

Measuring Knowledge in the Health Sector

Adam B. Jaffe

Brandeis University and
National Bureau of Economic Research

Measuring Knowledge in Learning Economies and Societies

OECD/NSF High-Level Forum

17 and 18 May, 1999

I. Introduction and Overview

In this paper, I will give a general background on work in the U.S. related to measurement of the state of knowledge and of the performance of innovation systems. I will then discuss specifically the state of such measurement efforts with respect to the health sector.

II. The Government Performance and Results Act and Related Measurement Efforts

The Government Performance and Results Act ("GPRA"), passed in 1993, mandates that every federal agency prepare a strategic plan by September 1997, and then regularly assess its performance relative to that plan and report the results to Congress. The plans must contain specific *performance goals*, expressed as measurable objectives against which the success of the agency can be objectively assessed. The plans must also identify the *performance indicators* that will measure the agency's success in achieving its goals. Thus the GPRA is an ambitious attempt to ensure that government programs are successful and cost-effective. Science and technology agencies view the GPRA as a major challenge.

The GPRA distinguishes between *outputs* and *outcomes*. Outputs are the direct production of agency activities and effort; outcomes are the effects or consequences that the program is intended to have. The Act mandates that performance objectives be specified for both outputs and outcomes, and that performance indicators be tracked for both. It is clear, however, that the intent of the Act is to cause agencies to establish goals with respect to

outcomes, and develop indicators that will permit them to determine if these outcomes are being achieved.¹

Beginning before GPRA, and accelerating in response to it, there have been a number of interagency efforts within the science and technology agencies in the U.S. to develop performance metrics or indicators. In 1992, the Interagency Committee on Federal Technology Transfer established a working group, chaired by the Commerce Department, on Technology Transfer Measurement and Evaluation. The group was charged with “a coordinated effort to develop a system to measure the effectiveness of Federal technology transfer and to assess the impact of Federal technology transfer programs on U.S. industry and the economy.” The Working Group issued a draft report in November 1994.²

The working draft proposed a “matrix” for characterizing technology transfer activities and measuring their impacts. The matrix allows for activities to be described and data to be collected using a set of terms and categories that (it is hoped) could be standardized across agencies. The four dimensions of the analysis are: the nature of the R&D; the development

¹ “While the Committee believes a range of measures is important for program management and should be included in agency performance plans, it also believes that measures of program outcomes, not outputs, are the key set of measures that should be reported to OMB and Congress.” United States Senate, “Government Performance and Results Act of 1993,” 103rd Congress, Report 103-58, 1993, p. 29.

² Interagency Committee on Federal Technology Transfer, Working Group on Technology Transfer Measurement and Evaluation, 1994. This draft has never been finalized.

phase; the technology application area; and the transfer mechanism. The Draft Report includes a proposed standard reporting form, in which technology transfer activities would be categorized along these dimensions, and a set of common data collected. The data specified to be collected includes some items that would be collected for all activities (mainly identifying information and the characterization along the four dimensions, as well as a description of the private-sector transferee), and then specific data to be collected for the different kinds of transfer mechanisms

The National Science and Technology Council (NSTC), in response to the GPRA mandates, established the Committee on Fundamental Science “to establish a broad framework for GPRA implementation in assessment of fundamental science programs” (Committee on Fundamental Science, 1996). Since this committee was working in direct response to GPRA, its framework mirrors that of the Act. It begins by describing performance goals at several levels. It describes “improved health and environment, national security, economic prosperity and quality of life” as the “overarching goals” of science in the national interest. In support of these overarching goals, the committee identifies “leadership across the frontiers of scientific knowledge” as “an enabling or intermediate objective.”

