2. The theory and practice of public-sector R&D economic impact analysis

Albert N. Link and John T. Scott

INTRODUCTION

This chapter summarizes the theory and practice of public-sector R&D economic analysis with specific reference to the National Institute of Standards and Technology's (NIST's) efforts to document the impact that their in-house R&D has had on society. Motivating this research is the general expectation and challenge for public institutions to be accountable for their use of public resources.¹ Economic impact analysis is one way that public institutions can quantify the social contribution of their activity. Impact analysis can also provide important lessons to management about the effectiveness of previous resource allocation decisions, and it can provide guidelines for future strategic planning.

To place R&D impact analysis in a broader perspective, we begin with a brief discussion of R&D evaluations. An evaluation of public-sector R&D programs is based on the criterion of efficiency.² The central question asked in an R&D evaluation is: How efficient are all attributes of a public-sector R&D program including the program's management, its strategic planning, and its investment strategy?³

One part of an R&D program evaluation is an economic impact analysis. The central question asked in an economic impact analysis is: How do the social benefits associated with the publicly supported R&D program compare to society's costs to undertake the program?⁴

Public interest in program evaluation is visible to the policy community as well as to the general public. For example, on 7 October, 2009, Peter Orszag, then Director of OMB, sent a memorandum to the heads of executive departments and agencies related to increased emphasis on program evaluations. Therein he wrote:

Rigorous, independent program evaluations can be a key resource in determining whether government programs are achieving their intended outcomes. . . . Evaluations can help policymakers and agency managers strengthen the design and operation of programs. Ultimately, evaluations can help the [Obama] Administration determine how to spend taxpayer dollars effectively and efficiently (Orszag, 2009b).

16 *The theory and practice of program evaluation*

Following this memorandum, on 1 June 2010 the Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science (STAR METRICS) initiative was announced. This is a multi-agency effort led by the National Institutes of Health (NIH), the National Science Foundation (NSF), and the White House Office of Science and Technology Policy (OSTP) to help the federal government document the value of its investments in R&D.

Our focus in this chapter is on impact analyses of NIST activities, retrospective analyses in particular. Although some of the NIST impact analyses reviewed below are partially prospective in nature (meaning that the time series of estimated expected benefits and costs extends into the future because the useful commercial lifetime of the technology studied extends beyond the date of the analysis), the analyses are still retrospective in the sense that they examine NIST programs from a historical investment perspective. However, we do discuss briefly prospective analyses in the concluding section of this chapter.

The remainder of this chapter is outlined as follows. We motivate the scope of our research in the second section with an overview of the theoretical and conceptual foundations for an economic impact analysis. Although academic in nature, there is both a management and strategic planning value to understanding the economic foundations upon which program evaluations and economic impact analyses are based. We stress these economic foundations for at least two reasons. First, individuals from a number of different disciplines are involved in the conduct of program evaluations and economic impact assessments. As such, discipline-specific terminologies for similar concepts are pervasive. And second, even within a given discipline, terminologies are misused because the evaluation questions are frequently misstated.

In the third section, we discuss 17 laboratory-based economic impact analyses, 16 of which were sponsored by the Program Office, and one that was sponsored by the Advanced Technology Program (ATP) at NIST.

In the fourth section, other widely used approaches for conducting an economic impact analysis are presented and illustrated with examples. The purpose of the discussion is to provide a general overview of other complementary approaches that management at NIST or other institutions engaged in public R&D might consider during the strategic planning phase that precedes the formulation and implementation of an economic impact analysis methodology.

In the final section, we offer guidelines about the conduct of future economic impact analyses, targeted toward NIST or other similar public-sector R&D agencies, especially guidelines related to real-time data collection by the research divisions to support any upcoming analysis.

ECONOMICS FOUNDATION FOR AN IMPACT ANALYSIS⁵

In spite of efforts in the United States over the past decade to implement standardized methodologies for assessing the economic impact of government research programs, no generally accepted approach yet exists. (Tassey, 2003, p. 1).

Often overlooked, as agencies attempt to meet Government Performance and Results Act (GPRA) reporting requirements in a cost-efficient manner, is that there is an analytical and theoretical foundation or established methodology for conducting an impact analysis. This foundation, which has long been germane to the economics literature, is gaining prominence in the policy evaluation arena.⁶

Griliches (1958) pioneered the application of fundamental economic insight to the development of estimates of private and social rates of return to public investments in R&D.⁷ Streams of investment outlays through time—the costs—generate streams of economic surplus through time—the benefits. Once identified and measured, these streams of costs and benefits are used to calculate rates of return, benefit-to-cost ratios, and other related metrics.⁸

In the simplest Griliches model, public-sector innovations are conceptualized as reducing the cost of producing a good sold in a competitive market at constant long-run unit cost, as shown in Figure 2.1. For any period, there is a demand curve for the good and, in the simplest model, a horizontal supply curve. Innovation lowers the unit cost of production, hence lowering the horizontal supply curve, increasing supply, and thereby, at the new lower equilibrium price, resulting in greater consumer surplus (the difference between the price consumers would have been willing to pay and the actual price they paid, summed over all purchases).⁹

The Griliches model for characterizing the benefits from a public-sector innovation has long been the traditional economics methodology for analyzing public-sector R&D programs. The Griliches model for calculating economic social rates of return adds the public and the private investments through time to determine social investment costs, and then the stream of new economic surplus generated from those investments is the benefit. Thus, the evaluation question that can be answered from such an analysis is: What is the social rate of return to the innovation, and how does it compare to the private rate of return?

This might not be the most appropriate question to ask from a public accountability perspective. Tassey (1997) developed the concept of using private and social hurdle rates to identify appropriate targets for government intervention. The fact that the social rate of return is greater than the

18 The theory and practice of program evaluation

Source: Link and Scott (2011).

Figure 2.1 Gain in economic surplus from process innovation in a competitive market in the case of long-run constant unit costs

private rate of return could validate the role of government in innovation if the private sector would not have undertaken the research; but the question above ignores, for example, consideration of the cost effectiveness of the public sector undertaking the research as opposed to the private sector, and thus an alternative methodology might be more appropriate. Two alternative methodologies are discussed below, the counterfactual methodology and the spillover methodology, although the counterfactual methodology is the focus of the remainder of this chapter.

Counterfactual Methodology

The Griliches methodology assumes as the counterfactual situation the status quo technology (and hence the status quo demands and costs) that existed without the public R&D investments and the new technology (and hence new demands and costs) that resulted. However, with reference to what Link and Scott (1998, 2011), building on Tasse (1997), term a “counterfactual analysis”, under a counterfactual economic impact analysis methodology, a different counterfactual scenario should be considered when publicly funded, publicly performed investments are evaluated, because typically the private sector would in some way have tried to replace the public investments and the technologies they produced. Link and Scott (1998) set out for the first time the distinctions between such a

“counterfactual analysis” and the traditional analysis of economic surplus in the economics literature.¹⁰ As explained there, in the extreme case that the public R&D output could be replaced with private R&D output of equal quality, holding constant the very stream of economic surplus that the Griliches model seeks to measure, and making no attempt to measure that stream, one should ask the counterfactual question: What would the private sector have had to invest to achieve those benefits in the absence of the public sector’s investments?

In the less extreme case, where barriers to technology and market failures (discussed below) prevent the private sector from replacing the public R&D output with private R&D output of equal quality, the counterfactual question is: What would the private sector have invested in the attempt to replace the output of the public sector’s investments and what would be the value lost because of the shortfall in the quality of the private sector’s replacement R&D output?

The answer to the counterfactual question gives the benefits of the public’s investments—namely, the costs avoided by the private sector.¹¹ With those benefits—obtained in practice through extensive interviews with administrators, federal research scientists, and those in the private sector who would have to duplicate the research in the absence of public performance—counterfactual rates of return and benefit-to-cost ratios can be calculated to answer the fundamental evaluation question: Are the public investments a more efficient way of generating the technology than private sector investments would have been?

The answer to this question is more in line with the public accountability issues implicit in GPRA, and certainly is more in line with the thinking of public sector stakeholders—or so we believe—who may doubt the appropriateness of government’s having a role in the innovation process in the first place.

Spillover Methodology

There are important projects where economic performance can be improved with public funding of privately performed research. Another useful methodology that has been used to evaluate such privately performed R&D that is subsidized by public funds is what Tassey (1997), and later Link and Scott (2011), termed the “spillover analysis”. It is not the approach used in the impact analyses discussed in the next section because they are all evaluating publicly performed R&D rather than R&D that is privately performed but publicly financed. The idea that public subsidy of privately performed R&D is justified by the positive externality associated with the spillover of knowledge generated by R&D investment is an old

20 *The theory and practice of program evaluation*

one, but the development of an implementable, interview-based, multiple-equation method for identifying social and private rates of return for publicly-subsidized and privately performed R&D—with an application to projects subsidized with ATP awards—is the unique contribution of Link and Scott (2001).

Under what Link and Scott (2011), building on Tassey (1997), refer to as the “spillover economic impact analysis methodology”, the question asked is one that facilitates an economic understanding of whether the public sector should be underwriting the private-sector firms’ research, namely: What is the social rate of return from the program (including spillovers) compared to the private rate of return? Or: What proportion of the total profit stream generated by the private firm’s R&D and innovation does the private firm expect to capture; and hence, what proportion is not appropriated but is instead captured by other firms that imitate the innovation or use knowledge generated by the R&D to produce competing products for the social good?

The part of the stream of expected profits captured by the innovative firm along with its costs determine its private return, while the entire stream is the lower bound on the social rate of return (that would be compared to the social hurdle rate—the opportunity costs of the public’s investment funds). In essence, this methodology weighs the private return, estimated through extensive interviews with firms receiving public support regarding their expectations of future patterns of events and future abilities to appropriate the value of R&D-based knowledge, against private investments. The social rate of return weighs the social returns against the social investments. The application of the spillover methodology to the evaluation of publicly funded, privately performed research is appropriate because the output of the research is only partially appropriable by the private firm, with the rest spilling over to society. The extent of the spillover of such knowledge with public good characteristics and its effect on private-sector rate-of-return estimates relative to a “hurdle rate” determines whether or not the public sector should fund the research.

Impact Analysis in Practice

While there is a rich economics-based theoretical foundation for impact analyses, there are pragmatic issues at play when applying the appropriate methodology. As the review of NIST impact analyses in the next section demonstrates, and as Tassey (2003, p. 15) perceptively noted, in practice the benefits estimated in an economic impact analysis of a specific R&D program or project are “frequently determined by data availability”. Because of data limitations, the economic impact analyses for NIST’s

laboratories, and also for the Advanced Technology Program (ATP) at NIST, have often been able to estimate only a very small subset of the benefits identified when applying the counterfactual methodology appropriate for analyzing publicly financed and publicly performed R&D and collecting data relevant to quantifying the economic impacts of its investments in infrastructure technology (that is, infratechnology) research.

