RAND EUROPE

THE ARTS
CHILD POLICY
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND HOMELAND SECURITY
TRANSPORTATION AND INFRASTRUCTURE
WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Browse Books & Publications

Make a charitable contribution

For More Information

Visit RAND at www.rand.org

Explore RAND Europe

View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

A historical reflection on research evaluation studies, their recurrent themes and challenges

Sonja Marjanovic, Stephen Hanney, Steven Wooding

Prepared for Project Retrosight

RAND EUROPE

The research described in this report was prepared for the Project Retrosight.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2009 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND Web site is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (http://www.rand.org/publications/permissions.html).

Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Contents

List of figures iii
List of tables iv

Summary 5

CHAPTER 1 **A historical overview of research evaluation studies 6**
1.1 Early landmark studies: the 1960s and 1970s 7
1.2 The evolution of studies concerned with innovation processes and outputs: from the 1980s to the present day 12
1.2.1 Process-oriented evaluation studies 12
1.2.2 Output-oriented evaluation studies 13

CHAPTER 2 **Methodological developments: case study-based approaches and conceptual frameworks in evaluation research 18**
2.1 The Payback framework 23
2.2 Economic and Social Research Council framework 24
2.3 Royal Netherlands Academy of Arts and Sciences framework 24
2.4 Framework for impact segmentation by audience 24
2.5 US Agency for Healthcare and Research Quality impact model 25

CHAPTER 3 **Consistent themes in research evaluation studies 26**
3.1 Scientific accountability and justification of investments into research 26
3.2 Lack of convincing evidence on success factors 27
3.3 Nature of the innovation process 27
3.4 Factors influencing innovation processes 28
3.4.1 Funding mechanisms and research settings 28
3.4.2 The relative contributions of basic and applied science to innovation 29
3.4.3 Knowledge-exchange environments 30
3.5 Diversity and range of benefits from research 30

CHAPTER 4 **Challenges in research evaluation studies 32**
4.1 Apparent contradictions between studies as an artefact of research design 32
4.1.1 Types of innovations examined 33
4.1.2 Timeframes 33
4.2 Biases in the selection of cases 34

4.3 Lack of clarity and unity in the definitions of concepts 34
4.4 Challenges to the ability to replicate methodology: unclear descriptions of techniques used for data collection and analysis 36
4.5 The challenge of setting boundaries in research 36
4.6 Challenges in the attribution of impacts 37
4.7 Issues of sector idiosyncrasies 38

CHAPTER 5 **Conclusion 39**

Bibliography 41
Appendix 48
A summary of key studies and their insights 48

List of figures

Figure 1. Time distributions in generational improvements and in the utilisation of research and development (R&D) events 8

Figure 2. dvanced Technology Program's evolving use of methods over its first decade 21

Figure 3. Advanced Technology Program Evaluation Logic Model 22

Figure 4. The research process in the Payback framework 23

Figure 5. Comparison of definitions used for types of research 35

List of tables

Table 1. Battelle study findings: the contributions of mission-oriented and non-mission-oriented research to innovation 9

Table 2. Advantages and limits of different evaluation methods 19

Summary

This report critically examines studies of how scientific research drives innovation which is then translated into socio-economic benefits. It focuses on research evaluation insights that are relevant not only to the academic community, but also to policymakers and evaluation practitioners – and particularly to biomedical and health research funders. First, it provides a predominantly descriptive *historical overview of some landmark studies* in the research evaluation field, from the late 1950s until the present day, and highlights some of their key contributions. Then, it reflects on the historical overview analytically, in order to discuss some of the *methodological developments* and *recurrent themes* in research evaluation studies. The report concludes by discussing the *enduring challenges* in research evaluation studies and their implications.

We emphasise that this report does not address *all* of the key studies in the research evaluation field. The evaluation literature today is so extensive that a selective approach is necessary to focus on those studies that we feel provide the most valuable insights in the context of *biomedical and health research* evaluation. However, we do go back in history to revisit some early landmark studies outside of the biomedical or health research sectors, which had a particularly strong influence on informing evaluation thinking, policy and practice.

We are interested in the evaluation of research projects and programmes occurring in universities, other public or private research settings or clinical environments. As Georghiou and Roessner (2000) emphasise: all evaluations are social processes, occurring in specific settings and conditions and addressing diverse objectives and stakeholder needs. The level of aggregation and the objectives being evaluated will influence the choice of an evaluation approach. A detailed examination of the evaluation of large-scale European collaborative research and development initiatives or national and European level innovation systems is beyond the scope of this report. Neither do we address the evaluation of programmes specifically aiming to encourage technology adoption and diffusion. This report also explicitly focuses on the evaluation of research outputs and impacts, rather than the evaluation of policy effectiveness.

Lastly, this report presents a historical context for the Project Retrosight. Retrosight is a multinational study investigating diverse impacts from cardiovascular disease research in three countries (Australia, Britain and Canada) over the last 20 to 25 years and the pathways through which the impacts are produced. The project is being led by RAND Europe and the Health Economics Research Group (HERG) at Brunel University.

CHAPTER 1

A historical overview of research evaluation studies

Evaluation can be defined as "a systematic and objective process designed to assess the relevance, efficiency and effectiveness of policies, programmes and projects" (Fahrenkrog et al., 2002: 16). There are four broad rationales for conducting research evaluation:

1 to increase the accountability of researchers, policymakers and funding bodies, by making the research process, its outputs and impacts more transparent;
2 to steer the research process towards desired outcomes – for example, by evaluating whether milestones have been reached and taking remedial action when necessary;
3 to provide a means for advocacy, such as the use of research evaluation outputs as a means of signalling the ability to conduct research or to fund it, based on past performance;
4 to provide an input into the research management process, through better understanding and learning from past experience of a research process. (For an alternative list, see Georghiou and Larédo, 2005.)

Interest in understanding the contributions of research to society blossomed in the 1960s and 1970s. A series of large, early studies aimed to demonstrate the importance of research by looking at the genesis of many individual innovations. A major theme was the comparison of basic and applied research contributions (e.g. Jewkes et al., 1958; Sherwin and Isenson, 1967; Illinois Institute of Technology, 1968; Comroe and Dripps, 1976; Battelle Laboratories, 1973; cf. Smith 1987b). Interest gradually shifted to smaller studies focused on better understanding the process through which research contributes to innovation (i.e. research translation pathways and variables; e.g. Evered et al., 1987; Narin, 1989; Spangenberg et al., 1990; Arundel et al., 1995; Roessner et al., 1998). In the 1990s, the theme of measuring research impact also began to capture increased scholarly attention (e.g. Mansfield, 1991; Herbertz and Muller-Hill, 1995 Buxton and Hanney, 1996; Martin and Salter, 1996; Dawson et al., 1998; Hanney et al., 2003a, 2003b; Wooding et al., 2004b).

We expand on some landmark studies below. First, we consider the early large studies of the 1960s and 1970s, many of which are of a scale that has not been repeated since (Section 1.1). Then we address the evolution of the research evaluation field from the 1980s until the present day (Section 1.2), considering the body of smaller-scale studies concerned with various factors influencing innovation processes and with diverse types of outputs.

1.1 Early landmark studies: the 1960s and 1970s

Many early research evaluation studies of the 1960s and 1970s share two common features: a core interest in the comparative contributions of basic and applied research to innovation, and a case study-based methodology. Some studies also consider the impact of different types of funding and research settings on innovation outputs, as well as the organisational conditions which influence innovation success. Finally, some early attempts to quantify the returns and socio-economic benefits from investments into research also can be traced to this period.

One of the earliest studies, *The Sources of Invention* (Jewkes et al., 1958), adopted a case study approach to examine the evolution of 61 innovations from diverse scientific disciplines. These included streptomycin, insulin and the electron microscope (e.g. Jewkes et al., 1958, 1969, reviewed in Smith, 1987b). The authors were interested in the types of funding and research organisation that are most conducive to innovation. They found that diverse stakeholders (e.g. government groups, industry groups, individual researchers) and different forms of organising research (e.g. supporting individuals, teams organisations, collaboration) all generate innovation. Because one cannot predict what is likely to work in terms of funding and structure, the authors concluded that there is no single infallible formula for success, and that variety should be supported.[1]

This set the scene for Project Hindsight (Sherwin and Isenson, 1967), a US Department of Defense-funded study which aimed to provide a justification for the amount of investment in defence research. The project set out to assess the value-for-money gained from research investments by comparing the relative costs and effectiveness of different generations of military hardware, and by seeking to determine the role of research in contributing to the generational improvements. This large-scale study examined 710 events contributing to innovations in 20 defence systems, tracing back their sources over a 20-year period.[2] Researchers compared succeeding generations of military hardware equipment and catalogued the events (i.e. research advances) which had enabled the development of the latter generation. They found that for the 710 events examined, 91% were technology development-related and only 9% scientific discovery-related. Within the 9% science events, the majority (97%) were applied science conducted either to meet a Department of Defense or commercial need (rather than 'undirected', 'basic', 'non-mission-oriented' research). The study also found that, on average, it took 13 years to commence development[3] of a successor generation of military equipment, from its predecessor. During this 13-year time period, 57% of the requisite scientific and technological advances (events) happened (see Figure 1). The Hindsight findings appeared to emphasise the importance of applied research for innovation (as discussed further in Section 4.1). The authors also suggested that industrial and government laboratory settings are key loci for applied research.[4] Universities contributed to only 12% of the events examined (e.g. Sherwin and Isenson, 1967).

---

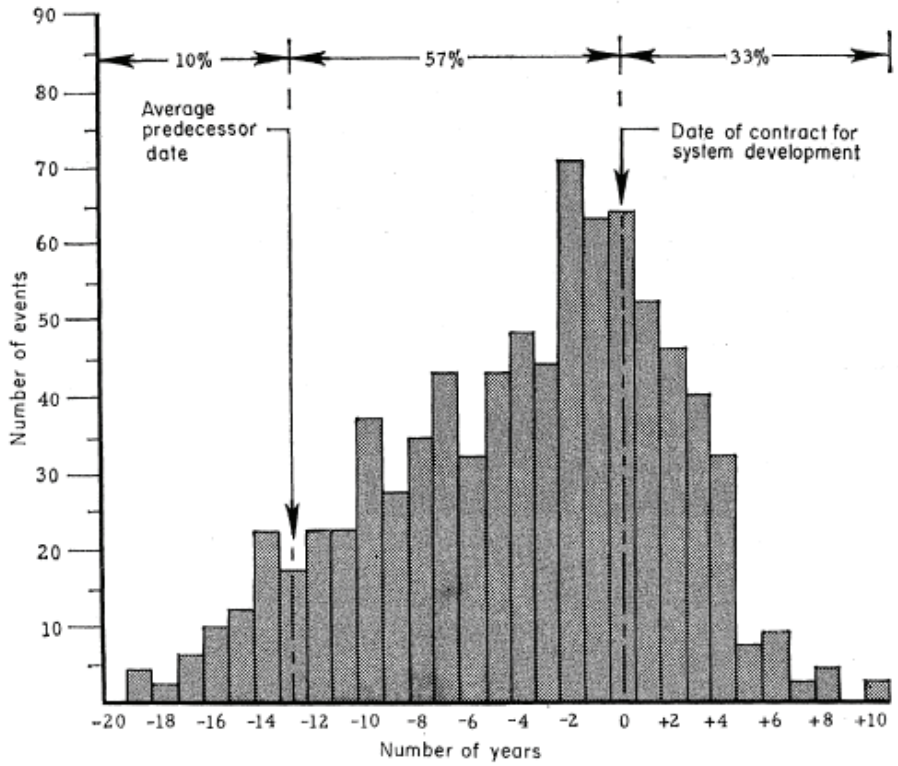
[1] Of the innovations they examined, 33 were generated by individual researchers, 21 resulted from either industrial or mixed government and industrial research, and seven could not be classified.

[2] The study expanded the equivalent of 40 professional person years of work, which would cost nearly GBP 5 million to fund today.

