, appearance in high-impact journals, h-index) Commercial (including patents, products readied for clinical testing, cost per patent) Economic (jobs created, additional business activity stimulated) Practice or policy (including changes to guidelines or law, changes in evidence-base prescription practices)
Tell the stories behind the metrics	Are there measurement artifacts that convey misleading impressions? Were the appropriate skills taught? Were students correctly matched to courses? Were the right faculty hired? Did we adequately research employer needs?	Are there measurement artifacts that convey misleading impressions? Is peer review identifying the best scientists and best projects? Is competition from industry, foundations, and agencies from other countries crowding us out? Are budgets appropriate and are they adequately flexible? Do they account for dynamic changes in technologies and prices? Are powerful economic forces keeping researchers and/or patients out of government-sponsored projects? Are economic forces preventing adoption and dissemination of evidence-based technologies?
Identify partners	Local employers, educators, experts in program evaluation	Academic leaders, professional societies, journal editors, other government agencies, industry researchers, research leaders, experts in program evaluation

(Continued)

Table 1. Continued

Case Unit of Accountability/Strategy	City Jobs Program	Research Funding Agency
Strategies and action plan	Modify and improve metrics Change hiring criteria for faculty Adopt technologies to improve teaching efficiency Systematize feedback with employers to enable continuous change and improvements	Modify and improve metrics Modify peer review criteria New program announcements (sometimes with dedicated review) or initiatives Workshops and working groups Joint programs with other agencies Establish special advantages (eg, to early stage investigators)

Emphasis is on performance accountability. The table provides examples but is not meant to be comprehensive.

employees. If we focus on these measures, we can demonstrate our value to the community, and equally important, we can make informed decisions about changes we can make that will improve our efficiency and effectiveness.

Results-Based Accountability

We have just applied an approach called “Results-Based Accountability (RBA),” which is “a disciplined way of thinking and taking action that can be used to improve the quality of life in communities, cities, counties, states, and nations. Results-based accountability can also be used to improve the performance of programs, agencies, and service systems.”¹⁰ During the past few years, a number of NIH leaders have learned about RBA as part of a formal NIH senior leadership program run in conjunction with the University of Maryland School of Public Policy. I summarize how we applied RBA to unemployment and job training in Table 1.

In RBA, we start by thinking about population ends, in this case an economically successful city operating at full employment. We then move to the role of our jobs-training program and come to realize that we should focus on means, in this case training unemployed workers and enabling them to secure new jobs. We identify measureable and meaningful performance metrics, such as the proportion of our trainees who secure jobs. RBA is a helpful construct because it forces us to separate population ends from program performance.

In Table 1 and Figure 1, I summarize a similar thought exercise about how we, a Federal funding agency that supports biomedical researchers, can use the RBA construct to evaluate our performance. We proceed through a series of steps as follows.

- 1) Identify our customers and articulate the services we provide (eg, provide funds to applicants who submit highly meritorious proposals).
- 2) Identify a relatively small number of “headline” performance metrics. Some metrics focus on process (eg, accruing patients to trials, completing application reviews on time, answering queries from applicants and grantees), whereas others focus on impact (eg, publications, citations, ancillary studies, new commercial products, changes in clinical practice).
- 3) For each metric, plot changes over time and forecast what is likely to happen if we make no changes in our processes.

- 4) “Tell the story” behind the metric. We ask a series of questions to identify factors that have improved or worsened performance.
- 5) Based on our story, identify partners and possible strategies to “turn the curve,” that is, attempt to change adverse trends.

Performance Metrics for Biomedical Research

To identify what performance metrics might be useful, we can do another thought experiment. Suppose we were an outright failure, that is, we funded universities and researchers who did absolutely nothing scientific with our money. How would we know? We would find no publications that could be linked to our grants. With no publications, there would be no citations. We might also find no evidence of new products being made ready for clinical testing and eventual commercialization (ie, no patents). We would see no trials going anywhere near completion, resulting in no publications, citations, or guidelines citing our research for recommendations on changes in practice.