In the next section, we discuss 17 NIST retrospective economic impact analyses, making the point about data limitations in the context of discussing the proper scope of economic impact analyses and the actual examples of the subsets of estimated benefits from NIST's R&D to generate infrastructure technology.

NIST R&D ECONOMIC IMPACT ANALYSES

Overview of NIST Economic Impact Analyses

Much of NIST's research focuses on infrastructure technology, or infratechnology. According to Tassey (2007, p. 112): "Infratechnologies leverage the development and efficient use of technology at all three major stages of economic activity: R&D, manufacturing, and commercialization." Measurement and test methods are examples of infratechnologies. They are required for the efficient conduct of R&D, control of production, and many market transactions. Infratechnologies are also the basis for technical and functional interfaces among the products that constitute a system or tiers in a supply chain. Finally, product acceptance testing protocols and standards assure consumers that technology-based products perform as specified.¹²

This focus of NIST's research has both an institutional basis as well as an economics basis. The concept of government's involvement in standards traces to the Articles of Confederation signed on 9 July, 1778—"Congress assembled, shall also have the sole and exclusive right and power of . . . fixing the standard of weights and measures throughout the United States"—and this responsibility was reiterated in Article 1 of the Constitution of the United States. More to the point, the Omnibus Trade and Competitiveness Act of 1988 stated:

The National Institute of Standards and Technology [shall] enhance the competitiveness of American industry while maintaining its traditional function as lead national laboratory for providing the measurement, calibrations, and quality assurance techniques which underpin United States commerce, technological progress, improved product reliability and manufacturing processes, and public safety. . . .

22 *The theory and practice of program evaluation*

From an economic perspective, infratechnologies have both public- and private-good characteristics; thus, they are often referred to as quasi-public goods and are jointly supplied by the public and private sectors. Infratechnologies have economic value only if they are uniformly and widely used. As such, the private sector will underinvest in infratechnologies because of its inability to appropriate fully the benefits from such investments. Thus, a theoretical basis for NIST's role in the provision of infratechnologies is based on the economic concept of market failure.

Market failure refers to the fact that the market—including both R&D-investing producers of a technology and the users of the technology—underinvests, from society's perspective, in a particular technology or technology application. Such underinvestment occurs because conditions or barriers exist that prevent organizations from undertaking or fully appropriating the benefits created by their investments.¹³

Table 2.1 lists 17 NIST-sponsored R&D laboratory and program impact analyses that are based on a counterfactual methodology, described above, for quantifying benefit information.¹⁴ Also listed in the table is the stage or stages of economic activity benefiting from the infrastructure technology research studied in each analysis: R&D, production and commercialization. Finally, the table shows for each analysis the benefit-to-cost ratio (described below) estimated for the program studied.

There are typically benefits for R&D because NIST is doing the infrastructure technology R&D and because R&D is more difficult without good measurement. There are benefits for production because there is better process control. Also, there are benefits for commercialization because products are of higher quality (yielding more value to consumers or more efficient operation for producers using intermediate goods benefiting from the infratechnologies) and of known, consistent quality (reducing the transactions costs associated with commercialization). Table 2.1 provides a summary view of the stages focused on in the analyses, and distinguishes quantitative evaluations from those that are qualitative only.

As previously stated, each of the economic impact analyses in Table 2.1 is grounded in the counterfactual methodology of what would have happened in the absence of NIST. Then, in the first instance, the benefits from NIST's program under consideration are the costs avoided (that is, the avoided costs of activities to replace NIST's program) by industry. The counterfactual methodology is typically one in which firms incur costs in their attempts to replace the NIST infrastructure technology that is no longer available in the counterfactual situation. The benefits from NIST's infratechnology are the costs avoided by industry—costs that

Table 2.1 NIST R&D program economic impact analyses and measured economic impacts

Economic impact analysis	Stage of economic activity	Benefit-to-cost ratio*
“Economic Assessment of the NIST Thermocouple Calibration Program”, Planning Report 97-1	R&D, production, and commercialization; commercialization for users of thermocouples evaluated qualitatively only	3-to-1
“Economic Evaluation of Radiopharmaceutical Research at NIST”, Planning Report 97-2	R&D, production, and commercialization; some benefits for each stage of activity not quantified or quantified partially	97-to-1
“Economic Assessment of the NIST Alternative Refrigerants Research Program”, Planning Report 98-1	R&D, production, and commercialization; commercialization benefits, including the value of better quality products, largely unquantified	4-to-1
“Economic Assessment of the NIST Ceramic Phase Diagram Program”, Planning Report 98-3	R&D primarily, but also production (detecting and diagnosing abnormalities in production), and commercialization (because delays in the introduction of new materials were avoided); commercialization benefits of faster development of new products not quantified, and benefits for some segments of the ceramics industry not quantified	10-to-1
“Benefit Analysis of IGBT Power Device Simulation Modeling”, Planning Report 99-3	R&D, production, and commercialization; production benefits for applications manufacturers and commercialization benefits of product quality for end users evaluated qualitatively only	23-to-1
“Economic Impact of Standard Reference Materials for Sulfur in Fossil Fuels”, Planning Report 00-1	R&D, production, and commercialization; R&D benefits evaluated qualitatively only	113-to-1
“Economic Impact Assessment: NIST-EEEL Laser and Fiberoptic Power and Energy Calibration Services”, Planning Report 00-3	R&D, production, and commercialization; some benefits at all stages of activity not measured, especially commercialization benefits related to product quality	11.3-to-1

24 The theory and practice of program evaluation

Table 2.1 (continued)

Economic impact analysis	Stage of economic activity	Benefit-to-cost ratio*
“The Economic Impacts of NIST’s Cholesterol Standards Program”, Planning Report 00-4	R&D, production, and commercialization; R&D benefits were not measured because it was assumed that industry would not have attempted to develop the accurate measurement technology provided by NIST, and the other benefits were not quantified completely because of insufficient data	4.5-to-1
“The Economic Impacts of NIST’s Data Encryption Standard (DES) Program”, Planning Report 01-2	R&D, production, and commercialization; R&D benefits not estimated because survey attempts failed, and other benefits were not quantified completely because of insufficient data	145-to-1
“Economic Evaluation of the Baldrige National Quality Program”, Planning Report 01-3	R&D, production, and commercialization; commercialization benefits because of a shortfall in performance quality not estimated because it was assumed that the R&D and production costs avoided—i.e. that would have been incurred absent the program—would have successfully achieved the same quality performance as achieved using the Baldrige Criteria	207-to-1
“The Economic Impact of Role-Based Access Control”, Planning Report 02-1	R&D, production, and commercialization; some benefits (because of sensitive proprietary information and because market penetration incomplete) at the commercialization stage not quantified	109-to-1
“The Economic Impact of the Gas-Mixture NIST-Traceable Reference Materials Program”, Planning Report 02-4	Production, commercialization; benefits for some parts of the supply chain not quantified	21-to-1
“Economic Impact Assessment of the International Standard for the Exchange of Product	R&D, production, commercialization; benefits quantified are because NIST’s participation reduced the	8-to-1

(continued overleaf)

Table 2.1 (continued)

Economic impact analysis	Stage of economic activity	Benefit-to-cost ratio*
Model Data (STEP) in Transportation Equipment Industries", Planning Report 02-5	development time but not quantified are benefits because NIST's participation resulted in higher quality of the infratechnology	
"Evaluation of ATP's Intramural Research Awards Program", NIST GCR 04-866	R&D, production, commercialization; some benefits from higher quality products not quantified	
Wavelength References for Optical Fiber Communications: SRM 2517a		267-to-1
Injectable Composite Bone Grafts		5400-to-1
Internet Commerce for Manufacturing		33-to-1
Polymer Composite Dielectrics for Integrated Thin-Film Capacitors		7-to-1
"Economic Analysis of NIST's Investments in Superfilling Research", Planning Report 08-1	R&D, production, commercialization; only the benefits of reduced R&D costs were quantified	6-to-1
"Economic Analysis of NIST's Low-k Materials Characterization Research", Planning Report 08-2	R&D, production, commercialization; only the R&D benefits were quantified	9-to-1
"Retrospective Economic Impact Assessment of the NIST Combinatorial Methods Center", Planning Report 09-1	R&D; commercialization benefits (resulting from the more efficient product R&D enabled by the infratechnology) for the users of the higher quality products not quantified	9-to-1

Notes:

Planning Reports are available at http://www.nist.gov/director/planning/impact_assessment.cfm. NIST GCR 04-866 is available at the preceding location under "Advanced Technology Program" and at http://www.atp.nist.gov/eao/eao_pubs.htm.

* Several other outcome metrics are in each Planning Report.

26 *The theory and practice of program evaluation*

would be incurred had industry tried to replace NIST's services in the counterfactual situation without NIST's program, provided in cooperative, public-private partnership between industry and NIST. Those costs would include not only the costs of activities to replace NIST's services, but also the lost economic value if the quality of industry's counterfactual alternative to the absence of NIST's services fell short of the quality of the NIST services being replaced. Such shortfalls in value are expected because the infratechnologies are quasi-public goods—they are provided through the public-private partnership of NIST and industry—and the ideal case typically entails a combined investment by NIST and industry. For that reason, the social costs of the NIST programs evaluated with economic impact analyses include not only NIST's costs, but the costs incurred by industry in its support of the NIST infratechnology programs.

It is this mixed case, where the private sector to some extent attempts to replace NIST's technology but does so incompletely, that is typically observed in the 17 economic impact analyses listed in Table 2.1. However, before reviewing the actual economic impact analyses, it will be important to set out clearly the two extreme and hypothetical, yet conceptually very important, special cases—the use of the status quo technology and the complete replacement case.

Status Quo Technology

In those cases where industry, because of severe barriers to technology development, would not have even attempted to replace NIST's program but instead would have simply worked with the technology available, the analysis is essentially the same as the traditional evaluation of social rate of return to public R&D investments in Griliches (1958). The counterfactual in the traditional method is the status quo without the innovation for which social rate of return is being evaluated. Thus, for Griliches' classic paper, the counterfactual is the old technology and associated unit cost of production before the innovation lowered unit cost.

Complete Replacement of Infrastructure Technology

In the case where industry does undertake to replace NIST's program, incurring the costs to replace it and doing so completely, the benefits of NIST's program are simply the costs avoided because industry did not have to incur the costs of establishing and operating the program in the private sector; there is no shortfall in the value generated with the replacement technology. Then, a benefit-to-cost ratio (discussed below) greater than

1.0 implies that the NIST program provided the infrastructure technology more efficiently than the private sector could have done. The costs to replace NIST's program are avoided by industry, and these avoided costs are the benefits from NIST's program. With a benefit-to-cost ratio greater than 1.0, those costs and hence the benefits exceed the total of NIST's and industry's costs for the NIST infrastructure technology program. Thus, in the extreme case of complete replacement of NIST's program by industry and the achievement of the same stream of economic surplus therefrom, the evaluation metrics compare the investment costs industry avoided (the benefits of NIST's program in the complete replacement scenario) with society's (NIST's and industry's) actual investment costs.