Having identified these goals, the document states that for the purpose of evaluating specific programs in individual agencies, “merit review based on peer evaluation will continue to be the primary vehicle for assessing the

excellence and conduct of science at the cutting edge.” The document does not advocate specific quantitative measures to be used in GPRA assessment of fundamental science, arguing that existing measures of research results (e.g., publication counts, citation counts, and rate of return and related economic measures) were developed for other purposes and capture only a subset of the spectrum of research outputs and outcomes.³

III. Inputs, Outputs, and Outcomes

The first step in designing systematic data collection is to define the various elements of the process that we wish to measure, and the nature of different kinds of measurements. The GPRA established a framework in which agencies use “inputs” to produce specific program “outputs,” which should have intended “outcomes” in the broader economy and environment. At a very general level, identifying inputs, outputs and outcomes is useful for any effort designed to develop indicators of the performance of the knowledge economy. This framework is applied to the health sector in Table 1.

While the distinction between outputs and outcomes is sometimes blurry, the conceptual dividing line is that outputs can be produced by the research process itself, while outcomes only arrive after the outputs of the research process interact with the economic and social system. Outputs are what the research sector produces, but outcomes are what we care about.

Table 1 indicates the kinds of performance indicators that might be collected to assess both outputs and outcomes. In almost all cases, outputs and outcomes themselves

³ The Committee Report also contains case studies of individual agency assessment efforts.

are largely unobservable. Instead, we measure observable constructs that we hope bear some relationship to the desired but unobservable concept. These indicators are either “proxies” that measure the concept of interest with some amount of error, or else other economic variables that we believe to be correlated with such a concept.

TABLE 1: Conceptual Framework for Indicators of Knowledge Performance

CATEGORY	CONCEPTS	INDICATORS
Inputs	Person-years, equipment-years	Expenditures
	Organizational capacity	Use of particular organizational practices
Outputs	Ideas, discoveries	Papers, prizes
	New Products	Patents, New Drug Applications
Outcomes or Impacts	Broad advance of human knowledge	Papers, citations, expert evaluations
	Improvements in health status and length of life	Outcomes studies, life expectancy
	Reduction in health-care expenditures	Outcomes studies, statistical analyses of health-care expenditures
	Economic output	Revenue growth, revenue from new products, profitability
	Productivity improvements	Productivity studies

A. Evaluating available proxies, correlates and other indicators

Since an indicator, by definition, is an imprecise measure of the underlying concept, its validity must be evaluated. Some criteria for a good indicator are characteristics of the indicator itself. For example, it should not be too expensive to collect and it should be reproducible or verifiable by independent observers. But most

issues in evaluating an indicator pertain to the relationship between the indicator and the underlying concept of interest. Since this underlying concept is unobservable, reliance on an indicator is inevitably somewhat a matter of faith. Nonetheless, it is useful to understand the issues.

First, an indicator should be as precise as possible, that is, it should bear a tight relationship with the underlying concept or have a high “signal-to-noise ratio.”⁴ Second, the indicator should be unbiased, meaning that the relationship between the indicator and the underlying concept does not vary systematically with particular characteristics. For example, if certain researchers get a disproportionately high number of citations simply because they are well known, citations would be a biased indicator of broad scientific impact. Third, the relationship between the indicator and the underlying concept should be stable over time. For example, if the widespread diffusion of computer-searchable databases has increased the number of citations made in every paper, then the relationship between citations and impact will not be stable over time, implying that reliable conclusions about the changing performance over time cannot be drawn from citations data.

Fourth, an indicator should be comparable across different environments. Note that an attribute can be precise without being comparable, and it can be comparable without being precise. Suppose that, on average, each paper in physics contains ten “ideas,” while each paper in economics contains on average one idea. Then the number of papers as an indicator for ideas is not comparable across these fields, but it might be

precise if the variation around the mean within a field is small. Conversely, it could be that the average number of ideas per paper is the same in both fields, but the variation is enormous. In this case the indicator is comparable but not precise.