[3] According to the Hindsight study, 'commencing development' is the system development contract date. This can be a contract for design, engineering development or production.

[4] Industrial labs contributed to 47% and government labs to 39% of examined events.

Figure 1. Time distributions in generational improvements and in the utilisation of research and development (R&D) events



Source: Sherwin and Isenson (1967)

In response to Hindsight's apparent 'advocacy' for applied research, supporters and practitioners of *basic* research began to conduct a number of studies of their own. The National Science Foundation first commissioned the TRACES study (Illinois Institute of Technology, 1968) and then an extension of this work, the Battelle study (Battelle Laboratories, 1973). These projects investigated how basic (i.e. 'non-mission-oriented') research contributes to practical innovations of economic or social importance. Both TRACES and the Battelle studies focused not only on a longer history window than Hindsight, but also on a smaller number of innovations and predominantly 'blockbuster' ones.

The TRACES study traced back 340 events (e.g. research papers and other recognitions of scientific advancement) that contributed to five innovations over a 50-year time period. The innovations were magnetic ferrites, the videotape recorder, contraceptive pill, electron microscope and matrix isolation techniques. The study found that 70% of the contributing events were basic (i.e. non-mission oriented) science-related, 20% applied (i.e. mission-oriented) research and 10% attributed to developmental work (which the authors consider to be a separate category from research). The findings also suggested that universities are a key locus of research that feeds into innovations. The study assumed that all events have equal contributing value.

The early studies described above relied on identifying an innovation (for example, a research breakthrough or product) and building a family tree of the research events that fed into it. Events

included research papers and journal publications, but also extended to patents, research board meetings, other forms of recognition of scientific advancement and knowledge transfer between R&D laboratories. To compare the contributions of different types of research (e.g. basic and applied), events that arose from each type of research were counted, making the assumption that each event was of equal value (the 'zero order' assumption). A process that builds up a family tree of causality in this way is bound to increase the weight of whatever research occurred earlier in time. This is because the roots of the tree become increasingly wider as the ancestors of each idea are followed back in time.

In 1973, the Battelle study attempted to move beyond this zero-order assumption and traced back key decisive and non-decisive events that led to diverse blockbuster innovations, looking back 50 years. The study revisited three innovations from TRACES (oral contraceptives, magnetic ferrites, videotape recorders), and explored five new ones (electrophotography, input–output economic analysis, organophosphorous insecticides, hybrid grains and small grains). It also explored the managerial, human resource and technological determinants of innovation success: the authors suggested that the recognition of an unmet need or technological opportunity, internal management characteristics and technical entrepreneurs were important factors. The Battelle findings indicated that both basic and applied research are important for innovation. However, when one compares all events and (key) decisive events, the fraction of applied (mission-oriented) research events changes little, but the fraction of basic (non-mission-oriented) research events falls substantially (see Table 1). One of the limitations of the Battelle study is that the criteria for distinguishing between decisive and non-decisive events are not detailed and clarified.

Table 1. Battelle study findings: the contributions of mission-oriented and non-mission-oriented research to innovation

Type of event	Fraction of mission-oriented research	Fraction of non-mission oriented research	Fraction of development work
(Key) decisive	45%	15%	39%
All events	38%	34%	26%

Source: Smith (1987b) (figures directly quoted)

Following the TRACES and Battelle studies, Comroe and Dripps (1977) focused specifically on biomedical research and looked at the evolution of 10 key innovations in cardiovascular-pulmonary medicine and surgery. They traced the research advances which led to them as well as their organisational sources. Research advances were captured through basic and applied research publications. The authors identified 663 key articles relating to the top 10 innovations and found the following: 61.5% were basic research-related (considering basic research as that which seeks to discover a mechanism of function); 20% were descriptive clinical work; 16.5% were concerned with new techniques, apparatus or procedures; and 2% were a synthesis of earlier work.

Of the basic research articles, more than two-thirds were attributable to non-mission-oriented research (i.e. conducted purely for the sake of knowledge). They also found that 67% of the key articles stemmed from colleges and universities, medical schools or hospitals. In addition, Comroe and Dripps looked at time-lags between the publication of findings and their implementation in

clinical practice, finding that in 39% of the cases this lag was between 21 and 50 years, and in 18%, more than 50 years.

The Comroe and Dripps project has become a seminal study into the research basis of key improvements in cardiovascular care. However, one of its shortcomings is the lack of sufficient and reliable evidence on how innovations and their contributing publications were identified and selected and as key.

As with TRACES and the Battelle study, Comroe and Dripps was in essence a response to Project Hindsight and aimed to provide a justification for public investment in basic research:

> We believe that the first priority should be to earmark a generous portion of the nation's biomedical research dollars to identify and then to provide long-term support for creative scientists whose main goal is to learn how living organisms function, without regard to the immediate relation of their research to specific human diseases. (Comroe and Dripps, 1977: 4)

Much earlier in history, some of the most influential advocates of scientific research emphasised the importance of supporting basic science. For example, the French physicist and chemist Marie Curie commented:

> We must not forget that when radium was discovered no one knew that it would prove useful in hospitals. The work was one of pure science. And this is a proof that scientific work must not be considered from the point of view of the direct usefulness of it. It must be done for itself, for the beauty of science and then there is always the chance that a scientific discovery may become like the radium, a benefit for humanity. (Curie, 1921)

In addition to studies looking at the contributions of publically funded research, a parallel stream of studies looked at the origins of innovation in the commercial sector. For example, the SAPPHO project (Rothwell et al., 1974) attempted to explore organisational issues that influence whether innovations are commercially successful or not. Based on the analysis of 72 paired cases of 'successful' and 'unsuccessful' innovations, the authors suggested some determinants of success. These included: the innovator's understanding of user needs; efficiency in product development; manager characteristics; efficiency of communications; and the nature of marketing and sales efforts.

The findings of the SAPPHO project were enriched by Rothwell's later (1977) review of the current state of research on industrial innovation and technically progressive firms. In this review, Rothwell expanded on the range of success factors identified by the SAPPHO study, adding factors such as: embracing innovation as a corporate-wide task; efficient planning and management techniques; after-sales service and user education; willingness to learn from others and adapt; government help; and firm size.

Rothwell (1977) acknowledged that innovation in businesses is a pluralistic and interactive process, and that the order and priority of various success determinants can vary substantially across heterogeneous industrial contexts.

Gibbons and Johnston (1974) examined the role of scientific research in technological innovations and contributions to industrial R&D. They adopted an interview-based approach in order to explore the types of information used to resolve technological challenges in the

development of an innovation, the number of "information units"[5] involved in developing each innovation and the sources of relevant information (cf. Smith, 1997). The authors examined 30 industrial innovations in the UK that involved significant technological change. They found that 34% of the knowledge inputs into an industrial innovation stemmed from outside the private sector. The major sources of scientific information that influenced the development of industrial innovations came from scientific journals and scientists at universities. Of the information that contributed to an innovation, 20% was attributed to basic scientific research. The major sources of technological information inputs came from trade literature and company employees. Gibbons and Johnston also found that scientists from universities or other public research organisations were central in the transfer of information to innovating firms, and that they translated information from scientific journals into a language that was understandable and relevant to developing an innovation in the private sector. Furthermore, public sector scientists provided direct contributions to industrial R&D, such as suggesting alternative designs or providing specialised facilities and services. The authors concluded that the contributions of basic research at universities and public research institutes to commercial innovation extend well beyond simply providing a firm with an exploitable scientific discovery (Gibbons and Johnston, 1974). *De facto*, Gibbons and Johnston's study rejected the linear model of innovation.

Finally, some early studies attempted to explore how to measure the socio-economic benefits and returns from investment into research. Griliches (1958) estimated the social rate of return to investment in hybrid corn research. The costs incurred by society included both research costs and the cost of additional resources devoted to the production of the hybrid seed. The benefits were the higher yields from hybrid corn. On the basis of various assumptions, the internal rate of return to the research expenditures was calculated to be between 35 and 40%. In the UK, the Council for Scientific Policy sponsored a study (Byatt and Cohen, 1969) to assess the feasibility of an approach to quantifying the benefits of basic research in terms of its impact on changes in technology and the development of new industries. Byatt and Cohen used discounted cash flow techniques to see what effect a notional delay in a scientific discovery would have on the net benefits from the discovery. The study found that although in some cases it was possible to trace the economic benefits of a scientific discovery, in general the method was not feasible. This conclusion was confirmed by Norris and Vaizey (1973). Many early studies of the rate of return from research, including those using case study and econometric approaches, were brought together later in a paper by Griliches (1992) and in reports for official bodies in the USA (e.g. Office of Technology Assessment, 1986) and UK (Office of Science and Technology and PREST, 1993).

Specifically in the health field, there have been several strands of analysis attempting to value research. One strand of research incorporates the human capital approach, and attempts to value health gains in terms of the value of production gained as a result of reductions in mortality and morbidity. In 1971, Weisbrod used this approach and an assessment of cost savings in order to calculate the rate of return on the research that led to the vaccine for poliomyelitis (Weisbrod, 1971). Taking into account the cost of the research and the resources devoted to applying the

[5] "A unit was defined in terms of the coherence of its content and source and involved for example information about the properties of materials or relevant general laws or appropriate test procedures" (Edquist and Johnson, 1997: 102).

vaccine, he calculated a multitude of rates of return based on various assumptions and time horizons. Weisbrod suggested that the 'most likely' rate of return seemed to be 11–12%. Mushkin (1979) used the human capital approach to estimate the value of the total reduction in morbidity and mortality in the USA between 1930 and 1975. Mushkin then considered the proportion of the reduction that could be accredited to biomedical research and produced a rate of return of 47%.

In the 1980s and throughout the 1990s, the body of studies into research evaluation grew. Studies were generally of a smaller scale to those of the 1960s and 1970s, but evolved to consider a broader range of factors influencing innovation processes and a broader range of outputs and possible impacts from research. We overview some key studies that are particularly relevant in the context of biomedical and health research evaluation (and to Project Retrosight) below.

1.2 The evolution of studies concerned with innovation processes and outputs: from the 1980s to the present day

As the field of research evaluation evolved, two major streams of work became prevalent in the literature: one concerned primarily with innovation process features, and the other with their outputs.

1.2.1 Process-oriented evaluation studies

In the process stream of studies, attention was devoted to identifying and exploring a diverse range of factors which influence innovation. The following examples illustrate this.

Evered et al. (1987) adopted a survey technique and bibliometric analyses to show that the early undergraduate education and training of medically qualified professionals in the UK influenced their careers and success as researchers.

Using the historiographic methods of the TRACES study, Narin (1989) investigated the impacts of alternative funding mechanisms and research settings on scientific advancements. He examined the evolution of 13 key advances in cancer research, whose development was supported by the US National Cancer Institute. Narin found that the type of research setting (e.g. large versus small, universities versus medical schools) did not have a major impact on the occurrence of advances. In terms of funding support, Narin's study suggested that support for research project grants and intramural programmes were most effective.

In 1992, the Australian National Board of Employment, Education and Training commissioned a study that looked at the effects of resource concentration on research performance. Through a combination of literature reviews and exploratory analysis of evidence on research performance from Australian universities, the study examined the influence of factors such as critical mass (e.g. size of teams), duration of funding and researcher time commitments (see Grant, 1997 for a review of this study). The authors found that groups of between three and five researchers (including postgraduates and technical staff) had a positive effect on research outputs, but that no further economies of scale were realised above that critical mass. However, somewhat in contradiction to this finding, the study also suggested that large, well-funded research groups with committed and involved leadership produced more publications of higher impact and international recognition than smaller groups.

Spangenberg et al. (1990) examined the links between organisational factors and the research performance of clinical units in university hospitals. They used publications, citations and international peer review as performance measures. Their study suggested that "the time spent on research, the volume of human resources, the degrees of craftsmanship of the research manager and the frequency of international communication were positively related to clinical research performance" (Spangenberg et al., 1990: 239).