A Summarizing Restatement

At the extreme of the case of zero replacement by industry of NIST's infratechnology investments, the shortfall in value in the counterfactual situation absent NIST's program is complete, and the evaluation metrics compare the stream of economic surplus—the benefits from having NIST's program rather than the status quo ante technology—with the investment costs of NIST's program. The latter case where the shortfall in value is assumed to be complete may be the appropriate counterfactual scenario when barriers to technology development and use are especially severe and the discussions with industry support the belief that market failure would have prevented the private sector from investing in socially valuable infrastructure technology that NIST could provide in cooperation with industry.

However, barriers to technology may not preclude private-sector provision of the infratechnology, and in that case the counterfactual will have as benefits of NIST's program the replacement costs avoided by industry. When those replacement costs exceed the costs of NIST's program, then the benefit-to-cost ratio will exceed 1.0, net present value will be positive, and the social rate of return will exceed the opportunity costs of the public's funds (that is, society's hurdle rate).¹⁵

One could argue that in Griliches' classic article, the counterfactual used for evaluating the public's investment in developing hybrid corn was not the right one for providing an analysis of economic impact. In particular, if large manufacturers of farm supplies including seeds would, in the absence of the government R&D program, have undertaken their own development programs, the relevant counterfactual would not be the technology of the status quo ante, but instead the situation where the stream of economic surplus generated by hybrid corn was captured by private investment. Then, for the evaluation of the public R&D, the benefits to

28 *The theory and practice of program evaluation*

weigh against the government's investment costs would have been not the stream of economic surplus from hybrid corn but instead the costs avoided by the private sector because it did not have to do the investment. The government's program would be judged a success if it performed the research at a lower cost than the private sector's cost in the counterfactual situation.

Mixed Case Counterfactual Analysis

A review of the analyses in Table 2.1 shows the typical case is the mixed case where in the counterfactual situation the private sector to some extent attempts to replace the technology of NIST's program but does so incompletely. The prevalence of the mixed case is not surprising given that the NIST programs are providing quasi-public goods in the context of public-private partnership to develop and apply infratechnologies. Although the economic impact analyses in these tables span 22 years, and although they were conducted by different academics and contractors with varying backgrounds and skill sets, some generalities about the scope of the economic impact analyses and the availability of data can be gleaned.

Reviewing the projects in Table 2.1 suggests a generic set of data and associated evaluation approaches when NIST is conducting an impact analysis, or when doing economic evaluations more generally. For those managing NIST laboratory projects, the taxonomies of data and approaches, revealed by a review of the 17 analyses listed in Table 2.1, are described here. From the descriptions here, managers can identify data available for their projects that could be routinely collected on an ongoing basis and then periodically used for impact analysis.

Data about Economic Impacts

The first observation about data to measure economic impacts is that those impacts will be observed in the supply chains with benefits attributable to the infratechnologies provided by NIST. The scope of an impact analysis, in terms of the data about impacts that it should set out to observe, will depend on the structure of the supply chain. That point can be explained with two extreme, hypothetical examples.

First, consider a case where the infratechnology is embodied in an upstream manufacturer's product and then used in the supply chain's downstream markets. Suppose that the upstream firm is a perfectly price-discriminating monopolist (that is, it can collect all of the area under its demand curve) selling to a perfectly competitive downstream market that buys the monopolist's product and embodies it in the product it produces

and then sells to the next stage of the supply chain. Assume moreover that all markets further downstream in the supply stream are also perfectly competitive. In this special case, collecting data about the infratechnology's effect on the first beneficiary in the supply chain will be sufficient to establish the benefits. In particular, one would gather data related to the outward shift in its demand curve because of the higher quality, and/or the downward shift in its average costs allowed by the NIST program's infratechnology, or about the private sector replacement costs to replicate NIST's technology and then, if there is a shortfall in quality, gather the data about the lost outward shift in demand and/or downward shift in costs.

Second, at the other extreme, imagine a case where the NIST program's developed infratechnology is used in a supply chain that is perfectly competitive at all stages of the supply chain. In that case, the social value of NIST's infratechnology can be captured by gathering data from the retail market for the product produced in the supply chain. In particular, one would gather data about the outward shift in the retail market's demand curve and/or about the downward shift in average costs or about the private sector's replacement costs to replicate the NIST program's infratechnology, and then, if there is a shortfall in quality, gather the data about the lost outward shift in demand and/or downward shift in costs.

Typically, the market structures of the supply chains benefiting from NIST's programs will not be either of these extreme cases, and consequently benefits occur and theoretically must be estimated through data collection and analysis at all levels of the supply chain.

However, the analyses in Table 2.1 reveal a limited ability of the evaluation teams to collect data of the theoretically desired scope.¹⁶ More often than not, the analyses use only a small part of the supply chain from which to gather benefit data, even when it is clear that the scope of the benefits extended throughout many parts of the supply chain. The analyses then acknowledge this limitation and observe that the evaluation metrics are conservative in the sense that they will systematically underestimate the true benefits of NIST's infratechnology investments.

To improve the ability to document benefits throughout the supply chain, it is recommended that, to the extent practicable, NIST project managers maintain key contacts with the users of their projects' outputs at all stages of the supply chain and let those contacts know the types of data that will be needed to estimate benefits. However, experience has shown that tiers in a supply chain beyond those with which NIST has had direct contact (and therefore where the beneficiaries recognize that NIST has made significant contributions) do not have an incentive to cooperate and, in fact, do not.¹⁷

30 *The theory and practice of program evaluation*

It is recommended that NIST project managers establish direct contact with those further down the supply chain who benefit from the NIST infratechnologies, explain those downstream benefits that are expected, and then request feedback about the extent to which those benefits are actually realized. In other words, with an educational outreach effort by NIST to those in tiers of a supply chain where traditionally there have not been direct contacts with NIST, the incentives to cooperate will be cultivated and NIST programs will be better attuned to industry's needs. Ideally, benefit data would be routinely gathered in real time on an ongoing basis. The availability of such data would allow periodic evaluations documenting impacts, thus facilitating NIST's ability to make possible adjustments that would better serve industry's needs and also allowing NIST to provide quality information about performance to support the mandate of the GPRA.

Numbers of respondents and statistical confidence intervals

Although several of the analyses in Table 2.1 provide upper and lower bounds for their estimated benefits, none of the analyses provides formal statistical confidence intervals for the estimates. In many cases a reason for this is that the numbers of respondents are too few to develop any formal statistics. The numbers of respondents providing information for the NIST economic impact analyses listed in Table 2.1 typically are far below what would normally be used in statistical analysis.¹⁸ However, if project managers follow the recommendations for maintaining in real time key contacts throughout the supply chains benefiting from NIST projects and communicating regularly about the types of data that will be needed for evaluations, there will be many more respondents and the very real problem of small numbers of expert opinions about key benefits can, in some cases, be mitigated dramatically. With larger numbers of respondents providing estimates of benefits of particular types, it will at times be practical to produce formal statistical confidence intervals for the estimated benefits.¹⁹

Types of impact data

The analyses in Table 2.1 rely on many types of impact data that NIST project managers could request in real time from the key contacts throughout the supply chains benefiting from their projects' outputs. It is useful to discuss these examples in groups for a generic supply chain benefiting from NIST's infrastructure technology investments.

The analyses in Table 2.1 illustrate a range of different supply chains, but for illustrative purposes the generic supply chain will have the first level of beneficiaries being firms upstream that manufacture an input

used downstream in the supply chain. The input produced by the first tier of beneficiaries could be an instrument that must be calibrated using the standards and calibration services developed at NIST, or it could be a reference material of field quality that is traceable to NIST standards. An example of a first tier of beneficiaries would be the suppliers of power meters and 249 nm excimer lasers used for photolithography allowing economic fabrication of miniaturized integrated circuits (see Planning Report 00-3). Another example would be the manufacturers of cholesterol measurement systems that are calibrated using NIST's cholesterol standard reference materials (SRMs) (see Planning Report 00-4).

The second tier of beneficiaries could be downstream manufacturers or service providers that must use the inputs purchased from the upstream firms to efficiently produce a product that will be sold to manufacturers or service providers further downstream in the supply chain. The second tier of firms, for example, might use an instrument that measures the wavelength of light to manufacture components that will be used in fiber optics communications systems.

For example, NIST standards and calibration services allow accurate calibration of instruments used in the manufacturing of specialized optical sources and detectors used in high performance communications systems. The NIST standards allow the manufacturers to characterize the frequency response of high-speed detectors that are essential to enabling many downstream areas of the telecommunications industry, including high-speed internet access. Or, the second tier of firms might use a reference material traceable to NIST standards to measure the sulfur content of a fuel that will be used in the production of energy. The goods produced by the second tier of firms benefiting from NIST's infrastructure technology are then used by firms further downstream; in the examples, those firms would be the providers of fiber-optic communications systems or the producers of energy.

Those firms in the third tier of the supply chain benefit from having inputs—the fiber-optic components or the fuel—with specifications traceable to NIST standards. Then, for the generic supply chain, the fourth tier will be the end users who also are beneficiaries of NIST's infratechnology. In the examples here, the end users would be the customers who use the services of a fiber optics communication system or those who use the energy produced with fuel with content meeting specifications traceable to NIST standards.

Table 2.2 provides examples of the types of data used in the analyses in Table 2.1 to estimate benefits of NIST's infratechnology investments for the various levels of the typical supply chain described in the preceding paragraph. Excepting the special cases discussed earlier in the text, benefits

32 The theory and practice of program evaluation

Table 2.2 Examples of impact data for a generic supply chain

Stage of the supply chain	Examples of impact data
First tier: Firms producing inputs—for example, materials with characteristics determined by characteristics data traceable to NIST	<p>Avoided cost of using foreign national laboratories, rather than NIST, for traceability to standards</p> <p>Avoided cost of using commercial consultants, rather than NIST, for technical support services</p> <p>Avoided cost for users of reference materials with higher operations and production expenses because of their lowered confidence in their reference materials (e.g. manufacturers of field quality calibration gas that would have greater measurement uncertainty without NIST-traceable reference materials (NTRM), or manufacturers of instruments to monitor emissions)</p> <p>Avoided cost of additional measurement equipment</p> <p>Avoided cost of calibrating and maintaining in-house measurement systems</p> <p>Avoided cost of verifying the accuracy of measurements for customers</p> <p>Avoided cost of measurement disputes, without traceability to NIST, with customers</p> <p>Avoided cost of lost value because of delay in product reaching market</p> <p>Avoided R&D costs—labor and equipment</p> <p>Avoided cost of inefficient production processes</p> <p>Avoided costs of maintaining quality control and assurance systems</p> <p>Avoided cost that would be incurred to prevent interoperability problems (throughout the supply chain) before they occur and that would be incurred to address interoperability problems after they occur (e.g. for users of STEP, the international standard designed to address interoperability problems encountered in the exchange of digital product information)</p>
Second tier: Firms producing inputs—for example, components, production equipment or other inputs with specifications	<p>Avoided cost of using foreign national laboratories, rather than NIST, for traceability to standards</p> <p>Avoided cost of using commercial consultants, rather than NIST, for technical support services</p> <p>Avoided cost for users of reference materials with higher operations and production expenses because of their lowered confidence in their reference materials, instruments, and system components (e.g. manufacturers of systems for monitoring and controlling emissions)</p> <p>Avoided cost of additional measurement equipment</p>