Fifth, indicators should not be susceptible to manipulation. Since indicators, once developed, may be used as a basis for assessment, and assessment may be the basis of individuals' promotion and agencies' budgets, we must be concerned about the possibility that the act of measurement may influence the process being observed. Specifically, it is fine if basing the assessment of a program's creation of new ideas on publications induces scientists to produce more ideas and therefore more papers, but it is not fine if scientists publish more papers with fewer ideas per paper.

Finally, an indicator should be subject to aggregation. Since we desire to assess the performance of individual firms, sectors and countries simultaneously, we want the indicator for the whole to equal the sum of the indicators for the constituent parts. This is partly an issue of comparability—adding physics papers and chemistry papers to get total scientific papers makes little sense if papers do not constitute comparable indicators in their fields—but it is also an issue of the relationship between the indicator and the underlying concept. Suppose, for example, we estimate the rate of return to NASA satellite research by measuring the economic value of commercial satellites, subtracting an estimate of all non-NASA investment that has contributed to the industry, and then compare the resulting net economic value to NASA expenditures. (Put aside the heroic data problems that would have to be solved to do this.) Suppose we then undertook the

⁴ Formally, if we think of the indicator as being produced as the sum of the underlying concept and an unobserved random error, we wish the variance of that

analogous exercise for the Defense Department. Individually, these might be sensible proxies for the economic benefit of each agency's program. Their combined estimate for the overall rate of return to federal satellite research, however, is clearly too high, because it double-counts the extent to which total economic benefits exceed the total investments of all parties.

Of course, no indicator satisfies all these criteria. There are often tradeoffs, such as the likelihood that indicators that measure output in a particular area with high precision will not be comparable across areas. Further, the criteria are not of equal consequence, because some issues can be dealt with using appropriate statistical techniques. For example, with appropriate external information, corrections can be made for instability over time or sometimes even for noncomparability across contexts. Finally, the significance of different problems will depend on the use to which the assessment is to be put. An agency tracking its own performance for internal management purposes should care a lot about stability over time but might not care much about comparability, while a researcher or congressional staffer who is trying to figure out if the rules governing cost-sharing affect the rate of return to cooperative research will need performance indicators that are comparable across different programs and agencies.

IV. Application to the Health Sector

While there has been no systematic evaluation of the knowledge performance of the health sector in any country, many of the indicators discussed in Table 1 have been analyzed in particular contexts. Before

error to be as small as possible.

discussing some of these examples, it is worth noting that the “Health sector” itself is highly complex and heterogeneous. It includes governmental biomedical research institutions, biomedical research in universities, pharmaceutical companies, medical device manufacturers, academic clinical researchers, plus the vast array of medical service providers. Most of the research of which I am aware has focused on the public research sector and the pharmaceutical industry. There is also some work on the relationship of overall medical sector performance to measures of research and innovation. I know of little work looking in detail on knowledge creation within health services organizations.

In an important series of papers, Henderson and Cockburn (1994), Cockburn and Henderson (1996 and 1998) and Cockburn, Henderson and Stern (1999a, 1999b) have analyzed various measures of research output of the pharmaceutical industry, as well as the determinants of the productivity of research of firms. They use as their primary measure of research output the production of “important patents,” where an important patent is defined as an invention that is patented in two of the three major “markets” of the U.S., Europe and Japan.⁵ They find evidence that research productivity is affected by economies of scope in research portfolios, by spillovers of knowledge across firms doing research in related areas, by the nature and structure of R&D organizations, and by the form of performance incentives

firms offer their researchers. Their recent work has emphasized the latter point particularly, demonstrating strong complementarity between basic and applied research and higher productivity for firms that encourage contact between their researchers and public research communities.

In related work, Henderson and Cockburn (1998) measure the “absorptive capacity” of firms’ research programs by looking at connections between public and private research. They show that organizational strategies that favor openness to the scientific community, rates of coauthorship between firm scientists and scientists in public research institutions, and productivity of firm research are all correlated. Their estimates imply a significant impact of public research on private research productivity.