The PACE Report (Arundel et al., 1995) explored the process of knowledge dissemination as a factor influencing research impact. The study suggested that conferences, recruitment processes, informal contacts and temporary human resource exchanges (among other factors) complement publications as important channels for knowledge exchange and scientific advancement.

Roessner et al. (1998) examined the events, people, interactions and conditions that led to the evolution of 12 significant engineering innovations. A further goal of the study was to document and evaluate the role and significance of National Science Foundation support in the emergence of the innovations. Their in-depth, case study-based research emphasised the importance of nurturing human capital and networks of knowledge exchange for innovation success:

> Indeed, if there is a single, consistent pattern that stands out … it is the critical role played by human capital in the form of individual inventors (e.g. MRI [Magnetic Resonance Imaging], Internet, optical fiber), technical entrepreneurs (e.g. CAD/EC, cellular phone) and students trained in the state-of-the-art who could continue to push technical advance in all three sectors of the economy (government, industry and university). The cases reveal clearly the importance of invisible colleges: scientists and engineers who share results and know-how via networks that span both cooperating and competing institutions. Isolation appears clearly as the enemy of innovation. (Roessner et al., 1998: 17–18)

Roessner et al. also provided a useful framework for considering how institutions support research through involvement in activities such as education, direct research funding, contributions to the knowledge base, research infrastructure, technology organisational leadership and the facilitation of interaction and communication. Similarly, Bozeman and Rogers (1999) emphasised that social networks of communication and collaboration need to be integral in evaluations of knowledge and innovation flows.

1.2.2 Output-oriented evaluation studies

The output-oriented stream of research evaluation studies was concerned with how research impact can be measured, considering various indicators such as publications, citations and monetary assessment. The following gives some examples.

Hertzfeld (1992) provided a review of economic methods used to examine the returns to space research and development. He classified economic approaches into three categories: (1) macroeconomic production function models; (2) microeconomic models of evaluating returns to the economy, based on estimates of consumer and producer surpluses that new technologies create; and (3) direct measures of outputs and their value based on survey data on patents, licenses, value royalties and sales. Hertzfeld suggested that:

> [The] most promising type of economic study for measuring returns to space R&D is the documentation of actual cases ... Case-studies provide relatively clear examples of returns … A well-structured series of case-studies coupled with good theoretical modelling aimed at integrating

> the data and incorporating those data with more economic measures of benefits may well be the way to establish reliable and 'believable' economic measures of returns to space R&D. (Hertzfeld, 1992: 168)

Mansfield (1991) looked at seven sectors of the US economy and attempted to quantify the social return on basic academic R&D by using survey data and econometric analysis. The study estimated the average rate of return to be 28%, but to vary significantly by sector. Efforts to explore how the socio-economic impacts of research funding can be quantified in monetary terms have proven to be notoriously difficult (this is discussed, for example, in Cozzens et al., 1994 and Martin and Salter, 1996). They call for a range of assumptions to be made (such as about the attribution of research impacts), many of which cannot be much more than reasonable hypotheses. In addition, there are difficulties with regard to integrating the complexity of innovation processes and their long timeframes into 'simple calculations'.

Another strand of analysis on the field of evaluation research is the 'Exceptional Returns' approach, which attempts to put a monetary value on life: examples of studies adopting such analyses include Funding First (2000) and Murphy and Topel (2003). The Funding First team developed a framework for assigning economic value to medical research advancements, through valuing the gains in life-expectancy that result from the research. Their study calculated that improvements in life-expectancy added approximately USD 2.6 trillion to the US economy per year in the period 1970–98. Reductions in mortality due to heart disease alone had increased the value of life by USD 1.5 trillion per year during the same period. In 2003, the Australian Society for Health Research commissioned Access Economics to undertake a replication of the US study (Access Economics, 2003). It attempted to calculate the economic returns from research by estimating the value of life and extensions of life. The authors then made assumptions about the fractions of wellness benefits (improvements in health) and longevity benefits (reductions in mortality) derived from research, and assumed that Australia's national contribution is equal to a fraction of the global research outputs (i.e. 2.5%). The report indicated that investments in R&D had surpassed all other sources of raising living standards and that cardiovascular research had highest rates of return on investment. However, the authors admitted that that there are no solid estimates but only intelligent guesses of the proportion of wellness and longevity gains that could be linked to medical research (they used 50% as a best guess). An updated report was published in 2008, in which somewhat different methods were used to calculate the rate of return. This was an attempt to take into account the lags between research being undertaken and the benefits accruing (Access Economics, 2008).

Dawson et al. (1998) attempted to compare UK biomedical research outputs to other international contexts, based on large-scale bibliometric analyses of publications and their citations. They also attempted to identify how research trends (e.g. priority areas of biomedical funding and publication) compared between public and commercial R&D contexts. The researchers found that the UK share of biomedical publications remained relatively static in the period 1988–95, but that shares of publications in the cardiology, genetics, nursing and ophthalmology subfields had increased. National and international collaboration also increased. The government remained the largest funder of biomedical research at the time (34%), but the investment of private foundations and industry grew over the years.

The long history of studies looking at evaluating research with measures of publication outputs, research training and research income levels has been considered in various reviews (e.g. a good

example is Cave et al., 1988). During the 1990s, a number of evaluation studies continued to explore how bibliometric analyses of impact could be refined. Rather than simply looking at publication and citation quantities, studies began to include measures such as 'citations per publication' as a measure of quality and 'research costs per citation' as a measure of efficiency (e.g. Herbertz and Muller-Hill, 1995; May, 1997). Moreover, there has been a long history of trying to categorise and analyse citations for various purposes. This literature was reviewed in a number of more recent studies which have tried to incorporate more sophisticated bibliometric analysis into a wider assessment of the socio-economic impact of research (e.g. Hanney et al., 2005, 2003a; Grant et al., 2003a). These studies have devoted attention to examining ways of categorising citations according to importance (e.g. critical or peripheral to advancement, considering factors such as citation half-life) and function (to review state-of-the-art, develop new knowledge, etc. using more sophisticated citation filters). As emphasised in a 1994 RAND Corporation review of methods available for reviewing fundamental science, "bibliometric indicators are only indirectly and loosely related to those aspects of economic or other goals of public improvement which are of greatest interest" (Cozzens et al., 1994: 44).

Hanney et al. (2003b) explored the possibility of looking at the basic and applied clinical research done in the past and following it forward to the present day. They combined qualitative (survey and interview) and bibliometric (publication and citation analysis) approaches in order to examine the impacts of diabetes research conducted by a team at Newcastle University in 1981 on clinical practice. The study demonstrated that a combination of bibliometric and qualitative approaches could be used successfully to gather data on the evolution and range of impacts from research conducted in the past. The authors also contributed to refining analyses of impact by categorising citations in various ways (across generations of papers, across scientific areas, according to cited half-life) and interpreting bibliometric data in the context of evidence from narratives.

Some output-oriented studies also consider process-related variables. For example, Martin and Salter (1996) examined the links between research funding and economic performance[6] – focusing specifically on the influence of basic research. Their study was rooted in the view that governments and the public expect more direct and tangible benefits from their investments into basic research. Based on estimates of the rates of return from public research investment, they attempted to understand the appropriate level of spending on basic research (as a variable influencing the efficiency of research translation processes) and to quantify it as a percentage of total public funding. Martin and Salter resorted to evidence from previous studies, using econometric assessment, survey and case study methodology. They demonstrated a positive relationship between investment in basic research and economic benefit, but could not quantify the size of the benefit. One of the complications to their efforts was the diversity of definitions used for the term basic research, across various studies. The authors also suggested categories of 'intermediary' benefits that researchers can consider in impact studies, which lead to economic benefits in the long-term. These include: new information; new instrumentation and methodologies; technological skills; access to expert networks; problem-solving skills; spin-off companies; and company attitudes to basic research.

[6] The definition of economic performance that they adopted is rather narrow and does not consider improvements in public health.

Martin and Salter's study also emphasised the importance of considering both success and failure rates (and case studies thereof), in efforts to draw realistic macro-level estimates of social returns on research investment.

Bozeman et al. (1999), through the Research Value Mapping Project, aimed to develop a method for investigating research benefits that captures the richness of data but still allows some quantitative examination of evidence. In Research Value Mapping, case studies were scored by multiple coders, with the scores tested for reliability and then used to test hypotheses about research. The authors also tried to develop various input and output metrices for analysing innovation pathways and to look at how input factors (e.g. size of team, collaboration, basic versus applied focus) influence outputs (e.g. publications, patents). Their study of research funded by the Office for Basic Energy Sciences concluded that basic research produces more patents than applied research, but they did not consider differences in the size of basic versus applied research funding.

Hanney et al. (2007) reviewed 46 papers (five of which were solely conceptual) which assessed the wider impacts from specific programmes of health research. The review included some of the studies described in this report (such as Weisbrod, 1971) that considered economic impacts, and included studies which concentrated on impacts on health policies and practice (e.g. a study of the impact of Health Technology Assessments in Quebec by Jacob and McGregor, 1997; and an analysis of the outcomes from the research funded by the Public Health R&D Committee of the National Health and Medical Research Council in Australia, by Shah and Ward, 2001). Overall, the review identified the methods used in various evaluations (desk research, questionnaires, interviews and case studies), the main models or conceptual frameworks applied, and assessed the strengths and weaknesses of the main approaches. The review confirmed that impact on knowledge generation was quantified more easily than that on policy, behaviour or health gain.

In reflection of the studies described thus far, it becomes apparent that some have been retrospective (tracing back innovations to research that fed into them), and that other more recent studies (including Project Retrosight) have attempted to follow research forward through time, looking at the impacts from bench to bedside (rather than from bedside to bench). Despite demonstrating the feasibility of both research that tracks backwards and forwards, studies have shown that both approaches are presented with the challenge of identifying a coherent starting point for analysis (i.e. boundary setting challenges), as well as challenges in the attribution of impacts to a distinct or piece of research (attribution challenges).

Lastly, in addition to the body of empirical studies we have described, a series of notable reviews on the state of the research evaluation field began to emerge by the mid-1980s. These acknowledged the contributions of previous scholarship and identified various themes in need of further exploration. For example, in a review of the TRACES study, Irvine and Martin (1984) emphasised that innovation is underpinned by the cross-fertilisation of knowledge from diverse disciplines and stages of research (e.g. basic and applied). The authors called for more research on the themes of knowledge exchange, interaction and cross-synthesis. Others highlighted the methodological difficulties in research evaluation studies (e.g. in attempts to calculate an economic valuation of research benefit) and alluded to the need for strengthening the methodological rigour of monetary impact assessments: "In broad terms, the conclusion was that the various possible approaches were all too simplistic and potentially misleading" (Martin and Salter, 1996: 58). In addition, Georghiou and Roessner (2000) have highlighted the inability of

economic returns evaluations to address the externalities that result from R&D activity. They have criticised multi-methodology evaluations for rarely implementing benchmarking and comparison group approaches. Furthermore, reviews of prior work have highlighted the void in integrative conceptual and methodological frameworks to guide empirical investigations of innovation processes, and raid this as a key agenda for future research (e.g. as discussed in Smith, 1987a, 1987b).[7]

It is apparent from the description of the various research evaluation studies described in this report that case study approaches have dominated the research field. In what follows, we briefly reflect on the evolution of case study approaches and overview some key conceptual frameworks which have assisted research evaluation efforts.

[7] In an attempt to contribute to the establishment of conceptual frameworks to assist empirical studies, Smith (then editor of the *British Medical Journal*) describes two models of innovation: a 'linear model' driven by curiosity and the 'market pull model' directed by market need (Smith, 1987b).

CHAPTER 2

Methodological developments: case study-based approaches and conceptual frameworks in evaluation research

Case study approaches in evaluation research have evolved over time. In addition, conceptual frameworks have been developed to assist in organising research enquiries and findings. Bozeman et al. (1999) explained how previous case study approaches have moved through four incremental phases:

1 historical descriptions tracing back innovations to their fundamental supporting inventions (e.g. *The Sources of Invention*);

2 research event-based case studies building a family tree of research events that led to an innovation (e.g. TRACES);

3 matched comparisons – taking 'matched' successful and unsuccessful innovations, tracing and comparing their development (e.g. the SAPPHO project);

4 conventional case studies (using action research, interviews, surveys, narrative descriptions) complemented with economic and bibliometric techniques to increase methodological rigour and objectivity.