(continued overleaf)

Table 2.2 (continued)

Stage of the supply chain	Examples of impact data
traceable to NIST standards	<p>Avoided cost of calibrating and maintaining in-house measurement systems</p> <p>Avoided cost of verifying the accuracy of measurements for customers</p> <p>Avoided cost of measurement disputes, without traceability to NIST, with suppliers or customers</p> <p>Avoided costs of implementation (risks and adoption costs) and interoperability across systems (e.g. software developers using infratechnologies supporting role-based access control (RBAC))</p> <p>Avoided cost of lost value because of delay in product reaching market</p> <p>Avoided R&D costs—labor and equipment</p> <p>Avoided cost of inefficient production processes</p> <p>Avoided costs of maintaining quality control and assurance systems</p> <p>Avoided cost that would be incurred to prevent interoperability problems (throughout the supply chain) before they occur and that would be incurred to address interoperability problems after they occur (e.g. for users of STEP, the international standard designed to address interoperability problems encountered in the exchange of digital product information)</p>
Third tier: End user firms—for example, firms (such as OEMs) producing systems or outputs (autos, TVs, computers) with specifications traceable to NIST standards	<p>Avoided cost of measurement disputes, without traceability to NIST, with suppliers or customers</p> <p>Avoided cost for end users that would have higher operations, maintenance and production expenses because of their lowered confidence in the accuracy of measurement technology (e.g. manufacturers using systems for monitoring and controlling their emissions)</p> <p>Avoided cost of additional measurement equipment</p> <p>Avoided cost of calibrating and maintaining in-house measurement systems</p> <p>Avoided cost of verifying the accuracy of measurements for customers</p> <p>Avoided cost of lost value because of delay in product reaching market</p> <p>Avoided R&D costs—labor and equipment</p> <p>Avoided cost of inefficient production processes</p> <p>Avoided cost of regulatory compliance for traceability to absolute standards (e.g. emissions control)</p>

34 *The theory and practice of program evaluation*

Table 2.2 (continued)

Stage of the supply chain	Examples of impact data
	<p>Avoided costs of maintaining quality control and assurance systems (e.g. clinical laboratories performing cholesterol tests)</p> <p>Avoided costs of implementation (risks and adoption costs) and interoperability across systems (e.g. end users of infratechnologies supporting RBAC)</p> <p>Avoided cost that would be incurred to prevent interoperability problems (throughout the supply chain) before they occur and that would be incurred to address interoperability problems after they occur (e.g. for users of STEP, the international standard designed to address interoperability problems encountered in the exchange of digital product information)</p> <p>Avoided loss in quality (product features, performance, reliability) of products as reflected in customers' decreased willingness to pay for the products [Note that care must be taken to avoid double-counting of this decreased willingness to pay; ideally the deterioration in quality at all levels of the supply chain and associated loss in customers' willingness to pay can be captured at the penultimate tier of the supply chain, with the cumulative effects of deterioration in quality throughout the supply chain reflected in the demand curve for the products of the end user firms by the decreased willingness to pay for their final product sold to the final consumers. At times the loss in value from quality deterioration may be captured in the downward shift of the demand curve for an upstream stage of the supply chain, but it must not then be also measured in the downstream markets.]</p>
Fourth tier: Customers of end user firms—for example, consumers using the final goods and services provided by the third tier firms	<p>Avoided cost of measurement disputes, without traceability to NIST, with suppliers</p> <p>Avoided costs of maintenance and repair (e.g. the users of refrigeration equipment with improperly functioning alternative refrigerants)</p> <p>Avoided costs of damage from, or costs of repeating, final consumption that had poor results because of inaccurate measurement (e.g. the patients at the end of the supply chain in the radiopharmaceuticals analysis; or, e.g. regarding costs of damage from poor measurement, the environmental damage without standard reference materials (SRMs) for the sulfur content of fossil fuels)</p>

occur and are ideally measured at all levels of the supply chain. One should note that the examples given for each tier in the generic supply chain are all taken from the actual benefits documented in the analyses summarized in Table 2.1; however, in many cases a particular type of benefit that has been measured for a particular tier of the supply chain could also occur at other parts of the supply chain.

One can observe that the examples given in Table 2.2 include many different types of benefits from NIST's programs. These are benefits that could be quantified and collected in real time and used by NIST for more effective management. The types of benefits include avoiding costs that would be incurred by the firm in the counterfactual scenario without NIST's investments: investment, operating and maintenance costs to develop, use and maintain infrastructure technology and provide infratechnology services such as standard reference materials (SRMs) and calibrations services and traceability to national standards. Benefits to the firm also include avoiding increased time to market in the absence of NIST's programs, avoiding the loss of valuable knowledge that would not have been developed without NIST's infratechnologies increasing the performance of R&D investments (such knowledge is reflected in part in knowledge metrics such as patents and publications). Benefits include avoiding a loss in quality of products and services—that is, avoiding a loss in the firm's performance in the absence of NIST's investments. Benefits also include avoiding costly increases in time to meeting third party regulations or simply more difficulty in detecting whether or not the regulations are being met and that difficulty requires costly countermeasures to ensure that in fact the regulations are met.

Outcome Metrics

The outcome metrics traditionally used in the NIST retrospective economic impact analyses in Table 2.1 are the internal rate of return (*IRR*), the benefit-to-cost ratio (*B/C*), and net present value (*NPV*).²⁰

Internal rate of return

The internal rate of return (*IRR*)—a real rate of return in the context of constant-dollar cash flows—is the value of the discount rate, i , that equates the net present value (*NPV*) of the stream of net benefits associated with a research project to zero. The time series runs from the beginning of the research project, $t = 0$, through a terminal point, $t = n$.

Mathematically:

$$NPV = [(B_0 - C_0)/(1 + i)^0] + \dots + [(B_n - C_n)/(1 + i)^n] = 0 \quad (2.1)$$

36 The theory and practice of program evaluation

where $(B_t - C_t)$ represents the net benefits associated with the project in year t , and n represents the number of time periods—years in the case studies below—being considered in the evaluation.

For unique solutions for i , from equation (2.1), the *IRR* can be compared to a value, r , that represents the opportunity cost of funds invested by the technology-based public institution. Thus, if the opportunity cost of funds is less than the internal rate of return, the project was worthwhile from an *ex post* social perspective.²¹

Benefit-to-cost ratio

The ratio B/C is the ratio of the present value of all measured benefits to the present value of all measured costs. Both benefits and costs are referenced to the initial time period, $t = 0$, when the project began as:

$$B/C = [\sum_{t=0 \text{ to } t=n} B_t / (1 + r)^t] / [\sum_{t=0 \text{ to } t=n} C_t / (1 + r)^t] \quad (2.2)$$

A benefit-to-cost ratio of 1 is said to indicate a project that breaks even. Any project with $B/C > 1$ is a relatively successful project as defined in terms of benefits exceeding costs.

Fundamental to implementing the ratio of benefits-to-costs is a value for the discount rate, r . While the discount rate representing the opportunity cost for public funds could differ across a portfolio of public investments, the calculated metrics in the analyses in Table 2.1 follow the guidelines set forth by the OMB (1992) in Circular A-94: “Constant-dollar benefit-cost analyses of proposed investments and regulations should report net present value and other outcomes determined using the Office of Management and Budget (OMB) recommended real discount rate of 7 percent.” The analyses listed in Table 2.1 are evaluating investment decisions—the allocation of capital across alternative investment options, and so the OMB-mandated 7 percent real discount rate has been used.²²

Net present value

OMB circular A-94 states (OMB, 1992, p. 3):

The standard criterion for deciding whether a government program can be justified on economic principles is net present value—the discounted monetized value of expected net benefits (i.e., benefits minus costs). Net present value is computed by assigning monetary values to benefits and costs, discounting future benefits and costs using an appropriate discount rate, and subtracting the sum total of discounted costs from the sum total of discounted benefits.

The information developed to determine the benefit-to-cost ratio can be used to determine NPV as:

$$NPV_{\text{initial year}} = B - C \quad (2.3)$$

where, as in the calculation of B/C , B refers to the present value of all measured benefits and C refers to the present value of all measured costs, and where present value refers to the initial year or time period in which the project began, $t = 0$ in terms of the B/C formula in equation (2.2). Note that NPV allows, in principle, one means of ranking several projects *ex post*, providing investment sizes are similar.

It could be argued that for the purpose of public consumption of the finding from an economic impact analysis, NPV should be calculated as the difference between the present value of benefits and the present value of costs where both present values are referenced not to an initial period but rather to the time period that corresponds to when the analysis is conducted, $t = n$. The reason for this alternative approach (that is, alternative point of time reference to that implied by equation (2.3)) is that the use of NPV in a retrospective economic impact analysis is different from its conventional use as described in corporate finance textbooks. In the latter case, objectives include comparing alternative investments subject to a corporate constraint on funds available for investments, or deciding which investments will add to the value of the corporation's stock. All estimates are in current dollars, as the "present" is the time of decision. In the case of a retrospective analysis, and all of the analyses in Table 2.1 are retrospective in scope, the analysis takes place some considerable amount of time after the project was initiated (the "present" time in the corporate finance use of this metric). Because the results of retrospective analyses are read by policy makers and other stakeholders some time—in most cases, many years—after the project was initiated, these stakeholders will compare results in a "current-dollar" context. Hence, the reference of NPV to the time of the analysis can allow stakeholders to compare the results with current investment options. In other words, the following formula for NPV should be considered:

$$\begin{aligned} NPV_{\text{year of the analysis}} &= [\sum_{t=0 \text{ to } t=n} B_t \times (1+r)^{n-t}] - [\sum_{t=0 \text{ to } t=n} \\ C_t \times (1+r)^{n-t}] &= (NPV_{\text{initial year}}) \times (1+r)^n \end{aligned} \quad (2.4)$$

After the experience with the investment, a positive $NPV_{\text{initial year}}$ shows that at the outset of the investment, the value of the stream of benefits

38 *The theory and practice of program evaluation*

exceeded the value of the stream of costs by the amount $NPV_{\text{initial year}}$. The intuitive story for $NPV_{\text{year of the analysis}}$ is that at the time the investment is put in place, its stream of benefits could, theoretically, be sold for the stream's present value, and then a portion of the proceeds equal to the present value of the costs could be invested to release the stream of costs needed for the project, leaving the $NPV_{\text{initial year}}$ as an excess of value above and beyond costs. That value could at that time have been invested, and then the resulting value from the project by the time of the analysis would have been $NPV_{\text{year of the analysis}}$.