Henderson and Cockburn also show that the productivity of pharmaceutical research has generally been declining in recent decades. They explore various hypotheses for why this might be true, including increased competition and exhaustion of technological opportunity. There evidence tends to reject these hypotheses, in favor of a secular increase in the cost of drug development, particularly associated with significant increases in the cost of clinical trials necessary for regulatory approval.

I have been working with several coauthors on exploring the use of patent citations data to measure the contribution of research by particular

⁵ For a general discussion of the use of patents as an invention or innovation

institutions or geographic areas to the innovative success of other institutions or areas (Trajtenberg, Henderson and Jaffe, 1997; Jaffe and Trajtenberg, 1996 and 1999; Jaffe, Henderson and Trajtenberg, 1998; Jaffe, Trajtenberg and Henderson, 1993; Jaffe, Fogarty and Banks, 1998; Jaffe and Lerner, 1999). This work shows that the number of citations *made* by an organization (“backward citations”) is indicative of the extent to which the organization builds on and synthesizes previous work, and the number of citations *received* by the patents of an organization (“forward citations”) is indicative of subsequent impact. Recently, I have been exploring the applications of network systems theory to citations patterns in order to formally incorporate all backward and forward interactions of an organization into measures of its “influence” or importance (Fogarty, Sinha and Jaffe, 1998 and 1999).

Another line of work using citations information looks at the contribution of publicly funded scientific research to commercial innovation. Francis Narin and his colleagues have utilized the citations made by patents to published scientific papers, classified by country as well as (within the U.S.) the institutional affiliation of the authors of the papers. Related to the health sector, they have shown specifically that research supported by the NIH is a major contributor to medical patenting, and that U.S. inventors utilize this scientific output more than inventors from other countries.

indicator, see Griliches (1990).

The next step downstream downstream from patents is to examine new pharmaceutical products submitted to regulatory authorities for approval. In terms of sources for raw data, the FDA in the U.S. publishes statistics on the number of new drug applications (“NDAs”) and new molecular entities (“NMEs”) applied for and approved each year in the U.S. (U.S. Department of Health and Human Services, 1998). A number of authors have examined the generation of NDAs as a measure of pharmaceutical innovative output (e.g., Peltzman, 1973; Grabowski and Vernon, 1990).

Of course, just as not all patents are equally important, not all NDAs are of equal value. Dranove and Meltzer (1994) attempt to identify “important” drugs, using a mixture of data on citations, sales, and introduction into multiple countries. They find that “important” drugs defined in this manner are approved more quickly by the FDA in the U.S.

Thomas (1996) examines several measures of performance of ethical drug producers in 9 major industrialized nations. His measures include share of exports, number of innovating firms, and worldwide share of new drug discoveries. He finds that the U.S. is the dominant worldwide innovator, with a share of non-domestic sales almost 3 times as great as any other country. Switzerland, Britain and Germany also have significant international shares, but France, Italy, Japan, Sweden and the Netherlands do not. Although the link between economic performance and innovative

performance in this study is largely implicit, Thomas does argue that superior performance is tied to an emphasis on research focused on developing the most significant new chemical entities, which in turn are more likely to be globally successful products.

Moving beyond the invention, regulatory approval and sale of pharmaceuticals, another interesting line of work looks at the valuation in financial markets of the “knowledge capital” that firms create through the research process. Griliches (1981), Cockburn and Griliches (1987), Hall (1993 and 1999), and Hall, Jaffe and Trajtenberg (1999) all look in various ways at using the correlation between the stock market valuation of firms and their R&D or patents as a way of measuring the value that the market implicitly places on the intangible asset that new knowledge represents.