Both qualitative and quantitative approaches have been pursued simultaneously. Some scholars have used very loosely structured narratives to try and develop an improved historical understanding of the research process and achieve descriptive depth. An example would be Wellcome Trust witness seminars, which bring together a group of experts involved in a particular advance to recollect and discuss, with the resulting debate being transcribed and published to produce a socially constructed history (e.g. Christie and Tansey, 2001). However, bibliometric and economic analyses also have permeated the literature. Others have developed methods to quantify various aspects of case studies (e.g. by scoring data), to facilitate more systematic analysis and hypothesis testing (e.g. Bozeman et al., 1999; Wooding et al., 2004a).

Evaluators use a variety of methods that share some common features, but each has its own purposes, advantages and disadvantages (see Table 2). Both qualitative and quantitative research methodologies provide their own insights, but have their own trade-offs between depth and breadth. For example, on the one hand, surveys and case study approaches allow for a more detailed tracing of research translation processes than econometric models, but are often difficult to generalise from (as discussed in Eisenhardt, 1989). On the other hand, econometric assessments suffer from difficulties in integrating detail into calculations, of using assumptions of

Table 2. Advantages and limits of different evaluation methods

Method	Brief description	Advantages	Limits
Survey	Asking multiple parties a uniform set of questions about activities, plans, relationships, accomplishments, value or other topics, which can be statistically analysed	▪ Provides an economical way to gather information about a programme and its participants that is not available through other sources ▪ Accommodates the use of control and comparison groups or the collection of counterfactual information ▪ Usually, diverse audiences can understand the approach and results	▪ Phone interviews work best when timeliness is important ▪ Mailed questionnaires often have low response rates ▪ Does not provide the richness of individual project detail that stakeholders tend to find interesting ▪ Responses are often subjective in nature and respondents may not be truthful
Case study – descriptive	Investigating in-depth a programme or project, technology or a facility, describing and explaining how and why developments of interest have occurred	▪ Many decision-makers read and process anecdotal cases more easily than quantitative studies ▪ Provides richness of detail ▪ Can be used to identify best practice experience	▪ The anecdotal evidence provided is generally considered less persuasive than quantitative evidence ▪ The results of one or more individual cases may not apply to other cases ▪ Can be difficult to aggregate findings
Case study – economic estimation	Adding to a descriptive case study quantification of economic effects, such as through cost–benefit analysis	▪ Focuses on ultimate outcomes and impacts rather than on outputs ▪ Provides quantitative estimates of results ▪ Uses financial methods	▪ The value of important benefits may not be estimated in monetary terms ▪ Needs to be carried out a long time after the project has finished
Econometric and statistical analysis	Using tools of statistics, mathematical economics and econometrics to analyse functional relationships between economic and social phenomena and to forecast economic effects	▪ Produces quantitative results with detailed parameters ▪ Demonstrates cause-and-effect relationships	▪ Difficult for non-specialists to understand, replicate and communicate ▪ Not all the effects can be captured in these highly quantitative methods
Sociometric and social network analysis	Identifying and studying the structure of relationships by direct observation, survey and statistical analysis of secondary databases to increase understanding of social or organisational behaviour and related economic outcomes	▪ Focuses on the understanding of the process of innovation ▪ Requires relatively modest inputs which can be obtained through survey, interview or existing databases	▪ Remains largely unfamiliar to most economists and programme stakeholders ▪ Results may not be very informative on a programme's performance
Bibliometrics	Bibliometric analysis use data on numbers and authors of scientific publications and on articles and the citations therein (and in patents) to measure the output of individuals or research teams, institutions and countries, to identify national and international networks and to map the development of new	▪ Widely applicable to evaluation of programmes with an emphasis on publishing or patenting ▪ Can address a variety of evaluation topics, including research output, collaborative relationships and patterns and intensity of knowledge dissemination ▪ Diverse audiences can understand the results ▪ Can be applied to a programme with a relatively short time-lag	▪ Treats only publications and patents as programme outputs and ignores other outputs and long-term outcomes ▪ Time must pass before extensive publication and/or patent citations can be observed ▪ Counts indicate quantity and impact of output, not quality ▪ Not all publications and patents are of equal importance

	(multidisciplinary) fields of science and technology	▪ High degree of credibility	▪ The propensities to publish and patent differ among scientific and technological fields
Historical tracing	Tracing forward from research to a future outcome or backward from an outcome to precursor contributing developments	▪ Produces interesting and credible studies documenting a chain of interrelated developments ▪ Sheds light on process dynamics	▪ Chains of events tend to be highly complex with many organisations and researchers involved
Expert judgement	Using informed judgements to make assessments	▪ Provides a relatively quick, straightforward, feasible and widely-accepted approach to assessment ▪ Offers the opportunity for an exchange of ideas which can lead to new perspectives	▪ Not much is known to the quality or accuracy of expert judgement as applied to R&D programme impact assessment

Source: Adapted and modified from Ruegg and Feller (2003)

uncertain reliability and validity and oversimplifying the dynamics of research translation processes.

There is widespread agreement that criteria for evaluating research need to integrate both qualitative and quantitative information bases, and that single indicators can be misleading (as discussed in Cronin and Normand, 1998; Bozeman et al., 1999; Georghiou and Roessner, 2000). Flawed methodology can lead to flawed and ambiguous findings and implications. The most powerful and rigorous methodologies would be those that combine a number of techniques and make use of triangulation: the idea that integrating a number of sources of evidence, a number of methods of analysis and even a number of investigators, increases the validity and reliability of conclusions (as discussed in Cronin and Normand, 1998; Bozeman et al., 1999; Buxton et al., 1999). Triangulation needs to be based on compatible assumptions about the nature of innovation processes and integrative conceptual and methodological frameworks:

> In the absence of such a framework, the result is a disparate array of indicators and measures. Each set reflects the influence of the underlying discipline, so preference for indicators and measures is guided by the variables of concern in the discipline. (Geisler, 2005: 270)[8]

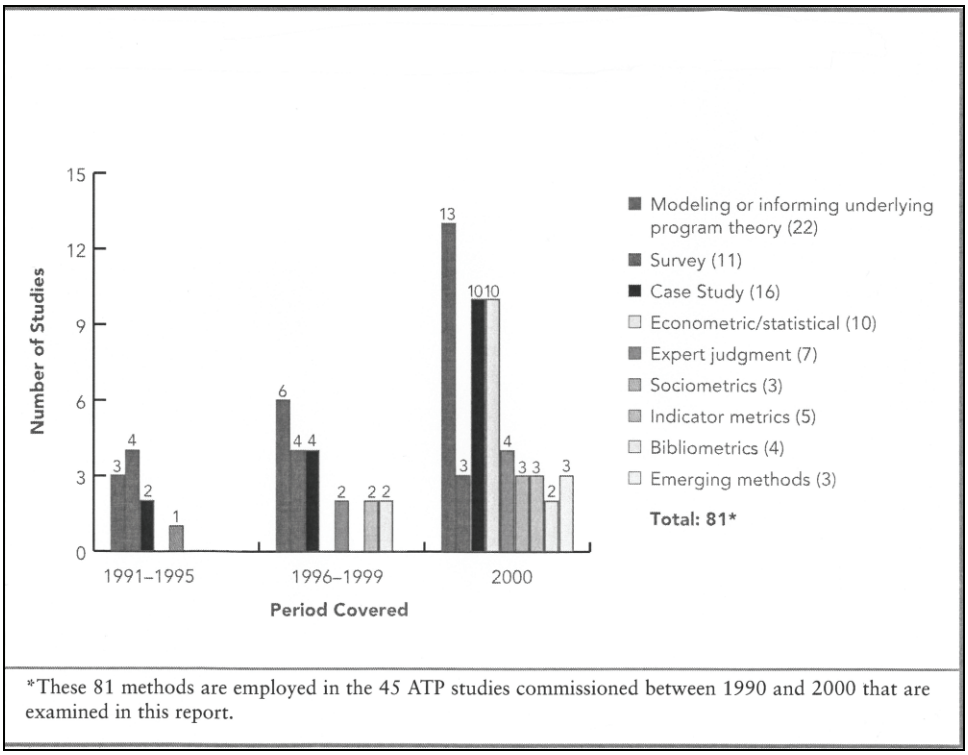
Towards this end, various research evaluation frameworks have been developed. These attempt to provide a more unifying template for conducting research evaluation studies and seek to facilitate the use of multiple methodologies and methods of analysis, as well as cross-case and cross-study comparisons (for discussion of the use of various research evaluation frameworks, see Brutscher et al., 2008).

The Advanced Technology Program's *Toolkit for Evaluating Public R&D Investment* (Ruegg and Feller, 2003) has emerged as one of the most influential reference works, practical aids and planning guides for practitioners of research evaluation across sectors and levels of aggregation. The toolkit was developed based on the analysis of findings from a body of 45 studies commissioned by the Program between 1990 and 2000. The associated report analysed the methods and techniques used and examines the outputs and impacts of those studies. These include: firm or industry effects; collaboration effects; spillover effects; interfaces and comparisons with other programmes; and measures of overall programme performance.

[8] For a more detailed discussion of theoretical limitations, see Geisler (2005).

The extensive toolkit of evaluation methods presented in the report (survey, descriptive and economic case study, bibliometrics, historical tracing, econometrics, expert judgement, social network analysis, cost index and a composite performance rating system constructed from indicator metrics) illustrates how the methods can be used to answer a variety of stakeholder questions (Figure 2). In addition, the use of analytical and conceptual modelling to explore a programme's underlying relationships and process dynamics is considered. Moreover, the report provides an evaluation framework (logic model) for programme administrators, public policymakers and evaluators (Figure 3).

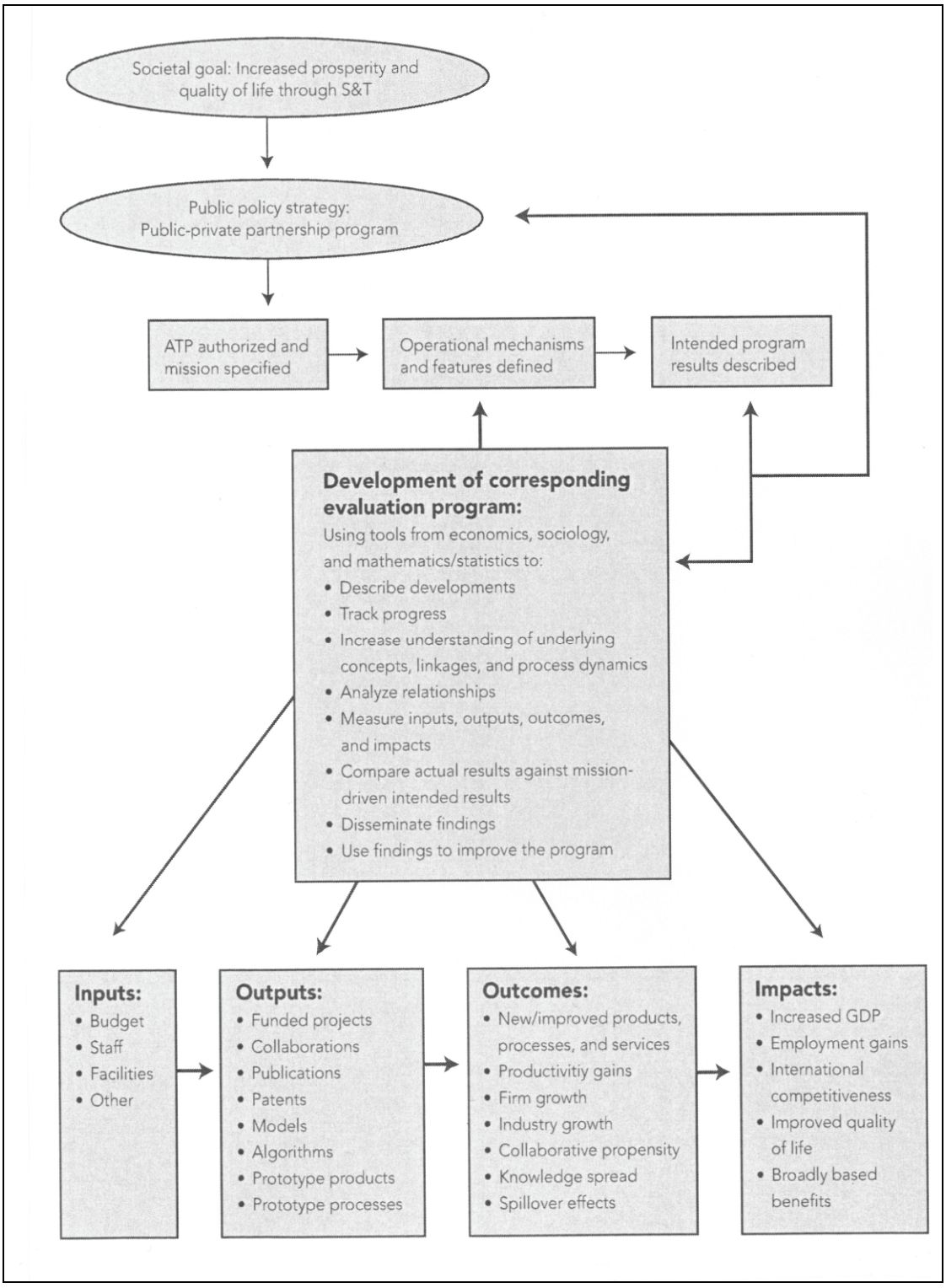
Figure 2. Advanced Technology Program's evolving use of methods over its first decade