This worst-case scenario is easy for all of us to assess. In the real world, it is more difficult to assess whether a given level of productivity is appropriate in quantity (numbers of

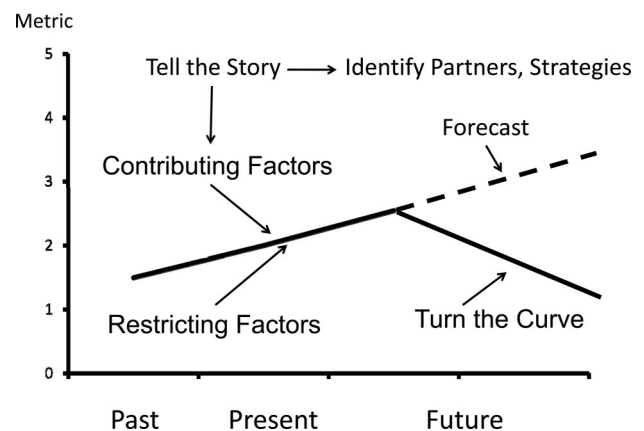


Figure 1. How to use performance metrics in results-based accountability. The X-axis represents time, whereas the Y-axis represents a quantitative metric, which could measure process or impact. After plotting the data, along with a forecast of what we think will happen in the absence of any change, we seek to “tell the story” behind the numbers by asking questions that identify contributing and restricting factors. Armed with the data and the story, we find partners and develop strategies that will “turn the curve,” producing better outcomes. Adapted.¹⁰

Table 2. Overview of Commonly Used Bibliometrics (Adapted⁶)

Metric	Definition(s)	Comments
Publications	Number in peer-reviewed journals Number of original research articles (ie, excluding reviews, commentaries, and editorials) Number in high-impact journals	Some publications may be missed because authors do not include grant number in their manuscripts Incentives for “salami slicing” Some consider focus on journal impact (as assessed by impact factor) to be a “mortal sin” Along with citations, can be used to measure the extent and impact of scientific teamwork Difficult to assess the relative contributions of multiple coauthors Difficult to assess the relative contributions of specific grants and contracts for publications that credit multiple grants and contracts Difficult to compare across fields
Citations	Total number Number in peer-reviewed journals Number per publication Number weighted for impact factor of citing journal Number weighted for number of coauthors Number weighted for scientific field	Highly correlated with scientific field, career stage, personal reputation May spread misinformation, reflecting network phenomena rather than genuine scientific impact ¹⁴ May not necessarily reflect positive impact on field
Impact Factor	For a journal, the impact factor for year y is the number of citations appearing in year y to articles published in years y-1 and y-2, divided by the total number of “citable articles” published in years y-1 and y-2	Not a valid measure of individual researchers or an individual article Even for high-impact journals, most citations come from a small proportion of articles Difficult to compare across fields
H-index	For a researcher (or laboratory, department, or any unit), the h-index is the number of publications that have been cited at least h times (eg, an h-index of 40 means a researcher has published at least 40 articles that have each been cited at least 40 times)	Measures the impact of a large body of work Correlates with other measures of scientific recognition, like Nobel prizes Can only increase with age, meaning it is unable to detect declining impact Difficult to compare across fields
Online accesses	Number of visits Number of downloads	Assesses impact beyond other scientists who publish articles Standards not yet developed

publications or inventions), quality (impact on the field of science or practice of medicine), and timeliness (how long it takes for work to “pay off”). During the past few years, there has been extraordinary interest in bibliometrics, science performance metrics based on publications and citations. I summarize some commonly used measures in Table 2.⁶ Some scholars have suggested that bibliometrics are powerful predictors of future scientific success and recognition¹¹ and that they can be used to assess teamwork in science.¹² Others express “deep concern” about the misapplication of metrics for purposes beyond which they were created¹³ and about the misuse of citation streams to support false beliefs.¹⁴ Remarkably, an almost identical conversation is taking place about education, in which there is an increasing emphasis on accountability of schools at all levels and concern that the items that can be measured may not reflect value and actually provide an incentive for schools, including universities, to reward the activities that improve test performance at the expense of the development of critical thinking and creativity in students.¹⁵