A number of authors have begun to try to measure the change in the performance of the health delivery sector that technology has made possible. Applying statistical and econometric techniques to data from both clinical studies and population studies, this work attempts to measure the changes in health outcomes for particular classes of medical problems, and to determine the portion of those change that is associated with new technology. Examples of this work include Cutler, McClellan and Newhouse (1998) on cardiovascular disease and Berndt, Busch and Frank (1998) on depression. A related branch of analysis hedonic analysis of drug prices. Cockburn and Anis (1998) have shown that changes in standard measures of efficacy and

toxicity do not appear to affect drug prices but do affect market shares for drugs used to treat arthritis. Measuring the output of the health care sector is complicated by the nature of health care markets and the principal/agent relationship between doctor and patient. Ellickson, Stern and Trajtenberg (1999) develop a formal analytical framework for measuring the value of health treatments taking these issues into account.

Much of the work described in the previous paragraph has been carried out in the context of attempts to derive appropriately quality-adjusted price indices for medical care. Recent papers discussing the conceptual and measurement issues that arise in attempting to measure prices and output in the health sector include Shapiro, Shapiro and Wilcox (1999) and Berndt, *et al* (1999).

A final strand of analysis goes to the most macro level by looking at the relationship *across* therapeutic categories between the introduction of new drugs and health outcomes. Lichtenberg (1996) found that disease categories in which more new drugs have been introduced have seen decreases in hospital utilization, surgery and hospital expenditure. Lichtenberg (1998) shows analogous correlation between new drug introduction and mortality reductions.

REFERENCES

- Anderson, J., N. Williams, D. Seemungal, F. Narin and D. Olivastro, "Human Genetic Technology: Exploring the Links Between and Science and Innovation," *Technology Analysis and Strategic Management* 8 (1996): 135-155
- Berndt, Ernst, S. Busch, and R. Frank, "Price Indices for Acute Phase Treatment of Depression," *National Bureau of Economic Research Working Paper* No. 6799 (1998)
- Berndt, Ernst, D.M. Cutler, R.G. Frank, Z. Griliches and J.P. Newhouse, "Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues," *National Bureau of Economic Research Working Paper* No. 6817 (1998)
- Cockburn, Iain, and A. H. Anis, "Hedonic Analysis of Arthritis Drugs," *National Bureau of Economic Research Working Paper* No. 6574 (1998)
- _____, and Z. Griliches, "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents," *American Economic Review*, 78:419-423 (1987)
- _____ and R. Henderson, "Scale, Scope and Spillovers: The Determinants of Research Productivity in the Pharmaceutical Industry," *Rand Journal of Economics* 27:1, 32-59 (1996)
- _____, "Absorptive Capacity, Coauthoring Behavior and the Organization of Research in Drug Discovery," *Journal of Industrial Economics* 46:2, 157-182, (1998)
- _____, R. Henderson and S. Stern, "Balancing Incentives: The Tension Between Basic and Applied Research," *National Bureau of Economic Research Working Paper* No. 6882 (1999a)
- _____, "Are Fixed Effects Fixed? Exploring the diffusion of Science-Driven Drug Discovery in Pharmaceutical Research," paper presented at the NBER Conference "Organizational Change and Performance Improvement," April (1999b)
- Committee on Fundamental Science, "Assessing Fundamental Science," <<http://www.NSF.gov>>, undated (participants in its development tell me it was completed in the fall of 1996).

Cutler, David, M. McClellan and J. Newhouse, "The Costs and Benefits of Intensive Treatment for Cardiovascular Disease," *National Bureau of Economic Research Working Paper* No. 6514 (1998)

Dranove, David, and D. Meltzer, "Do Important Drugs Reach the Market Sooner?" *Rand Journal of Economics* 25:3, 402-423 (1994)

Ellickson, Paul, Scott Stern, and Manuel Trajtenberg, "Patient Welfare and Patient Compliance: An Framework for Measuring the Benefits from Pharmaceutical Innovation" NBER Working Paper No. W6890 January 1999

Fogarty, Michael S., Amit K. Sinha, and Adam B. Jaffe, "ATP and the U.S. Innovation System—A Methodology for Identifying Enabling R&D Spillover Networks with an Application to Microelectro-mechanical Systems (MEMS)," mimeo, 1998

_____ "Sustaining The 'New Economy': California As A Source Of New Technology," a paper prepared for the Public Policy Institute of California, February, 1999.