*These 81 methods are employed in the 45 ATP studies commissioned between 1990 and 2000 that are examined in this report.

Source: Ruegg and Feller (2003)

Figure 3. Advanced Technology Program Evaluation Logic Model



Source: Ruegg and Feller (2003)

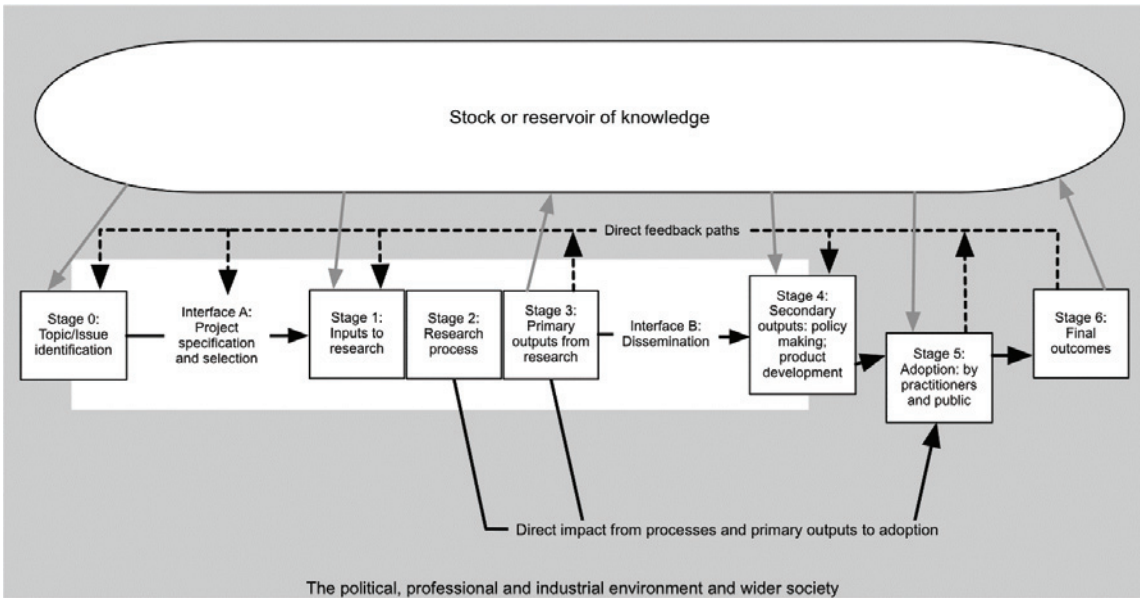
Below we briefly overview some of the key evaluation frameworks which have been used in biomedical research contexts.

2.1 The Payback framework

The Payback framework is the research evaluation framework applied in Project Retrosight. It was developed by the Health Economics Research Group (HERG) at Brunel University (e.g. Buxton and Hanney, 1996)[9] and subsequently refined in collaboration with RAND Europe (e.g. Hanney et al., 2004). It represents a tool for evaluating a comprehensive range of potential *outputs* from research and (unlike most other frameworks) provides a way of conceptualising the *process* through which outputs are created (i.e. a logic framework). The Payback framework considers five categories of payback: (1) knowledge production; (2) benefits for future research and research use; (3) informing policy and product development; (4) health sector benefits; and (5) broader economic benefits.

A range of methods can be used to assess individual impact categories. This includes expert review, bibliometrics, economic analysis, document review and interviews (as discussed in Buxton et al., 1999). Figure 4 summarises the stages of the Payback framework's logic model element.

Figure 4. The research process in the Payback framework



Source: Hanney et al. (2004)

[9] The model is an adaptation of an earlier version developed by HERG to assess the 'payback' of health services and biomedical research.

The Payback framework has been applied to a number of different contexts. As an example, Buxton and Schneider (1999) explored its possible application to a Canadian research organisation that funded basic biomedical and early clinical studies alongside health services research. Prior to this, the framework was applied generally to health service research in the UK context only. In addition, the model has informed analysis of health research systems on behalf of the World Health Organization (WHO; e.g. Hanney et al., 2003a; Pang et al., 2003). It has been used most recently in assessments of the payback from Health Technology Assessment programmes in the UK and the Netherlands (Oortwijn et al., 2008), and to explore the payback on arthritis research funded by the Arthritis Research Council (e.g. Wooding et al., 2004a, 2004b).

2.2 Economic and Social Research Council framework

This framework considers the types of outputs which can be expected from specific research – for example, codified or tacit knowledge, skills; the channels for diffusion and dissemination of outputs to relevant audiences – for example, social networks, publication, the mobility of researchers; the forms of impact – for example, the use of knowledge and skills to solve specific problems, influence policy or technical decisions; direct or indirect impacts; and confirmatory or justification roles (Molas-Gallart et al., 2000).

> Various permutations between outputs, forms of impact and diffusion channels result in many different ways – impact mechanisms – in which economic and social research can have an impact beyond the academic world. (cf. Hanney et al., 2007: 30)

2.3 Royal Netherlands Academy of Arts and Sciences framework

This framework (Royal Netherlands Academy of Arts and Sciences, 2002) focuses on output metrics for healthcare research, and emphasises two key evaluation criteria: (1) research relevance for healthcare providers and healthcare delivery processes; and (2) relevance for policymakers and policymaking. The developers suggest various categories of outputs for evaluation, including: content analysis (e.g. publications, guidelines, protocols); citation analysis; authorship analyses; product development analyses; funding benefits; publicity analyses; memberships analysis; teaching; analyses of output implementation strategies – for example, advisory roles, interactions with external stakeholders, feedback from target groups; and 'independence' – explained as the operationalisation of research questions, methodology, analyses and results in various contexts.

2.4 Framework for impact segmentation by audience

This framework suggests that different target audiences merit different measures of impact, and that impact assessment processes should include steps to identify the target group, select tailored measures depending on who is promoting research use among the target audiences, select measures according to resource availability, identify data sources and collect assessment data (Lavis et al., 2003).

2.5 US Agency for Healthcare and Research Quality impact model

This framework suggests the need to consider multiple levels (stages, types) of impact (Stryer et al., 2000):

- level 1: research findings – findings of studies which identify problems, generate and test hypotheses, establish new research methodologies, establish the effectiveness of various interventions;
- level 2: impact on policy – when a policy or programme is developed as a result of the research;
- level 3: impact on practice – when there is a documented change in what patients or practitioners do;
- level 4: impact on healthcare outcomes – changes in clinical outcomes, economic benefits, quality of life and satisfaction.

The Appendix to this report summarises the major studies reviewed in this report, together with the key insights that they provided.

CHAPTER 3 Consistent themes in research evaluation studies

We have looked across more than half a century of research through a combination of approaches, including content analysis of empirical studies and evidence from systematic reviews. A number of consistent ideas and themes emerge, which include:

- scientific accountability;
- the lack of convincing evidence on innovation 'success determinants';
- the nature of innovation processes; and
- factors influencing innovation processes and the diversity and range of benefits from research.

We expand on each below.

3.1 Scientific accountability and justification of investments into research

Scientific accountability is often thought of as a recent theme in innovation research. However, in fact the overarching purpose of many early studies (e.g. Project Hindsight, TRACES, the Battelle study) was to justify research investments by various public funders. Hence, the idea of scientific accountability is far from a novel theme. For example, it is articulated clearly in the opening paragraphs of Project Hindsight, 40 years ago:

> No matter how much science and technology may add to the quality of life, no matter how brilliant and meritorious are its practitioners and no matter how many individual results that have been of social and economic significance are pointed to with pride, the fact remains that public support of the overall enterprise on the present scale eventually demands satisfactory economic measures of benefit. The question is not whether such measures should be made, it is only how to make them. (Sherwin and Isenson, 1967: 1571)

The concept of scientific accountability has been refined subsequently, but the underlying ethos has remained much the same over time. Later studies often refer to the concept through the notion of a 'social contract', within which government and society expect evidence of direct and tangible benefits from their investments in research. Many argue that science needs to demonstrate not only the benefits that it enables, but also that it is producing those benefits efficiently and effectively (as discussed in Martin and Salter, 1996).

As Trochim points out:

> There is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback. (2006: 1)

However, this is a challenging process:

> Studies that seem critical sometimes fail to influence short-term decisions and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. (Trochim, 2006: 1)

3.2 Lack of convincing evidence on success factors

A complementary theme to that of accountability is one bemoaning the dearth of insights into innovation success factors (e.g. as discussed in Grant, 1997; Blackman, 2004). This is not only a theme but also presents an enduring challenge in research evaluation studies.

While individual studies have provided piecemeal understandings of the various factors influencing research processes and contributed towards insights relevant for science policy, as yet there is no comprehensive analysis of which research models work and which do not. However, perhaps such an analysis is impossible, since science translation and innovation processes are plausibly too complex, multifaceted, pluralistic and dependent on idiosyncratic external contexts (e.g. socio-economic, political, cultural, etc.) to be amenable to prescriptions for success and to broad generalisations. As the former Medical Research Council task force secretary David Smith commented: "We've been looking … for any hard data that helps us look at the relative merits of different models … and we're not finding it" (quoted in Blackman, 2004: 17). To date, the feasibility of identifying success factors remains a contentious topic.

3.3 Nature of the innovation process

Often, studies interested in understanding the contributions of research to society have devoted attention to understanding how innovation pathways unfold. In evaluations, the choice of what to measure, how and when to measure it and how to interpret the results of an evaluation is dependent on the underlying model of innovation evaluators are using "implicitly or explicitly" (Georghiou and Roessner, 2000: 658).

It is widely agreed that the so-called linear model of research – where basic, curiosity-driven research feeds into applied research, which in turn feeds into development and an innovation product – is neither the only nor the pre-eminent form of innovation. One alternative pipeline model is the 'market pull model', in which a recognised market need drives the innovation process (as discussed in Smith, 1987b). In fact, both the linear and market pull models are in essence linear in nature. According to these models, innovation occurs in a 'programmed sequence' of events. However, innovation scholars tend to agree that the evolution of many innovations falls in the grey area between these two, extreme pipeline models. In addition, both models can seep into certain innovation pathways. For example, an academic laboratory doing curiosity-driven research (for example, in pathogenesis) might collaborate with an industry group seeking research to meet a specific market need (a new treatment for a particular disease). Their joint research can lead to an innovation (such as a new pharmaceutical drug).