An Incomplete Example

Using an internal NIH-based search tool, I used the keyword “myocard*” to identify 1267 R01 grants that were awarded between 1990 and 2009 through the NHLBI cardiovascular research division and that were classified as “nonhuman.” To date, these grants and their successful competitive renewals account for \$2.2 billion in funding. They have led to at least

15 656 publications that have garnered 448 830 citations (approximately 29 citations per publication). The average cost per publication is \$143 276, and the average cost per citation is \$4997. These figures compare favorably to national estimates of cost per publication for all academic-based research, which has increased from \$186 567 in 1998 to \$308 641 in 2008.¹⁶

Figure 2 shows the behavior of selected process and impact metrics over time. There has been a marked increase in the number of projects funded and total funding (even after accounting for inflation¹⁷), coincident in part with the NIH doubling in the late 1990s and early 2000s. To allow for fair comparisons, the output metrics related to publications and citations are shown for each project from its start until 5 years later. The number of publications increased commensurate with the increase in funding, whereas the cost per publication remained remarkably constant. However, the number of citations has not kept pace, with a decrease in the number of citations per publication and an increase in cost per citation.

Our next step, which is critical, is to ask a series of questions that will reveal the story behind these data. This is an exercise that is beyond the scope of this article. We might start by asking why the citation rate is going down. Is it because the field is losing luster? Or, perhaps, is it because we funded a number of relatively new investigators during the doubling, whose projects have not yet matured to the point at which they can garner many citations?

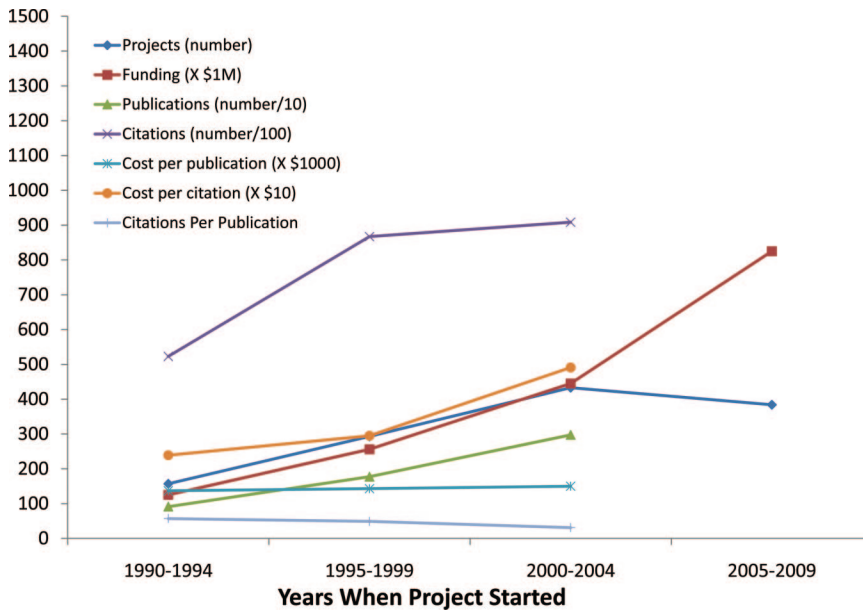


Figure 2. An example showing the behavior of selected performance metrics for a portfolio of nonhuman R01 grants that responded to a keyword of “myocard” and that were administered by the NHLBI cardiovascular research division. Process measures (projects and funding) are shown for projects starting between 1990 and 2009. Impact measures (publications, citations, and costs) are only shown for projects starting between 1990 and 2004, to allow all projects the opportunity to garner citations up to 5 years after their start. To assure fair comparisons, we only consider publications and citations occurring within 5 years of project initiation. All dollar figures are adjusted for the Biomedical Research and Development Price Index (BRDPI).¹⁷