OTHER TECHNIQUES RELEVANT TO AN ECONOMIC IMPACT ANALYSIS

The economics-based methodologies discussed above provide analyses that are theoretically sound and more commonly used with reference to R&D- and technology-based programs, especially for analyses of R&D programs funded by US agencies. However, other retrospective techniques, some of which are also economics-based or can be used to evaluate an economics-based measure of performance, have been used to assess or to support economic impact analyses of public-sector R&D programs. Following Polt and Rojo (2002), these other techniques include econometric models, productivity models, benchmarking analysis, innovation surveys, expert panels and peer review, and network analysis.²³ Each of these other techniques is discussed below. In the concluding section of this chapter we suggest that, when possible, these techniques be considered as tools to complement the economics-based methodologies discussed earlier.

Econometric Models

An ideal analytical approach [for a retrospective evaluation] is the construction of a time series of economic activity of affected industries that includes a period before government intervention. At some point in the time series, a government funded project . . . occurs and the subsequent portion of the time series reflects the technical and economic impacts of the intervention. (Tassey, 2003, p. 15).

Econometric models (more specifically models that are estimated using econometric methods) quantify the level of a predefined performance variable before and after the public-sector R&D program being considered had an effect.²⁴ The time series data that are used in these models relate to policy-relevant economic units such as a sector, an industry, a firm, an

organization or institution, or an individual. Generally, the data, when available, pertain to individual firms' performances before and after the public-sector R&D program.

Define a performance variable for the i th firm as P_i . Consider two series of data. The first is a time series of data on the observed performance of k firms, $i = 1$ to $i = k$, before and after the effect from the public-sector R&D program. After the public-sector R&D program is operating, each of the k firms will be affected, but not necessarily in the same degree. If performance data are available from time periods $t = 0$ to $t = n$, and if the public-sector R&D program became effective at time period $t = t^*$, then the relevant comparison is between the performance of the k firms before the R&D program, P_i , for $t = 0$ to $t = t^* - 1$, and their performance after the R&D program, P_i , for $t = t^*$ to $t = n$.

The second series of data could be on the performance of affected and non-affected firms (that is, matched pairs of firms) after the public-sector program was initiated at t^* . If performance data are only available from time $t = t^*$ to $t = n$ for k affected firms, P_i , $i = 1$ to $i = k$, and for m non-affected firms, P_j , $j = 1$ to $j = m$, then, for each matched pair of firms, the relevant comparison is over time between P_i and its matched P_j .²⁵ The counterfactual situation, that is the situation without the public-sector R&D program, is the performance of the m non-affected firms.

In the case of the first series of data that quantifies pre- and post-R&D program performance, pooled cross-sectional and time series data could be used to estimate a model that takes the general form:

$$P_{i,t} = a_0 + a_1 RD_{t^*} + \text{control variables} + \varepsilon \quad (2.5)$$

where $P_{i,t}$ represents the relevant performance variable of the i th firm at time t ; RD represents the public-sector R&D program being evaluated that was initiated or became effective at time t^* — RD takes on a value of 0 for the time period before t^* and a value of 1 at t^* and afterwards;²⁶ and ε is a normally distributed random error term.

Estimated regression parameters from equation (2.5) allow one to interpret the economic impact of the public-sector R&D program. For example, the estimated coefficient on RD in equation (2.5) quantifies the impact of the R&D program on the average performance of the sample of k firms. If the estimated value of a_1 is positive and statistically significant, then the public-sector R&D program had a measurable positive impact on firm performance, all other factors held constant. If the public program is hypothesized to change the extent to which various control variables affect performance, then interaction terms would be added to the specification with each interaction multiplying RD with a control variable for

40 *The theory and practice of program evaluation*

which impact would be affected. For example, the public R&D program might have not only an “intercept effect” but also have a “slope effect” for a variable such as the firm’s own R&D investment, which could be more effective given the public R&D investments.

In the case of the matched pairs of firms in which the counterfactual situation is approximated by the performance of firms not affected by the target R&D program, cross-sectional time series data could be used to estimate a model that takes the general form:

$$P_t = b_0 + b_1 E + \text{control variables} + \varepsilon \quad (2.6)$$

The variable E divides the sample of firms into those affected by the R&D program and those matched pairs that are not affected: E takes on a value of 0 for the m non-affected matched firms and a value of 1 for the k affected matched firms. If the estimated value of b_1 is positive and statistically significant, then the R&D program had a measurable positive impact on firm performance relative to the performance of “similar” firms not affected by the program, other things held constant.²⁷

There are a number of important data issues related to the use of the econometric models discussed above, and one of those issues relates to how the performance variable, P , is measured. If P is measured in terms of the stated goals of the public-sector R&D program, then the use of econometric models might be an appropriate tool for an economic assessment. If, however, there are spillovers and they are the focus of the public R&D program, then measuring firm performance, P , rather than a broader measure of performance will not work for an evaluation of the public R&D.

To illustrate, one notable economic assessment analysis by Busom (2000) is based on a model that is, in concept, equivalent to that in equation (2.6) above. She examined a sample of 154 Spanish firms that conducted R&D in 1988. About 45 percent of the firms received public support of their R&D through an agency of the Spanish Ministry, and a stated goal of this public-support program was to leverage private R&D. Other factors held constant, Busom found that participation in the public program did increase private R&D effort.²⁸

Productivity Models

Productivity models are a special case of a performance model (that is, P in the previous section) that has been estimated using econometric tools. The special case is highlighted here because many of the innovation-based policy responses to the productivity slowdown in the early 1970s and

then again in the late 1970s were based on what economists frequently call production function analysis. A production function is a mathematical representation of the relationship between a firm's (or other unit of analysis) output and the inputs that generate that output. For example, it is generally assumed that a firm's flow of labor (L), its stock of physical plant and equipment or capital (K), and its stock of technical knowledge (T) are relevant inputs in the production of a firm's output (Q) as:

$$Q = F(L, K, T) \quad (2.7)$$

Under a set of stylistic assumptions, such as the functional form for $F(\bullet)$ and the relationship between T and the firm's investments in R&D, an econometric model can be derived from which one can estimate the rate of return to the firm's investments in R&D. An estimate of the rate of return to investments in R&D could be useful for an economic impact analysis if the model is estimated before and after an R&D program.²⁹ Generally, however, such productivity models have not been used for this purpose but rather for justification of future public-sector R&D programs to support firm-level investments in R&D.

Benchmarking Analysis

Benchmarking analysis involves the comparison of the performance of firms affected by a public-sector R&D program relative to a theoretical objective or goal, to the best practice of all of the firms being studied, or to some other exemplary standard. For example, if the performance of k firms affected by the public-sector R&D program is denoted as $P_{i, i=1 \text{ to } i=k}$, and if the theoretical objective or goal of the program is for firm performance to reach the level P^* ,³⁰ then the relevant comparison is the performance of each of the k firms to the benchmark P^* , that is the relevant comparison is between P^* and $P_{i, i=1 \text{ to } i=k}$.

Because benchmarking analysis is designed not only for impact analysis, but also for improved program management, k firm-specific indices can be calculated as $(P^* - P_i)$ and then each index can be compared to a set of firm characteristics. In other words, one could quantify the characteristics of firms that are related to their performance being "closer" (that is, $(P^* - P_i) > 0$ but relatively small in value) or "farther away" (that is, $(P^* - P_i) > 0$ but relatively large in value) from the theoretical objective or goal of the program.

As an example, the interaction between public sector R&D (mostly within universities) and firm performance has been called by Polt et al. (2001) a dimension of "industry-science relations (ISR)". In the 1990s

42 *The theory and practice of program evaluation*

throughout Europe two paradoxical trends were being observed: high performance in science and deteriorating industrial competitiveness. In response to this widely observed “European paradox”, the European Commission, DG Enterprise, and the Austrian Federal Ministry of Economy and Labor commissioned a benchmark analysis of ISR in eight EU countries (Austria, Belgium, Finland, Germany, Ireland, Italy, Sweden and the UK) and the United States and Japan. Examining multiple indicators on ISR, such as university–industry research collaboration, faculty consulting with industry, patent applications, and new business start-ups related to public research, Polt et al. (2001) analyzed factors that determined the highest level of university performance in each dimension of ISR. In other words, the performance of ISR across universities within each country was benchmarked against the most efficiently performing country in each dimension (for example, the most efficient countries in terms of new business start-ups related to public research are those that support campus-based infrastructures to provide management and financial support for start-ups).

Innovation Surveys

Innovation surveys, although referred to by Polt and Rojo (2002) and others (for example, Licht and Sirilli, 2002) as an evaluation methodology, are in our opinion a data collection tool that can be used for both an economic assessment and an economic impact analysis. Large publicly administered innovation surveys are an effective means to collect data related to various aspects of the innovation process from a national perspective.³¹ The surveyed units are generally firms or enterprises, but aggregation to an industry or national level is not uncommon.

European countries have sponsored broader and more detailed innovation surveys than has the United States. Noteworthy are the Community Innovation Surveys (CIS) throughout the European Union. The surveys began in 1992 and have continued more or less on a biannual basis. A hallmark of the CIS efforts is their breadth of coverage of multiple dimensions of the innovation process (for example sources of information that firms use to enhance their innovation strategy) and their documentation of a variety of government innovation policy schemes (Licht and Sirilli, 2002). The National Science Foundation’s Survey of Industrial Research and Development pales in comparison to the CIS efforts in terms of its ability to collect information to quantify multiple R&D spending dimensions of the innovation process.

Information collected through the CIS efforts lends itself to the estimation of econometric models like those in equations (2.5), (2.6) and (2.7)

above. Information collected through NSF's R&D survey could be used in econometric models like those in equations (2.5), (2.6) and (2.7), subject to accessibility to firm responses, but only when other external information is imposed on the model (for example, time series R&D data from the NSF survey can be used in a model like that in equation (2.5) to test for changes in R&D spending pre- and post-policy periods).

Expert Panels and Peer Reviews

Following the United Nations' definition (2005, p. 17):

Peer review can be described as the systematic examination and assessment of the performance of an entity by counterpart entities, with the ultimate goal of helping the reviewed entity improve its policy making, adopt best practices, and comply with established standards and principles.

The peer review process is widely used to assess the quality of scientific endeavors ranging from manuscripts submitted for publication in a scholarly journal to the social impact of a public sector program's research programs (for example, US Department of Energy, 1982, 1991; Office of Naval Research, 1989). To generalize, experts are asked to review public sector research projects on a number of dimensions. In the case of the US Department of Energy (1982) evaluation of basic energy sciences projects, several evaluation factors were considered, ranging from the scientific merit of the research to the expected impact of the research on the energy mission of the agency.