It is also likely that innovation processes follow more complex positive and negative feedback paths than either the linear or market pull models would suggest. This issue of non-linearity is

addressed in much innovation process literature, including in conceptual frameworks such as the Payback framework (see Figure 2). In addition, sometimes innovations are underpinned sometimes by pure intellectual curiosity-driven research, and sometimes by research initiated with the explicit goal of addressing an unmet need. More recently, innovation scholars have addressed the serendipity of scientific discovery and innovation, where both chance and choice interplay in influencing innovation processes (for example, as discussed by de Rond and Thietart, 2007).[10]

Unexpected discoveries can also play an important role in innovation. As the famous US novelist and scholar, Isaac Asimov, stated: "The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny'."

3.4 Factors influencing innovation processes

Below we summarise some of the commonly examined variables in innovation processes which are thought to affect the nature, type and timing or research outcomes. These variables include: funding mechanisms and research settings; types of research; and knowledge-exchange environments.

3.4.1 Funding mechanisms and research settings

Research evaluation studies have repeatedly examined a number of issues revolving around the impact of funding mechanisms and research settings on innovation. Frequently considered issues include whether:

- funding geared at individuals (e.g. established researchers) or at specific ideas is more conducive to innovation;
- project, programme or institution funding is more efficient;
- funding allocated at the level of research groups (teams) organisations or collaborations is more likely to lead to innovation;
- specialised interdisciplinary institutes or generalised universities are more effective in encouraging innovation and supporting knowledge exchange.

As implied in Section 3.2, there has been much debate on these issues. Blackman (2004) suggested that the current trend in biomedical research organisation leans towards building institutes for integrative biomedical science – housing biologists, physicists, computer scientists and engineers under one roof. This change, he explained, is driven by the changing nature of biological science itself and by an increasing awareness of the cross-fertilisation of disciplines in biological science advancement. By contrast, May (1997) suggested that specialised institutions are less effective in terms of citation production than generalised universities. However, May's

[10] On reflection, even the 1987 example that Smith (1987b) cites in presenting the market pull model of innovation processes alludes (implicitly) to notions of serendipity – albeit not as an explicit source of innovations. Smith discusses the emergence of the Interscan aircraft landing system: in 1968, the International Civil Aviation Organisation requested submissions to design a new aircraft landing system. By coincidence, an Australian radiophysics department with threatened sustainability, saw responding to this tender as a way of remaining financially viable. They then used the technology of radio astronomy to develop the new landing system.

findings were contradicted by a study on Biotechnology and Biological Sciences Research Council institutes, which suggested that institutes outperform universities in a number of scientific fields (as discussed in Miflin, 1997). The importance of cross-fertilisation or interaction between different disciplines and stages of research is rarely disputed, but it is likely that alternative forms of research organisation each have their own merits as well as drawbacks. Blackman (2004) proposes that specialised interdisciplinary institutes are beneficial:

1 for research areas that generally do not fit into the teaching divisions or research agendas of universities;

2 which can be set up quickly in response to specific scientific opportunities and needs; and

3 which they can cluster diverse projects cost-effectively against core big bits of kit.

By contrast, the Howard Hughes Medical Institute's former president, Thomas Cech, suggests that universities may have the advantage of the liberty to experiment with "fresh ideas and the wonderfully naive but important questions" (quoted in Blackman, 2004: 19).

3.4.2 The relative contributions of basic and applied science to innovation

Another question of interest to researchers and policymakers alike concerns the respective contributions of basic research and applied research to innovation. In other words, does basic or applied science merit more funding and institutional support?

The relative benefits of basic versus applied research are a contentious issue. Investigating respective contributions is complicated further by the use of a startling variety of different definitions of the concepts of basic versus applied research, and by the challenge of setting boundaries between the terms (we will return to this issue later). However, most scholars agree that it is far more difficult to trace the influence of more basic research (as discussed in Martin and Irvine, 1989). The majority of the benefits from basic research accrue on a timescale greater than 20 years (as shown in the studies of Sherwin and Isenson, 1967; Illinois Institute of Technology, 1968; Comroe and Dripps, 1977; also discussed in Grant et al., 2003a).

A number of studies have examined the influences of 'technical entrepreneurs' or 'innovation champions' on innovation processes and outputs (e.g. Hanney et al., 2003b). These terms refer to people who act as a bridge, connecting curiosity-driven (basic) and directed (applied) research inputs into innovations. The concepts generally refer to people whose interests span from applied research or practice through to interests in basic research, allowing them both to identify needs from practice, and shape the fundamental research needed to underpin advances. These individuals are identified first in the early studies (e.g. Battelle Laboratories, 1973; Rothwell, 1977) in both public and private sector research, and continue to appear in slightly different forms in more recent work (e.g. Roessner et al., 1998; Buxton and Schneider, 1999).

Most studies on innovation processes agree that crediting breakthroughs to an isolated basic versus applied research event (or to one individual or organisation, for that matter) is unrepresentative of the vast majority of innovations – most of which are built from the combinations of many incremental developments in understanding – and hence entail the exchange and dissemination of knowledge from a variety of R&D sources (projects and researchers).

3.4.3 Knowledge-exchange environments

With the above in view, the theme of knowledge-exchange environments is another recurring topic of interest in studies of innovation processes, with the key question of interest being how effective knowledge exchange can be encouraged and facilitated.

This theme is implicit in studies examining the effectiveness of alternative models of research funding and organisation. For example, facilitating fruitful communications is central to the justifications for the support of interdisciplinary institutes (as discussed in Blackman, 2004), as well as research collaborations. Roessner et al., (1998) emphasise the importance of what they call "invisible colleges" for innovation: "scientists and engineers who share results and know-how via networks that span both cooperating and competing institutions. Isolation appears clearly as the enemy of innovation" (Roessner et al., 1998: 6). Georghiou and Roessner argue that

> innovation gives rise to diverse scientific and technological outcomes including, among others, scientific knowledge, new technological devices, craft and technique, measures and instruments and commercial products. But it is the collective arrangements and their attendant dynamics, not the accumulated innovation products, that should be more properly considered the main asset when assessing the value of research. (2000: 662)

They suggest that

> rather than using the project or a program as a unit of analysis in evaluation, the most appropriate evaluation unit would be a set of individuals connected by their uses of a particular body of information for a particular type of application – the creation of scientific or technical knowledge.

They term this set of individuals "the knowledge value collective" (2000: 662).

Intensive communications are key in the acquisition, assimilation, transformation and utilisation of information and new knowledge. Hermens (2001) argues that frequent and close interactions are particularly important in the transfer of tacit (or implicit) knowledge. The establishment of enduring learning networks often requires experimentation and repeated interactions between individual members of a network (e.g. individuals, research groups organisations). Beyond cumulative learning, repeated interactions also facilitate the establishment of relationships of trust and legitimacy through time. Social relationships can help in efforts to access, recombine and deploy information on advancements that feed into innovation processes (as discussed in Nonaka, 1994; Van den Bosch et al., 1999). In this context, organisational learning and knowledge networks also represent social networks, which take time to establish and need to be nurtured (as discussed in Powell et al., 1996).

3.5 Diversity and range of benefits from research

Perhaps the most common theme in research evaluation studies is that of benefit assessment. Stakeholders in innovation processes are diverse (e.g. funders, academia, industry, the public, etc.) and are likely to bring a mix of common and unique performance goals (disciplinary interests, personal, professional and institutional objectives) into the innovation equation. 'Success' may mean something different to distinct stakeholders (as discussed in Marjanovic, forthcoming 2009). Furthermore, what counts as evidence (of benefits) can be different for different stakeholders and individuals.

The field of research evaluation has been characterised by a growing realisation that the benefits of research are wider than the knowledge produced and published in academic papers. How far evaluation should focus on these impacts has been long debated (as discussed, for example, in Kogan and Henkel, 1983; Kogan et al., 2006). However, support for adopting the 'wider view' is backed by a number of empirical studies: for example, in studies looking at a variety of knowledge-dissemination routes (e.g. Arundel et al., 1995). In addition, the diversity of output and impact categories considered in evaluation research has increased. In the context of healthcare research, impact assessments are trying increasingly to consider a range of outputs, from knowledge-production indicators through to research capacity-building indicators, through to economic indicators and health benefit measures (these developments are illustrated in Buxton and Hanney, 1996; Martin and Salter, 1996; Bozeman et al., 1999).

However, there are other impacts from science that studies have rarely considered and which are extremely difficult to assess (for example, impacts on social relationships, values and beliefs). A typical example would be the impact of knowledge advances into HIV transmission and prevention of the disease on combating the stigmatisation of affected individuals.

The unintended consequences of scientific research are rarely considered in impact evaluation studies, but can be important for policymakers. The use of infectious agents as biological weapons, the ethical debate surrounding genetically modified foods, the regulation of stem cell research applications and the inequity gap in access to new healthcare technologies and products are examples of some negative impacts (as discussed in de Rond and Marjanovic, 2006).

Having reflected on common themes in evaluation research, we now turn to discussing enduring research challenges.

CHAPTER 4 Challenges in research evaluation studies

In reviewing the previous research evaluation studies it has become clear that there are a number of challenges to carrying out research evaluation, many of which have persisted throughout time. These pertain to:

- apparent contradictions between the conclusions of various studies;
- biases in the selection of cases to examine in research;
- a lack of clarity and unity in the definitions of explored concepts (across studies);
- unclear descriptions of study methodology and associated difficulties in the ability to repeat them;
- the challenge of setting boundaries in research;
- challenges in impact attribution; and
- issues of sector idiosyncrasies.

We expand on each challenge below.

4.1 Apparent contradictions between studies as an artefact of research design

Selecting the appropriate type of innovation to study and adequate timeframe to examine in research are important for drawing reliable theoretical and policy-relevant insights from empirical evidence.

In some cases, the apparently contradictory findings between various studies (e.g. Project Hindsight and TRACES) may be no more than an artefact of research design. Two common reasons behind the apparently contradictory insights and key messages between studies may lie in the differences between the types of innovations they examined, and the timeframes they considered. This suggests the need for caution in the interpretation of a study's findings, in criticisms of its headline conclusions and in the use of findings for informing policy. We illustrate these differences below.

4.1.1 Types of innovations examined

Project Hindsight looked at generational changes in military hardware (for example, how a mark 56 mine changed into a mark 57 mine) at the expense of exploring the types of more radical developments that have allowed aircraft to be replaced by unmanned drones. Hindsight has been criticised for underplaying the role of research that causes fundamental changes in approach (this limitation was discussed in Kreilkamp, 1971). As the authors of Project Hindsight themselves acknowledge, the findings on the relative importance of applied versus basic research contributions to innovation may have differed, had the project examined non-engineering innovations and looked back at the origins of innovation over a longer retrospective time period.[11] Also, they may have differed, had the emphasis been on a mix of 'blockbuster' (i.e. more radical) innovations rather than incremental improvements in military hardware.

In contrast, some of the later studies (including Comroe and Dripps, 1977 and TRACES) tried to identify high-profile blockbuster innovations (such as the development of drug treatments for hypertension and oral contraceptives). These studies were criticised as well, this time for underplaying the role of incremental innovation in bringing about social benefit. The shopping trolley, for example, has been cited as an example of a trivial innovation, but one that enabled the building of supermarkets and catalysed the massive changes in society which have resulted (as discussed in Walsh, 1973). Moreover, studies looking at blockbuster innovations have been criticised for overemphasising technological innovations (e.g. magnetic ferrites, oral contraceptives) at the expense of neglecting other types of innovation that lead to social benefit. Some studies have attempted to address this criticism by examining the genesis of more abstract innovations, such as input–output economic analysis (as discussed in Walsh, 1973).

4.1.2 Timeframes

All studies acknowledge that the contributions of research often occur and manifest themselves over long timescales. In fact, the varying timescales examined in evaluation studies help towards explaining the differences between their frequently discrepant headline conclusions and implications (e.g. such as discrepancies between the relative contributions of basic and applied research to innovation outputs). The following illustrate this.