Closing Thoughts

As stewards of taxpayer monies, we at NHLBI expect to be held accountable for our decisions and policies. No differently than scientists, hospital administrators, and for-profit company executives, we look to performance metrics as tools—nothing more or nothing less—to help understand the stories behind the effectiveness of our work and to help us make the best informed decisions. We look to colleagues in NIH, in the extramural scientific community, in medical editing, and in professional societies to work with us as we develop greater degrees of sophistication in program evaluation, ultimately to reach the goal of a society free of the ravages of cardiovascular disease.

References

1. Carmichael M, Begley S. Desperately seeking cures: how the road from promising scientific breakthrough to real world remedy has become all but a dead-end. *Newsweek*. 2010; May 15. Available at <http://www.newsweek.com/2010/05/15/desperately-seeking-cures.html>. Accessed December 27, 2010.
2. Wade N. A decade later, genetic map yields few cures. *New York Times*. 2010; June 12. Available at <http://www.nytimes.com/2010/06/13/health/research/13genome.html>. Accessed December 27, 2010.
3. Editorial. Faltering cancer trials. *New York Times*. 2010; April 24. Available at http://www.nytimes.com/2010/04/25/opinion/25sun1.html?_r=1&scp=1&sq=Faltering%20Cancer%20Trials&st=cse. Accessed December 27, 2010.
4. Kolata G. Grant system leads cancer researchers to play it safe. *New York Times*. 2009; June 27. Available at <http://www.nytimes.com/2009/06/28/health/research/28cancer.html?scp=1&sq=Grant%20System%20Leads%20Cancer%20Researchers%20to%20Play%20it%20Safe&st=cse>. Accessed December 27, 2010.
5. Galani RJ. Letter regarding National Heart, Lung, and Blood Institute and the American Recovery and Reinvestment Act of 2009: 1 year later. *J Am Coll Cardiol*. 2011; in press.
6. Van Noorden R. Metrics: a profusion of measures. *Nature*. 2010;465:864–866.
7. NSF. STAR METRICS: new way to measure the impact of federally funded research. 2010; June 1. Available at http://www.nsf.gov/news/news_summ.jsp?cntn_id=117042. Accessed December 27, 2010.
8. Abbott A, Cyranoski D, Jones N, Maher B, Schiermeier Q, Van Noorden R. Metrics: Do metrics matter? *Nature*. 2010;465:860–862.
9. Adler R, Ewing J, Taylor P. Citation statistics: a report of the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Sciences (IMS). Available at <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>. Accessed December 27, 2010.
10. Friedman M. Trying hard is not good enough: how to produce measurable improvements for customers and communities. Bloomington, IN: Trafford; 2005.
11. Hirsch JE. An index to quantify an individual's scientific research output. Available at [arXiv:physics/0508025 \(physics.soc-ph\)](http://arxiv.org/abs/physics/0508025). Accessed 2005.
12. Borner K, Dall'Asta L, Ke WM, Vespignani A. Studying the emerging global brain: analyzing and visualizing the impact of co-authorship teams. *Complexity*. 2005;10:57–67.
13. [Anon]. Assessing assessment. *Nature*. 2010;465:845–845.
14. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*. 2009;339:b2680.
15. Rothstein J. Teacher quality in educational production: tracking, decay, and student achievement. *Q J Econ*. 2010;125:175–214.
16. National Science Foundation report on science and engineering indicators: 2010, 8–94. Available at <http://www.nsf.gov/statistics/seind10/pdf/c08.pdf>. Accessed December 27, 2010.
17. Biomedical Research and Development Price Index (BRDPI): Fiscal Year 2009 update and projections for FY 2010-FY 2015. 2010; Feb 1. Available at http://officeofbudget.od.nih.gov/pdfs/FY11/BRDPI_Proj_Feb_2010.pdf. Accessed December 27, 2010.