Ormala (1994), for example, notes that evaluation panels are used widely throughout Europe for public sector policy and program evaluation, especially those related to public sector R&D evaluation. One especially noteworthy effort was, according to Ormala, the Commission of the European Communities evaluation of the first European Community R&D Framework Program.³²

The National Academy of Sciences (1999, p. 38) recommended that federal agencies, in compliance with GPRA, should rely on expert review to assess the quality of basic research that they fund: "Federal agencies should use expert review to assess the quality of research they support, the relevance of that research to their mission, and the leadership of the research."

The Academy (1999, p. 39) was also of the opinion: "The most effective way to evaluate research programs is by expert review. The most commonly used form of expert review of quality is peer review. This operates on the premise that the people best qualified to judge the quality of research are experts in the field of research."

44 *The theory and practice of program evaluation*

CONCLUSIONS

Although prospective analyses are beyond the scope of this chapter, thinking about them adds useful perspective to the discussion of retrospective analyses. There are examples of economic impact analyses in Table 2.1 where the time series of expected benefits and costs are extended into the future because the NIST infrastructure technology being evaluated had a commercial lifetime extending beyond the date of the analysis. Nonetheless, those analyses are retrospective because they evaluate NIST programs that started in the past; the programs are expected to generate benefits into the future; prospective analyses would entail evaluations of potential public investment in new projects.

In thinking about altogether new technology programs and projects, according to Martin (1995, pp. 139–40):

It is widely agreed that new technologies . . . will have a revolutionary impact on the economy and society over the coming years. . . . [Government is thus under pressure to answer the question:] how can one identify the most promising research areas and the emerging technologies on which to concentrate resources and, hence, derive the fullest socio-economic benefits? . . . Foresight represents one response to these pressures. [Foresight] is the process involved in systematically attempting to look into the longer-term future of science, technology, the economy and society with the aim of identifying the areas of strategic research and the emerging generic technologies likely to yield the greatest economic and social benefits.

Technology assessment/economic impact, in contrast, is a process to project the economic impacts associated with the allocation of resources toward new science and technology.

Foresight differs from technology assessment/evaluation in at least one important dimension. Foresight is a management tool, not an assessment or evaluation tool, method, or methodology.³³ In contrast, technology assessment/evaluation is a policy tool used to approximate the economic impacts associated with science and technology resource allocations.³⁴

In the United States, technology assessment as a policy tool gained visibility through the Congressional Office of Technology Assessment (active from 1974–95).³⁵ Technology assessment/evaluation is a tool that could effectively be used for prospective economic impact analyses, but prospective analyses have not been conducted to any great extent, and are conspicuously absent from the NIST analyses summarized in Table 2.1.³⁶

Because the policy tool of technology assessment/evaluation examines resource allocations to science and technology, a key question for such

evaluations of economic impact, whether retrospective or prospective, is: Why should government rather than the private sector fund and perform the activity being evaluated? The analyses reviewed in Table 2.1 answer this fundamental question in a variety of ways depending on the context of each particular analysis, but the short answer is “market failure”. Each of the analyses in Table 2.1, based on discussions with industry experts, found significant sources of market failure underlying the need for the publicly financed and publicly performed infratechnology investment being evaluated.

Whether economic impact analyses are retrospective or prospective, identification of the market failure that provides the reason for public investment plays an integral role in the analyses themselves because understanding the market failure helps to formulate the crucial counterfactual scenario that must be clearly stated if benefits from the public investment are to be identified. Are barriers to technology so severe that the private sector would not even attempt to replicate the public investment (extant investment for a retrospective analysis and proposed investment in a prospective analysis)? In that case, the counterfactual scenario is the status quo ante technology absent the public investment, and a “traditional” economic impact analysis in the sense of Griliches (1958) is appropriate.

Or, would the private sector, absent the public investment and despite barriers to technology, undertake investment to provide a substitute for the infratechnology generated with the public investment? In that case, what Link and Scott (2011) call “counterfactual analysis” is the appropriate approach for an economic impact analysis, with the benefits of the public’s investment being the costs avoided by the private sector—the investment costs the private sector would have incurred to replicate the public infratechnology and any shortfall in the value of the counterfactual privately developed infratechnology from the value of the technology developed publicly (often by means of the cooperative efforts of NIST and industry in public–private partnership).

As explained earlier, the “counterfactual analysis” is equivalent to the “traditional” approach when that approach is appropriate for the evaluation of publicly financed and performed R&D, that is, when barriers imply the private sector would not attempt to provide the infratechnology. Thus, the “counterfactual analysis” actually covers both approaches. As the discussions in the second and third sections explain and as the economic impact analyses in Table 2.1 illustrate, in practice both approaches are useful—both the traditional approach, with its counterfactual of the status quo ante technology, and the approach for which the counterfactual scenario entails private investment to provide the infratechnology that the public program provides.

46 *The theory and practice of program evaluation*

The timing of an economic impact analysis is important—sufficient time must have passed for the effects of the infratechnology to be observed. For example, the analysis of the Thermocouple Calibration Program, Planning Report 97-1, was premature. The analysis was supposed to estimate the benefits of a new international standard for calibration data. But, the international body did not understand that industry was committed and happy with the existing standard; that is, industry did not yet need to shift to the new standard. Similarly, in the analysis of role-based access control, Planning Report 02-1, the analysis was premature because there had not yet been sufficient market penetration of the infratechnology being evaluated.

The timing of an economic impact analysis is also important because the time during which the greatest effects of the infratechnology were realized must not be so far in the past that gathering good information about the benefits is not possible. An analysis can be done too late; in contrast to the thermocouple analysis, that of NIST's Cholesterol Standards Program, Planning Report 00-4, was done too late because as a practical matter it is not typically possible to gather the quantitative information about benefits in a period about which the institutional memory of respondents—even if appropriate respondents can be found—has faded.

Benefit data for most economic impact analyses came from detailed interviews or surveys, but possibilities for using published data (in conjunction with or in place of gathering new survey data) should be considered and used when appropriate information is available, as in the analysis of NIST's Data Encryption Standard (DES) Program, Planning Report 01-2.

When benefit data are gathered by survey, the study team should be given access to the addresses and other contact information for the entities surveyed. For example, in the analysis of the Baldrige National Quality Program, Planning Report 01-3, the American Society for Quality (ASQ) sent the survey to its members, but the study team did not have access to the email list of the recipients of the survey. The predictable result was the low response rate because although the ASQ could follow up with a mass email to its members, the opportunity for the study team to implore each member to participate was absent.

Before an infratechnology investment is undertaken, NIST should be sure that a significant underinvestment gap exists. The relatively small reduction in R&D costs found in the analysis of NIST's investments in superfilling research, Planning Report 08-1, suggests as a methodological lesson that when deciding to undertake an investment project, project managers should be sure that there is a significant underinvestment gap.

NIST project managers can maintain key contacts with the users of

their projects' outputs at all stages of the supply chain and communicate with those contacts about the types of data that will be needed to estimate benefits. Ideally such data would be routinely gathered in real time on an ongoing basis. The availability of such data would allow periodic evaluations documenting impacts, thus facilitating NIST's ability to make possible adjustments that would better serve industry's needs and also allowing NIST to provide quality information about performance to support the mandate of the GPRA.

When NIST initiates a new program and also in real time as program directors and project managers gather information useful for evaluating their projects, they can also work with their industry contacts to develop understanding of the nature of the market failure to which the public provision of infratechnology investment is responding. Developing that understanding will inform the direction of the infratechnology investment in a way that increases its beneficial economic impact.

A project manager's key contacts in industry throughout the relevant supply chain can also provide access to greater numbers of industrial respondents to surveys about the economic impact of the project. Key contacts could be asked on an ongoing basis about other industrial users of the infrastructure technology, and project managers could maintain lists of industrial users in real time and on an ongoing basis. With larger numbers of respondents to provide estimates of benefits of particular types when surveys of industry are administered, it will be possible to produce formal statistical confidence intervals for the estimated benefits.

The other techniques, discussed in the fourth section, relevant for an economic impact analysis can be used to complement the approach of the economic impact analyses that has been discussed in the third section.

Should NIST standardize the supply chain used in the economic impact analyses? In our opinion, the answer is no. Instead, the place to begin is with the abstract, general understanding of how the economic impact of NIST's infrastructure technologies occurs throughout the idealized supply chain based on the four-tier model identified in this chapter. Then, with that idealized model in mind, the project leaders deciding on infratechnology investments or the study team planning an economic impact analysis can develop the actual supply chain where benefits are realized.

Should NIST replace the current approach that is often driven by the availability of data for assessment of benefits with a more standardized application of the cost/benefit methodology? As a practical matter, the answer in our opinion is no, but future economic impact analyses should continue to identify qualitatively, as the past analyses have done, all of the benefits and costs, ensuring that the metrics presented are conservative and explaining why that is so. The current approach can only be characterized

48 *The theory and practice of program evaluation*

as data driven in the sense that all data that can be reliably collected and reasonably quantified is included in the calculations of metrics, but the analyses are not data driven in that they try to identify qualitatively all of the benefits and costs. Thus, the quantitative metrics are data driven (as of course must be the case), but the overall evaluation is not. The result is that estimated cost–benefit ratios can differ for reasons other than differential impacts between research programs, and for that reason it is crucial that each analysis continue to identify all effects qualitatively and state clearly what is missing from the estimated evaluation metrics and emphasize that readers should not simply compare benefit-to-cost ratios or other metrics across the analyses. Instead, the conservative metrics, along with a good understanding of what is missing from the metrics, must be in mind when comparing the economic impacts of the analyses. While some analyses are able to quantify impact for a single tier of the supply chain, others quantify benefits for three tiers. Some analyses are able to extrapolate benefits to draw conclusions regarding impacts at the industry level while others are able only to characterize the direct impacts on survey respondents. Some analyses are able to quantify future benefits while others must restrict the analysis to quantification of past benefits even if the NIST technical outputs remain state-of-the-art. The end point for benefits to be realized in the future must of course be determined by a conservative estimate of the future impact horizon, and the likely proportion of future benefits that is conservative may be very different across analyses given the information available.

Summarizing, by following the recommendations about gathering benefit and cost data for the economic impact of an infratechnology project on all affected parts of the supply chain, NIST will have, to the extent practicable, the type of data that it needs to evaluate the project. As a review of the previous economic impact analyses shows, the type of data will often differ by project and by the tier of the supply chain affected. In our experience, benefits extending into the future are even more difficult for industry's beneficiaries to estimate, but if the recommendation for ongoing, real-time gathering of benefit and cost data are followed, a better fix on the benefits and costs and the expected commercial lifetime will be available than has been available for the extant analyses. The expected commercial lifetimes vary across the different infrastructure technologies because the pace of technological change varies widely across different industries and even within industries across different types of infratechnology applications. Real-time and ongoing data collection will also make possible more accurate extrapolations of benefits from the respondents to the industry-wide level. The previous economic impact analyses have differed in the extent to which they extrapolate, simply because it is not

always possible to develop a good understanding of how representative an individual respondent is of the entire industry. The real-time and ongoing maintenance of industry contacts and data collection will, because such benefits are often more difficult to quantify, also enable more analyses that measure the economic impact of basic measurement research (as contrasted with the evaluation of the calibration services that take such science as given) such as in the evaluation of the investment in developing the basic science of wavelength references for optical fiber communications in the analysis of ATP's Intramural Research Awards Program (in NIST GCR 04-866). As the analysis shows, the industrial benefits from developing basic measurement science can be immensely important for innovation and the productive evolution of technology.