Project Hindsight catalogued 'research events' reaching back 20 years from the contract date for a new generation of military hardware. It reported that 75% of the events were accreditable to 'applied science'. However, the TRACES study looked back 50 years and reported that only 33% of events were mission-oriented (applied) and that non-mission (basic) research peaked 20 to 30 years pre-innovation. The TRACES study results suggest that if the timeframe were reduced to 20 years, then the findings and implications would not be so strikingly dissimilar to Project Hindsight.[12] Comroe and Dripps' (1977) study looked back even further than TRACES, sometimes including research from more than 50 years earlier in their chronology of articles for the development of electrocardiography (for example, it includes a key article by Benjamin Franklin published in 1752). Unfortunately, Comroe and Dripps did not provide a time distribution of the key papers that they identified. Although they identify 111 representative 'lags'

[11] Hindsight looked back 20 years for the innovations it examined, which may have been too short to trace the basic science inputs into innovation.

[12] One might argue that sectoral differences play an influence on the discrepant findings, but there is no evidence to draw from at this stage to reach such a conclusion.

between phases of scientific discovery and application, they do not specify how these lags were selected. However, 57% of these lags they identify pertain to a period of more than 21 years, and 18% to a period of more than 50 years.

The findings of Comroe and Dripps on the contributions of universities and non-mission-oriented research to innovation appear to contradict those of Project Hindsight.[13] However, the differing timescales and magnitude of events described in the two studies compromise the validity of inferences made through comparisons. Comroe and Dripps looked at a longer timescale than Project Hindsight (50 versus 25 years), and the events they considered are much larger in magnitude than many of those looked at in Project Hindsight (many of the Hindsight engineering-related events are not publication-prone). This again suggests that the difference in findings between distinct studies may be more an artefact of timescale than an indication of contradictions and disagreement.

4.2 Biases in the selection of cases

In addition to concerns about the type of innovations (i.e. blockbuster versus incremental) which have been investigated, concerns have been raised about how the particular innovations have been selected. Often, selection processes and criteria are not clear from study reports, and even in studies where processes of selection are explained, biases still exist. At times, selection has been specified by the funders of a study whose internal agendas (e.g. to show impact from the projects that they fund and legitimise their organisation) appear to have influenced selections (and findings) by leaning towards highly successful and high-impact innovations, or to a specific type of research.[14] As discussed by Kreilkamp (1971), concerns about bias have touched on how particular 'research events' or publications are classified (e.g. in deciding what is basic or what applied research). If research is funded by donors who aim to legitimise a strategy of investing in basic research, there can be a tendency to 'overemphasise' the contributions of basic research to innovation, and vice versa.

4.3 Lack of clarity and unity in the definitions of concepts

A lack of clarity in the definitions of concepts explored in research has impeded cross-case comparison efforts and the informative value of studies for policymaking. For example, most early studies do not address the distinctions between an innovation, an invention and a discovery. Different studies also assume disparate meanings of terms such as basic versus applied, undirected versus directed and non-mission versus mission-oriented research. Figure 5 shows how distinct studies adopt different definitions for the same types of research stages.

[13] Project Hindsight suggested that the contributions of universities were relatively low (i.e. 12%), and of non-mission-oriented research, that it was minimal (i.e. 0.3%)

[14] For example, issues of potential bias apply to the evaluation studies by Sherwin and Isenson (1967); Illinois Institute of Technology (1968); Battelle Laboratories (1973); Comroe and Dripps (1977); Roessner et al., (1998).

Figure 5. Comparison of definitions used for types of research

<table>
<tr><th></th><th colspan="3">Basic</th><th colspan="2">Applied</th><th colspan="3">Developmental</th></tr>
<tr><td rowspan="2">Project Hindsight</td><td colspan="5">Science events: theoretical or experimental studies of new or unexplored natural phenomena.</td><td colspan="3" rowspan="2">Technology events: the conception or demonstration of the possibility of performing a specific elementary function with the use of established concepts, principles or materials; the measurement of the behaviour of materials and equipment as required for design; the development of new manufacturing techniques.</td></tr>
<tr><td colspan="3">Undirected: where the object of the work is the advancement of knowledge, without regard to possible application</td><td colspan="2">Applied/directed: where the object of the work is to produce specific knowledge or an understanding of a phenomenon which is needed for some particular use or uses. [Includes findings of research aimed at one problem but used for another]</td></tr>
<tr><td>TRACES</td><td colspan="4" rowspan="2">Non-mission-oriented</td><td colspan="3" rowspan="2">Development and application</td></tr>
<tr><td>Battelle</td><td colspan="2">Mission-oriented
[Does not seem to include research aimed at one problem but used for another]</td></tr>
<tr><td rowspan="2">Comroe and Dripps</td><td colspan="3" rowspan="2">Basic research is where the investigator attempts to determine the mechanisms responsible for the observed effects. It can be clinically-oriented or non-clinically-oriented, whether the research is investigator-initiated or commissioned; undirected or directed…</td><td colspan="2" rowspan="2">Clinically-oriented is research not concerned with the basic biological, chemical or physical mechanism (eg purely observational work or application of a procedure practised in animals, to man)</td><td colspan="2">Developmental work/engineering to create improve or perfect apparatus for…</td><td rowspan="2">Review and synthesis of published work</td></tr>
<tr><td>…unrelated to the solution of a clinical problem</td><td>…related to the solution of a clinical problem</td><td>…research use</td><td>…clinical or other use</td></tr>
</table>

Although many studies in the research evaluation field classify the research giving rise to 'events'[15] (e.g. publications), none of the studies consider the context in which the research took place. For example, Project Hindsight notes one of its events as the formulation of a new version of neoprene with a high tear strength (research event 0448). The study classifies this event as applied research, because it was carried out in response to a specific engineering need. However, the context of this work is not considered. Where did the knowledge to allow the formulation of the new neoprene come from? Was the reformulation possible because the laboratory was also carrying out basic research into the properties of elastomers (i.e. rubber)? These issues are not explained in the Project Hindsight official documents (e.g. such as in Naval Ordnance Laboratory – White Oak, 1966 document). A consideration of the research context could add value to the methodological robustness of future studies.

4.4 Challenges to the ability to replicate methodology: unclear descriptions of techniques used for data collection and analysis

A recurring challenge in innovation process and evaluation research has been that of devising and documenting robust and repeatable methods.

For example, this is evident in the work of Grant et al. (2003b), who failed to repeat successfully the methodology of Comroe and Dripps (1976) in their study of neonatal care research. To elaborate, Comroe and Dripps (1976) neither clearly explained how they identified the key advances in cardiovascular medicine which they examined, nor how they identified the key areas of knowledge that supported these advances. Furthermore, their method to identify key publications was based almost entirely on the opinions of the authors of the publications themselves. The relatively recent work of Bozeman and Rogers (1999) on the Research Value Mapping project also fails to present adequately how this method was applied and the analysis that lies behind the study's conclusions on programmes funded by the Office of Basic Energy Sciences (as discussed also in Bozeman and Klein, 1999).[16]

Clearly, a lack of descriptive strength in methodological explanations of studies poses challenges to the establishment of a robust evidence base which can inform science policy.

4.5 The challenge of setting boundaries in research

All research entails a process of boundary setting, both in data collection and analysis. As Runde and de Rond point out: "In the absence of restriction, we would have extreme difficulty in deciding where to go and where to begin and might become paralysed by the choices" (2007: 16). The challenge to investigators lies in setting the most appropriate boundaries to inform their overarching research agenda.

Most of the studies discussed in this report have been retrospective, working back from the final product or impact to the research that supported it. This approach has the advantage of providing

[15] Events included research papers, but also extended to patents, research board meetings, various recognitions of scientific advancement and other means of know-how transfer between R&D laboratories.

[16] This despite detailing a subtle potentially valuable method of quantifying the inputs and outcomes of case studies.

a clear 'finishing line' for each innovation, although there can be discussion about what this line should be. In the context of drug development for example, should the finishing point be that of first therapeutic use, or the point of widespread drug adoption (and perhaps improvement)?

However, the challenge with retrospective studies lies in deciding on a 'starting line'. The approach makes it difficult to consider the full range of research from which innovation could have arisen (as discussed in Smith, 1987a). For example, it is plausible that basic research provides far more of the findings that feed into new medical treatments than it is accredited? When considered in terms of money spent, basic research appears to be less efficient in providing medically relevant findings and medical innovations. However, in terms of efficiency arguments, it is equally plausible to argue that even if clinical (applied) research contributes the majority of findings to medical advances, it may be much more expensive than basic research and hence less cost-efficient. Both of these speculations are of course extremes and threatened by over-simplification. However, they allude to the importance of considering both basic and applied research as complements – cooperators rather than competitors – when assessing the cost-efficiency of research (as discussed in Mansfield, 1991).

In contrast with retrospective studies, the use of approaches that work prospectively (i.e. that work forward from research rather than back from innovation) is more rare in impact evaluation studies, but is gaining popularity in more recent times (some examples include the studies by Hanney et al., 2003b; Wooding et al., 2004a, 2004b). So far, prospective approaches (which are adopted in the Retrosight project) have always taken a historical view – taking research from the past and following it through to the present day. Prospective research itself is faced with the challenge of setting boundaries, because it is difficult to determine adequate timeframes for an endpoint. To our knowledge, no research funder has put in place a comprehensive system yet to monitor the developments accruing from funded research on a prospective basis and for appropriate timeframes for the translation of research into innovations. As Trochim points out: "Many groups and organizations struggle with how to build a good evaluation capability into their everyday activities and procedures" (2006: 1).

All studies acknowledge that the contributions of research often occur and manifest themselves over long timescales, and that timescales can have an important influence on the findings and messages that evaluation studies provide. In fact (as was illustrated in Section 4.1), the varying timescales examined in evaluation studies help towards explaining the differences between the discrepant headline conclusions and implications of some studies (such as insights on the relative contributions of basic and applied research to innovation outputs).

4.6 Challenges in the attribution of impacts

A recurring challenge in research evaluation studies is the issue of assessing the impact of a particular study, project or type of research (e.g. basic versus applied) on the generation of an innovation. Bibliometric approaches (e.g. citation analysis) have attempted to assist attribution efforts. However, when – as is generally the case – an innovation is generated through contributions from diverse research projects over time, attribution is by no means

straightforward. The matter is complicated further by the frequent difficulties of delineating a research project from other projects and studies that have influenced it or were influenced by it.[17]

Buxton et al. (1999) emphasise attribution challenges in the context of accrediting research outputs to a particular funder, in cases where the research is supported by funding from multiple agencies. Underpinning the attribution challenges are difficulties in gathering the required information. As the authors comment:

> The UK Department of Health may wish to identify the separate and additional impact its funding makes to payback from R&D. In this case it is necessary to find a means of apportioning funding to outputs and outcomes … Information will be needed about the different funding streams contributing to each publication, each grey output, each higher degree and each patent. This information should (or could) be supplemented by an estimate of the proportion of each output able to be attributed to each funding source. This is an ambitious requirement, raising conceptual and practical problems. It is likely to be difficult for researchers to attribute proportions of each output to different sources and it may be resented by researchers … It may, however, be a necessary task. It could be made more palatable if it were required by all funders and could in the long-term perhaps become information that was routinely noted in publications. We propose the feasibility and acceptability needs to be tested. (Buxton et al., 1999: 43)

4.7 Issues of sector idiosyncrasies

It is not clear how similar or different innovation processes are across sectors. Many of the studies which have influenced thinking in biomedical research policy have not been conducted on biomedical research (for example, Project Hindsight is related to the military sector). The broader variables influencing innovation are likely to apply across research disciplines (e.g. influence of funding mechanisms, research organisation, etc.) and there is significant cross-fertilisation of ideas and insights from studies across sectors. However, there is also a growing recognition that innovation processes are context-dependent and pluralist (as discussed in de Rond and Marjanovic, 2006). A study examining the National Science Foundation's contribution to enabling innovation suggested that differences between public and commercial research dynamics, and between innovation processes across sectors (such as telecoms and medical technology), are significant (as discussed in Roessner et al., 1998).