Griliches, Zvi, "Market Value, R&D and Patents," *Economics Letters* 7:183-187 (1981)

_____, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 92:630-653 (1990)

Hall, Bronwyn, "Industrial Research during the 1980s: Did the Rate of Return Fall?" *Brookings Papers on Economic Activity Micro (2)*: 289-344 (1993)

_____, "Innovation and Market Value," *National Bureau of Economic Research Working Paper* No. 6984 (1999)

_____, A. Jaffe and M. Trajtenberg, "'Market Value and Patent Citations: A First Look," presented at the NBER Conference Celebrating Zvi Griliches' 20 Years as Director of the Productivity Program, March 1999

Henderson, Rebecca and I. Cockburn, "Measuring Competence: Exploring Firm Effects in Pharmaceutical Research," *Strategic Management Journal* 15: 63-84 (1994)

Interagency Committee on Federal Technology Transfer, Working Group on Technology Transfer Measurement and Evaluation, "Collective

- Reporting and Common Measures” (Draft for Comment), November 1994.
- Jaffe, Adam B., B. Banks and M. Fogarty, “Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation”, *Journal of Industrial Economics*, (June 1998).
- Jaffe, Adam B., R. Henderson and Manuel Trajtenberg, “Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988, 1965-1988,” *Review of Economics and Statistics*, (February, 1998).
- Jaffe, Adam and Joshua Lerner, “Privatizing R&D: Patent Policy and the Commercialization of National Laboratory Technologies,” presented at the NBER conference “The Patent and Innovation,” January, 1999.
- Jaffe, Adam B and Manuel Trajtenberg, “Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographical Boundaries,” *Proceedings of the National Academy of Sciences*, 93: 12671-12677, (November 1996).
- Jaffe, Adam B., and Manuel Trajtenberg, “International Knowledge Flows: Evidence from Patent Citations,” *Economics of Innovation and New Technology* (1999)
- Jaffe, Adam B., Manuel Trajtenberg and R. Henderson, “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *Quarterly Journal of Economics* (1993).
- Kettler, Hannah E., Competition through Innovation, Innovation Through Competition, London: Office of Health Economics (1998)
- Lichtenberg, Frank, “The Effect of Pharmaceutical Utilization and Innovation on Hospitalization and Mortality,” *National Bureau of Economic Research Working Paper No. 5418* (1996)
- _____, “Pharmaceutical Innovation, Mortality Reduction and Economic Growth,” *National Bureau of Economic Research Working Paper No. 6569* (1998)
- Narin, Francis, “Patents as Indicators for the Evaluation of Industrial Research Output,” *Scientometrics*, 34 (1995): 489-496.

- Narin, Francis and Kimberly Hamilton, "Bibliometric Performance Measures," *Scientometrics* 36 (1996): 293-310
- Narin, Francis, K.S. Hamilton and D. Olivastro, "The Increasing Linkage Between U.S. Technology and Public Science," *Research Policy*, 26, (1997): 317-330.
- Peltzman, Sam, "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments," *Journal of Political Economy* 81: 1049-1091 (1973)
- Shapiro, Irving, M.D. Shapiro and D. Wilcox, "Quality Improvement in Health Care: A Framework for Price and Output Measurement," *National Bureau of Economic Research Working Paper* No. 6971 (1999)
- Thomas, Lacy G., "Industrial Policy and International Competitiveness in the Pharmaceutical Industry," in Helms, Robert B., ed., Competitive Strategies in the Pharmaceutical Industry, Washington: The AEI Press (1996)
- Trajtenberg, Manuel, Rebecca Henderson and Adam Jaffe, "University Versus Corporate Patents: A Window on the Basicness of Invention," *Economics of Innovation and New Technology* (November 1997).
- U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Improving Public Health Through Human Drugs, CDER 1997 Report to the Nation (1998)