NOTES

1. The concept of fiscal accountability in the United States is rooted in the fundamental principles of representation of the people, by the people. Following Link and Scott (1998, 2011), issues of fiscal accountability can be traced at least to as far back as President Woodrow Wilson's reforms, and in particular to the Budget and Accounting Act of 1921. This Act of 10 June, 1921 not only required the President to transmit to Congress a detailed budget on the first day of each regular session, but also it established the General Accounting Office (GAO) to settle and adjust all accounts of the government. More recently, the Government Performance and Results Act (GPRA) of 1993 stated that: "Congressional policymaking, spending decisions and program oversight are seriously handicapped by insufficient attention to program performance and results." GPRA mandated that the head of each government agency submit to the Director of the Office of Management and Budget (OMB): "a strategic plan for program activities... [The plan] shall contain . . . a description of the program evaluations used in establishing or revising general goals and objectives, with a schedule for future program evaluations." GPRA was at the time of its passage: "the most significant advance in bringing accountability to government programs... Unfortunately, the implementation of this law has fallen short of its authors' hopes. Agency plans are plagued by performance measures that are meaningless, vague, too numerous, and often compiled by people who have no direct connection with budget decisions" (President's 2004 Budget, pp.48–9). In response, President George W. Bush's 2004 budget presented a new assessment, not evaluation, tool: the Program Assessment Rating Tool (PART). There have also been policy discussions to replace PART with a new performance improvement and analysis framework called Putting Performance First (Orszag, 2009a).
2. Many point in the United States to President George H.W. Bush's 1990 US Technology Policy as our nation's first formal domestic technology policy statement, although, as Kahin and Hill (2010) point out, US innovation policies trace as far back as President George Washington's 1790 address to Congress. However, US Technology Policy, like its predecessors, has failed to articulate a foundation for government's role in supporting innovation and technology. Rather, it has implicitly assumed that government has a role, and then sets forth the general statement (Executive Office of the President 1990, p.2): "The goal of US technology policy is to make the best use of technology in achieving the national goals of improved quality of life for all Americans, continued economic growth, and national security." President William Clinton took a major step

50 The theory and practice of program evaluation

forward from this 1990 policy statement in his 1994 “Economic report of the President” by articulating first principles about why government should be involved in the technological process (Council of Economic Advisers, 1994, p. 191): “The goal of technology policy is not to substitute the government’s judgment for that of private industry in deciding which potential ‘winners’ to back. Rather, the point is to correct market failure”. Relatedly, Martin and Scott (2000, p. 438) observed: “Limited appropriability, financial market failure, external benefits to the production of knowledge, and other factors suggest that strict reliance on a market system will result in underinvestment in innovation, relative to the socially desirable level. This creates a *prima facie* case in favor of public intervention to promote innovative activity.” For a detailed discussion about the economic justification for government’s role in the innovation process see Link and Scott (2005, 2011).

3. NIST defines an evaluation in this broad manner. What NIST refers to as an economic impact analysis is to many, including Link and Scott (1998, 2011) and the references therein, an economic impact evaluation. However, the NIST terminology is maintained throughout this chapter.
4. It is important to distinguish between assessment and impact analysis, although the terms are frequently used interchangeably. A distinction is offered herein, as discussed below, to define the boundaries of this chapter, although (for example in the titles of the economic impact analyses reviewed in Table 2.1) the distinction is not made uniformly throughout this chapter. Policy assessment is based primarily on the criterion of effectiveness, not efficiency, and the question asked is: Has the R&D program met its stated goals and objectives, and have its designated outputs been achieved?

Also, regarding effectiveness versus efficiency, some areas of government, notably regulation, have the explicit requirement to demonstrate efficiency.

5. This section draws directly from Link and Scott (2011), which is based in part on Link and Scott (1998). Both sources refer to the following discussion under the rubric of an economic evaluation as noted in note 3 above.
6. It is important to emphasize the difference between the terms *methodology* and *method*. The terms are often used interchangeably, although it is incorrect to do so. A *methodology* is the theoretical foundation or practices within a discipline that determine or guide how to engage in an inquiry; a *method* is a tool or technique used to implement the inquiry.
7. The Mansfield et al. (1977) seminal article applied the Griliches methodology to private sector innovations, and it expanded the methodology to accommodate imperfect competition in certain circumstances.
8. These metrics are discussed in the third section.
9. Additionally, for market settings more complicated than the simplest model, the Griliches model accounts for producer surplus, measured as the difference between the price the producers receive per unit and the actual marginal cost, summed over the output sold, minus any fixed costs. Social benefits are then the streams of new consumer and producer surpluses—economic value above and beyond the opportunity costs of the resources used to create value, while private benefits for a firm that invests in innovation are the portions of the streams of producer surplus appropriated by the investor. Not all of the appropriated producer surplus is necessarily new because the surplus gained by one producer might be cannibalized from the pre-innovation surplus of another producer or from pre-innovation consumer surplus. Social and private costs will, in general, also be divergent.
10. See specifically pages 12–16 in Link and Scott (1998).
11. Observe that this does not assume that the private sector would make the same level of investment, and it does not assume that the resulting R&D output would be of equal quality. The costs avoided are the costs that the private sector would have spent on developing replacement technology and the loss in value because the replacement technology was not of equal quality to the technology developed with the public R&D. Observe that if the barriers to technology and market failure are sufficiently severe that

the private sector would not have attempted to replace the public R&D investment at all, then the counterfactual is the status quo ante technology and there are no private sector replacement costs and only the lost value from the shortfall in quality of the technology used with the public investment—exactly the counterfactual of the traditional analysis.

12. This description of infratechnologies draws on Tassey (2007). He notes (pp. 112–13): “if you cannot measure, you cannot do R&D; if you cannot test, you cannot control the production process for yield and quality; if you cannot pass a variety of complex data from machine to machine and from company to company, you cannot control cost or schedule; and if you cannot provide a customer with standardized test data for a product, you cannot assure that buyer that performance specifications have been met.”
13. The causes for an underinvestment in R&D or in technology are discussed in detail in Tassey (2007) and Link and Scott (2011). Link and Scott (2011) discuss eight factors or barriers to technology that lead to technological market failure: (1) High technical risk means the outcomes of the firm’s R&D might not be technically sufficient to meet its needs. This might cause market failure, given that when the firm is successful, the private returns fall short of the social returns. An underinvestment in R&D will result. (2) High technical risk can be related to high commercial or market risk, when the requisite R&D is highly capital intensive. Such investments could require too much capital for a firm to fund the outlay; thus, the firm will not make the investment, even though it would be better off if it had been able to finance the investment, and so would society. (3) Many R&D projects are characterized by a lengthy time interval until a commercial product reaches the market. The time expected to complete the R&D, and the time until commercialization of the R&D results, are long; thus, the realization of a cash flow is distant and in conjunction with differing private and social discount rates can result in market failure. (4) It is not uncommon for the scope of potential markets to be broader than the scope of the individual firm’s market strategies, so the firm will not perceive economic benefits from all potential market applications of the technology. (5) The evolving nature of markets requires investment in combinations of technologies that, if they existed, would reside in different industries that are not integrated. Because such conditions often transcend the R&D strategy of individual firms, such investments are not likely to be pursued. (6) The nature of the technology may make difficult the assignment of intellectual property rights. (7) Industry structure can raise the cost of market entry for applications of the technology. (8) Situations can exist where the complexity of a technology makes agreement with respect to product performance costly between buyers and sellers. Infrastructure technology investments by NIST can allow many such barriers to be overcome—reducing costs of entry or costs of agreement about performance so that the benefits outweigh NIST’s costs. Stated alternatively in the language of the “counterfactual method” described in the second section, the social cost for the investments in infratechnologies are lower with NIST and industry working together in public–private partnership than with the private sector attempting to accomplish the same ends without NIST.
14. These analyses were sponsored by the Program Office at NIST except for one that was sponsored by ATP. For additional impact assessments sponsored by ATP, see http://www.atp.nist.gov/eao/eao_pubs.htm.
15. This argument is developed in greater detail in Link and Scott (2011).
16. This statement is not intended as a criticism of the evaluators’ abilities or due diligence. Rather, it underscores the difficulty in collecting retrospective data subject to resource constraints.
17. Even when it is possible to interview the beneficiaries in a downstream tier of the supply chain, they are often unable to quantify the benefits they receive from NIST’s program. For example, in the analysis (Planning Report 97-1) of NIST’s Thermocouple Calibration Program, benefits for domestic users were not quantified because they are more indirect and difficult to quantify than the benefits for the wire suppliers (providing materials for thermocouple assemblies) and thermocouple suppliers.

52 The theory and practice of program evaluation

18. Even when all of the beneficiaries are interviewed, the numbers are often quite small. For example, in the analysis of the radiopharmaceutical research program at NIST (Planning Report 97-2), there are only seven respondents even though those seven respondents were 100 percent of the US manufacturers participating in the program.
19. One detailed example of the use of formal statistical confidence intervals in a journal article based on a NIST-sponsored analysis is Leech and Scott (2008).
20. Tassey (2003) refers to these metrics as corporate finance measures; they have traditionally been discussed in corporate finance textbooks' development of a firm's capital budgeting decision.
21. For the archetypal illustrative case to explain the concept of internal rate of return, the net cash flows of benefits minus the costs begin as negative and then become positive, with just a single reversal in the sign for the series of flows. There is then at most a single real solution for the internal rate of return. Of course, for actual investment projects there can be multiple reversals in the signs for the net cash flows, and then using the internal rate of return concept for a cost-benefit analysis requires some additional work with the time series of benefits and costs. For an actual example with an explanation of how to treat the issue, see Link and Scott (1998, p. 46).
22. Commenting on the 7 percent real discount rate, OMB (2003, p. 33) observed: "The 7 percent [real] rate is an estimate of the average before-tax rate of return to private capital in the US economy. It is a broad measure that reflects the returns to real estate and small business capital as well as corporate capital. It approximates the opportunity cost of capital, and it is the appropriate discount rate whenever the main effect of a regulation is to displace or alter the use of capital in the private sector. OMB revised *Circular A-94* in 1992 after extensive internal review and public comment." Further, OMB (2003, p. 33) observed: "The pre-tax rates of return better measure society's gains from investment. Since the rates of return on capital are higher in some sectors of the economy than others, the government needs to be sensitive to possible impacts of regulatory policy on capital allocation." However, OMB (2003, p. 33) also observed: "The effects of regulation do not always fall exclusively or primarily on the allocation of capital. When regulation primarily and directly affects private consumption (e.g., through higher consumer prices for goods and services), a lower discount rate is appropriate." Hence, if one were evaluating a policy where, instead of alternative uses of investment capital in public R&D investment decisions, the issue evaluated were a regulatory policy (for example, for health care) that would directly and primarily affect the stream of real income to consumers (for example, alternative health plans with streams of different magnitudes and different timings), then the OMB has directed (OMB, 2003, pp. 33-4) that "for regulatory analysis" (p. 34), rather than an evaluation of an investment, the real discount rate of 3 percent should be used and then compared to the results using the 7 percent real discount rate. OMB explains that for consumers' decisions, 3 percent better approximates their real rate of time preference.

OMB (2003, p. 33) explicitly stated that the 7 percent real required rate of return that is based on the average rate of return to private capital investment is "a default position"; yet, the market failure story recognizes that for investments (not just "regulatory policy") the social rate of return and the private rate of return can (and are expected to) diverge, with the social required rate of return being less than the private hurdle rate. As it turns out, in practice, a 7 percent social hurdle rate for public investments is not inconsistent with that logic because the 7 percent is based on the average. However, for the R&D investment projects we have evaluated in case studies, the firms report higher private hurdle rates. OMB appears to be taking the least controversial approach by using for the social hurdle rate for investments an average return for private capital investments and by advising consideration of the variance in private returns in different activities. Clearly, as we have noted, there is no reason society should be constrained in its assessments of value by prices determined in markets where there are market failures and the prices give the wrong signals. Hence, the private rate of return on investment should not be expected to equal the social opportunity cost

of investment funds; the private rates of return may be based on prices that do not reflect social value. We know that with positive externalities such as non-appropriated spillovers that benefit those who did not invest, social rates of return can be high when private rates of return are low. Moreover, the private rate of return can be high even when the social rate of return is low or even negative. For example, in the context of R&D investment, the results of a privately profitable R&D investment may simply cannibalize previously existing economic surplus, causing the investment to have a negative social rate of return for a period of time. OMB's approach is a solution in the absence of a practical way to determine what the theoretical social hurdle rate should be in any given situation.

23. These techniques are widely cited by the European Commission (Polt and Rojo, 2002) as methodologies. In our opinion they are a potpourri of methodologies, methods and data collection tools rather than all being methodologies, and they vary greatly in their relation to the theory of value and opportunity costs that underlies economics-based methodologies. All of these techniques are discussed because of their general visibility.
24. As Jaffe (2002) explains, identifying the effect of the public program is difficult.
25. There are many issues related to how one defines a matched firm (for example, size, level of own R&D, industry), but that discussion is beyond the scope of this chapter.
26. The specification in equation (2.5) is simplified for purposes of an introductory explanation because it assumes the R&D program's impact occurred at time t^* and that impact remained constant through subsequent time periods. More sophisticated variations of equation (2.5) are possible.
27. Again, if the program is hypothesized to affect the impacts of control variables (that is, if it will have "slope effects" as well as "intercept effects"), then interaction terms that multiply E with each affected control variable would be entered in the specification.
28. This finding could be interpreted as showing that public R&D complements private R&D rather than substitutes for it (that is, crowds it out). See David et al. (2000) and the references therein.
29. Hall et al. (2010) provide an excellent review of the academic literature on such productivity models.
30. P^* could be the performance level of the most efficient firm in the sample, the average performance of all firms in the sample, or some other performance index related to best practices.
31. Such publicly administered surveys also have economic value in the sense that they are within the public domain (controlled for confidentiality issues), and thus, analyses based on the collected data can be replicated.
32. The Framework Program was initially designed to assist small- and medium-sized enterprises, technology-based enterprises in particular, compete in globalized markets.
33. The importance of foresight as a strategic management tool is not independent of government's role in the innovation process, which is based on the economic concept of market failure. Market failure refers to the market—including both the R&D-investing producers of a technology and the users of the technology—underinvesting, from society's standpoint, in a particular technology or technology application. Such underinvestment occurs because conditions exist that prevent organizations from fully realizing or appropriating the benefits created by their investments. To elaborate on the concept of market failure, consider a marketable technology to be produced through an R&D process where conditions prevent the R&D-investing firm from fully appropriating the benefits from technological advancement. Other firms in the market, or in related markets, will realize some of the benefits (economic profits—revenues in excess of the opportunity costs of the resources used by these other firms) from the innovation, and of course consumers will typically place a higher value on a product than the price paid for it. The R&D-investing firm will then calculate, because of such conditions, that the marginal benefits it can receive from a unit investment in such R&D will be less than could be earned in the absence of the conditions reducing the appropriated benefits of R&D below their potential, namely the full social benefits. Thus, the R&D-investing

54 *The theory and practice of program evaluation*

firm might underinvest in R&D, relative to what it would have chosen as its investment in the absence of the conditions.

There are a number of factors that can explain why a firm will perceive that its expected private rate of return will fall below its hurdle rate, even when the social rate of return exceeds the social hurdle rate. See Link and Scott (2011) for a more detailed explanation, and see note 13 for enumeration and brief discussion of eight factors that constitute barriers to innovation and technology. The eight factors lead to a private underinvestment in R&D.

34. The importance of technology assessment/evaluation as a policy tool is not independent of identifying which of the barriers to innovation and technology brought about the market failure and then understanding the extent to which allocated resources will overcome those barriers.
35. Schot and Rip (1997) offer examples of technology assessment used in the Netherlands.
36. An exception is Leech and Scott (2008) which provides an economics-based prospective analysis to look ahead to the expected impact of intelligent machine technology and to explain that new public infratechnology investments would increase that future impact.

REFERENCES

- Busom, Isabel (2000), 'An empirical evaluation of the effects of R&D subsidies', *Economics of Innovation and New Technology*, **9**, 111–48.
- Council of Economic Advisers (annual), 'Economic report of the President', Washington, DC: US Government Printing Office.
- David, Paul A., Bronwyn H. Hall and Andrew A. Toole (2000), 'Is public R&D a complement for private R&D?', *Research Policy*, **29**, 497–529.
- Executive Office of the President (1990), 'US technology policy', Washington, DC: Office of Science and Technology Policy.
- Griliches, Zvi (1958), 'Research costs and social returns: hybrid corn and related innovations', *Journal of Political Economy*, **66**, 419–31.
- Hall, Bronwyn H., Jacques Mairesse and Pierre Mohnen (2010), 'Measuring the returns to R&D', in Bronwyn H. Hall and Nathan Rosenberg (eds), *The Economics of Innovation*, Amsterdam: Elsevier, pp. 1033–82.
- Jaffe, Adam B. (2002), 'Building programme evaluations into the design of public research-support programs', *Oxford Review of Economic Policy*, **18**, 22–34.
- Kahin, Brian and Christopher T. Hill (2010), 'United States: the need for continuity', *ISSUES in Science and Technology*, **26**, 51–60.
- Leech, David P. and John T. Scott (2008), 'Intelligent machine technology and productivity growth', *Economics of Innovation and New Technology*, **17**, 677–87.
- Licht, Georg and Giorgio Sirilli (2002), 'Innovation surveys', in Gustavo Fahrenkrog, Wolfgang Polt, Jaime Rojo, Alexander Tübke and Klaus Zinöcker (eds), *RTD Evaluation Toolbox*, Seville: European Commission, pp 71–81.
- Link, Albert N. and John T. Scott (1998), *Public Accountability: Evaluating Technology-based Institutions*, Norwell, MA: Kluwer Academic Publishers.
- Link, Albert N. and John T. Scott (2001), 'Public/private partnerships: stimulating competition in a dynamic market', *International Journal of Industrial Organization*, **19**, 763–94.
- Link, Albert N. and John T. Scott (2005), *Evaluating Public Research Institutions: The US Advanced Technology Program's Intramural Research Initiative*, London: Routledge.
- Link, Albert N. and John T. Scott (2011), *Public Goods, Public Gains: Calculating the Social Benefits of Public R&D*, New York: Oxford University Press.
- Mansfield, Edwin, John Rapoport, Anthony Romeo, Samuel Wagner and George Beardsley (1977), 'Social and private rates of return from industrial innovations', *Quarterly Journal of Economics*, **91**, 221–40.

- Martin, Ben R. (1995), 'Foresight in science and technology', *Technology Analysis and Strategic Management*, **7**, 140–68.
- Martin, Stephen and John T. Scott (2000), 'The nature of innovation market failure and the design of public support for private innovation', *Research Policy*, **29**, 437–47.
- National Academy of Sciences (1999), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, Washington, DC: National Academy of Sciences.
- Office of Management and Budget (OMB) (1992), 'Circular No. A–94: Guidelines and discount rates for benefit–cost analysis of federal programs', Washington, DC: Government Printing Office.
- Office of Management and Budget (OMB) (2003), 'Circular No. A–4: Regulatory analysis', Washington, DC: Government Printing Office.
- Office of Naval Research (1989), 'Significant accomplishments: Office of Naval Research sponsored programs', Washington, DC: Office of Naval Research.
- Ormala, Erkki (1994), 'Impact assessment: European experience of qualitative methods and practices', *Evaluation Review*, **18**, 41–51.
- Orszag, Peter R. (2009a), 'Testimony of Peter Orszag before the Subcommittee on Financial Services, Committee on Appropriations, on May 20, 2009', Washington, DC: House of Representatives, US Congress.
- Orszag, Peter R. (2009b), 'Memorandum for the Heads of Executive Departments and Agencies', Office of Management and Budget, Executive Office of the President, 7 October.
- Polt, Wolfgang and Jaime Rojo (2002), 'Evaluation methodologies', In Gustavo Fahrenkrog, Wolfgang Polt, Jaime Rojo, Alexander Tübke and Klaus Zinöcker (eds), *RTD Evaluation Toolbox*, Seville: European Commission, pp.65–70.
- Polt, Wolfgang, Christian Rammer, Helmut Gassler, Andreas Schibany and Doris Schartinger (2001), 'Benchmarking industry–science relations: the role of framework conditions', *Science and Public Policy*, **4**, 247–58.
- President's 2004 Budget (2003), available at: <<http://www.gpoaccess.gov/usbudget/fy04/pdf/budget/performance.pdf>>.
- Schot, Johan and Arie Rip (1997), 'The past and the future of constructive technology assessment', *Technological Forecasting and Social Change*, **54**, 251–68.
- Tassey, Gregory (1997), *The Economics of R&D Policy*, Westport, CT: Quorum Books.
- Tassey, Gregory (2003), 'Methods for Assessing the Economic Impacts of Government R&D', NIST Planning Report 03-1, Gaithersburg, MD.
- Tassey, Gregory (2007), *The Technology Imperative*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- US Department of Energy (1982), 'An assessment of the basic energy sciences program', Washington, DC: Office of Energy Research.
- US Department of Energy (1991), 'Procedures for peer review assessments', Washington, DC: Office of Energy Research.
- United Nations Development Programme (2005), 'Peer assessment of evaluation in multilateral organisations', Koege, Denmark: Ministry of Foreign Affairs of Denmark.