Studies of health sector research contributions to society still suffer from a lack of sufficient empirical evidence on innovation processes within specific subsectors of health sector research. Project Retrosight, for which this report sets the historical context, is concerned with cardiovascular research. The paucity of studies in research evaluation that draw evidence from cardiovascular research implies that, to a large extent, the idiosyncrasies that may apply to innovation in this particular biomedical research field are unknown. We hope to have unravelled some of them through our study.

[17] These issues are discussed, for example, in Irvine and Martin (1984); Smith (1987b); Hanney et al. (2003a, 2003b, 2005).

CHAPTER 5 Conclusion

In reviewing the papers discussed above, it is striking that many of the issues grappled with when evaluating research 40 years ago remain relevant today. We hope to have provided an informative, evidence-based overview of past research contributions and of recurrent themes that are in need of refinement and clarification. We also brought to the surface the diverse challenges facing those who evaluate research and those who use evaluation-based evidence to make strategic decisions (for example, donors and policymakers). By doing so, we hope to have made the case for more methodologically robust research evaluation approaches, and to have provided direction for studies which, like project Retrosight, place research reliability and validity high on their agenda.

There has been continuous and gradual progress and improvements in evaluation techniques and approaches. Nevertheless, as we have shown, many challenges and limitations persist. Evaluation is an imperfect process. However, even with their limitations, research evaluation approaches can still make a valuable contribution to evidence-based decision making – as long as we remain cautious and aware of the limitations when drawing conclusions from the findings.

There are a number of other practical concerns and trade-offs that are associated with conducting comprehensive and rigorous evaluations. Among others, these include constraints in budget, time and data availability. Indeed, the growing sophistication of research evaluation techniques over time, as well as attempts at more comprehensive coverage of the possible impacts from research, themselves can create the need for trade-offs (for example, between the rigour of an evaluation that adopts multiple methodologies and sources of evidence, and the need to minimise evaluation costs). In addition, the significant resources devoted to research evaluation can be underutilised. One of the most effective ways to make sure that evaluation findings are used is to disseminate them in a way that is of direct practical relevance and utility for specific stakeholders. Another important factor is to recognise that evaluation is one of many different sources of information and influence on decision-makers.

At the beginning of this report, we discussed four broad objectives of research evaluation:

1 to increase accountability;

2 to steer the research process towards desired outcomes;

3 to provide a means for advocacy; and

4 as an input into the research management process.

Implicit to achieving these objectives is the need for building stakeholder relationships. This includes open and honest communication about research needs, sensitivities, the evaluation

processes and findings between those funding, and evaluating and conducting research. Usually without such communication, the intended benefits of an evaluation cannot be fully harnessed. Similarly, effective communications are likely to help build a culture of trust and buy-in for evaluations across different interest groups.

Another interesting issue is the relationship between evaluation and the changing role of science in modernity. There are unresolved tensions between an evaluation culture and the serendipitous aspects of scientific discovery, and about how the two reconcile. Research evaluation helps to ensure accountability in the spending of public money. However, it is interesting to consider what evaluation-intensive systems imply for researcher's ability to experiment and take risks, which can lead to unexpected advances and discoveries. Striking an appropriate balance between the need for accountability on one side, and autonomy on another, is by no means straightforward.

Clearly, many unanswered questions remain and will continue to challenge as well as inspire those studying the field of research evaluation and evaluation practitioners alike. Overall, we have witnessed a growing diversity over time in the types of research being conducted, and an increased demand for the evaluation of diverse research. Within that context, perhaps the key consideration is the need to ensure that the evaluation approaches used are appropriate both for the type of research being evaluated, and for the aims of the specific evaluation.

Bibliography

Access Economics, 2003. *Exception Returns: The Value of Investing in Health R&D in Australia.* Canberra: Australian Society for Medical Research.

Access Economics, 2008. *Exception Returns: The Value of Investing in Health R&D in Australia II*. Canberra: Australian Society for Medical Research.

Alberta Heritage Foundation for Medical Research, 2003. *Assessment of Health Research Fund Outputs and Outcomes: 1995–2003.* Edmonton: Alberta Heritage Foundation for Medical Research.

Arundel, A., van de Pall, G. and Soete, L., 1995. *PACE Report: Innovation Strategies of Europe's Largest Industrial Firms: Results of the PACE Survey for Information Sources, Public Research, Protection of Innovations and Government Programmes*. Maastricht: Limburg University Press.

Bamberger, M., Rugh, J. and Mabry, L., 2006. *Real World Evaluation*. Thousand Oaks, CA: Sage Publications.

Battelle Laboratories, 1973. Interactions of Science and Technology in the Innovative Process: Some Case Studies. *National Science Foundation Report NSF C667*. Columbus, OH: Battelle Columbus Laboratories.

Benson, J.K., 1977. Organisations: a Dialectical View. *Administrative Science Quarterly,* 22(1): 1–21.

Berger, P.L. and Luckmann, T., 1966. *The Social Construction of Reality*. Garden City, New York: Anchor Books.

Berra, S., Guillamón, I., Méndez-Vásquez, R., Suñen-Piñol, E., Sánchez, E. and Pons J.M.V., 2006. *Avaluació de l'impacte de La Marató de TV3 en la recerca biomèdica a Catalunya.* Barcelona: Fundació La Marató de TV3. Agència d'Avaluació de Tecnologia i Recerca Mèdiques.

Blackman, S., 2004. The Right Research Mix. *The Scientist*, 18(4): 17–21.

Bozeman, B. and Klein, H.K., 1999. The Case Study as Research Heuristic: Lessons from the R&D Value Mapping Project. *Evaluation and Program Planning*, 22(1): 91–103.

Bozeman, B. and Rogers, J., 1999. Use and Transformation: An Elementary Theory of Knowledge and Innovation (draft paper). Atlanta, GA: School of Public Policy, Georgia Institute of Technology.

Bozeman, B., Rogers, J., Dietz, J.S. and Park, J., 1999. *The Research Value Mapping Project: Qualitative–quantitative Case Studies of Research Projects Funded by the Office of Basic Energy Sciences*. Atlanta, GA: Georgia Institute of Technology.

Brutscher, P., Wooding, S. and Grant, J., 2008. *Health Research Evaluation Frameworks: An International Comparison.* RAND Europe publication TR-629-DH.

Buxton, M. and Hanney, S., 1996. How Can Payback from Health Services Research Be Assessed? *Journal of Health Service Research and Policy,* 1(1): 35–43.

Buxton, M. and Schneider, W.L., 1999. *Assessing the 'Payback' from AHFMR-funded Research*. Edmonton: Alberta Heritage Foundation for Medical Research.

Buxton, M., Croxson, B. and Hanney, S., 1999. Assessing the Payback from Department of Health Research & Development: From *Ad Hoc* Studies to Regular Monitoring, *HERG Research Report 27*. Uxbridge: Health Economics Research Group, Brunel University.

Byatt, I. and Cohen, A., 1969. An Attempt to Quantify the Economic Benefits of Scientific Research. *Science Policy Studies 4*. London: HMSO.

Cave, M., Hanney, S., Kogan, M. and Trevett., G., 1988. *The Use of Performance Indicators in Higher Education.* London: Jessica Kingsley.

Christie, D.A. and Tansey, E. M., 2001. Origins of Neonatal Intensive Care in the UK. *Witness Seminar Transcript*. London: Wellcome Institute for the History of Medicine.

Comroe, J.H., 1977. *Retrospectroscope: Insights into Medical Discovery*. Menlo Park, CA: Von Gehr Press.

Comroe, J.H. and Dripps, R.D., 1976. Scientific Basis for the Support of Biomedical Science. *Science,* 192(4235): 105–111.

Comroe, J.H. and Dripps, R.D., 1977. *The Top Ten Clinical Advances in Cardiovascularpulmonary Medicine and Surgery 1945–1975*, DHEW Publication (NIH) 78-1521. Washington, DC: Public Health Service, National Institutes of Health and US Department of Health, Education and Welfare.

Council of Medical Sciences, 2002. *The Societal Impact of Applied Research: Towards a Quality Assessment System.* Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Cozzens, S., Popper, S., Bonomo, J., Koizumi, K. and Flanagan, A., 1994. *Methods for Evaluating Fundemental Science.* RAND/CTI DRU-875/2-CTI, Washington, DC: RAND.

Cronin, E. and Normand, C., 1998. *Methodologies for Evaluating R&D Programmes: A Literature Review.* London: Department of Public Health and Policy, London School of Hygiene and Tropical Medicine.

Curie, M., 1921. On the Discovery of Radium. Speech given at Vassar College, Poughkeepsie, NY, 14 May.

Dawson, G., Lucocq, B., Cottrell, R. and Lewison, G., 1998. Mapping the Landscape: National Biomedical Research Outputs 1988–95. *PRISM Policy Report 9.* London: Wellcome Trust.

de Rond, M. and Marjanovic, S., 2006. The Legitimacy of Messiness: Interdisciplinary Research, Systems of Innovation and Strategic Alliances. In O. Shenkar and J.J. Reuer (eds) *Handbook of Strategic Alliances*, pp. 413–40. Thousand Oaks, CA: Sage Publications.

de Rond, M. and Thietart, R.A., 2007. Choice, Chance and Inevitability in Strategy. *Strategic Management Journal,* 28(5): 535–51.

Edquist, C. and Johnson, B. (1997). Institutions and Organisations in Systems of Innovation. In C. Edquist (ed.) *Systems of Innovation: Technologies, Institutions and Organizations*, pp 1-519. London and Washington, DC: Pinter/Cassell Academic.

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. *Academy of Management Review*, 14(4): 532–50.

Evered, D.C., Anderson, J., Griggs, P. and Wakeford, R., 1987. The Correlates of Research Success. *British Medical Journal* (clin. res. edn), 295(6592): 241–6.

Fahrenkrog, G., Polt, W., Rojo, J., Tübke, A. and Zinöcker, K., 2002. *RTD Evaluation Tool Box: Assessing the Socio-economic Impact of RTD Policies*. IPTS Technical Report Series. Seville

Freeman, C., 1974. *The Economics of Industrial Innovation*. Harmondsworth: Penguin.

Funding First, 2000. *Exceptional Returns: The Economic Value of America's Investment in Medical Research.* New York: Lasker Foundation.

Geisler, E., 2005. The Measurement of Scientific Activity: Research Directions in Linking Philosophy of Science and Metrics of Science and Technology Outputs. *Scientometrics,* 62(2): 269–84.

Georghiou, L. and Roessner, D., 2000. Evaluating Technology Programs: Tools and Methods. *Research Policy,* 29(4-5): 657–78.

Georghiou, L. and Larédo, P., 2005. *Evaluation of Publicly Funded Research: Report on the Berlin Workshop*. [Accessed 19 August 2009: http://www.internationales-buero.de/_media/Report_on_Evaluation_Workshop.pdf]

Gibbons, M. and Johnston, R., 1974. The Role of Science in Technological Innovation. *Research Policy,* 3(3): 220–42.

Gibbons, M. and Georghiou, L., 1987. *Evaluation of Research: A Selection of Current Practices*. Paris: Organisation for Economic Co-operation and Development.

Grant, J., 1997. *Factors Leading to Success: A Literature Review*. London: Wellcome Trust.

Grant, J., Lewison, G. and May, R.M., 1997. Government Funding of Research and Development. *Science,* 278(5339): 878–80.

Grant, J., Green, L. and Mason, B., 2003a. Basic Research and Health: A Reassessment of the Scientific Basis for the Support of Biomedical Science. *Research Evaluation,* 12(3): 217–24.

Grant, J., Green, L. and Mason, B., 2003b. From Bedside to Bench: Comroe and Dripps Revisited. *HERG Research Report 30*. London: Brunel University.

Griliches, Z., 1958. Research Costs and Social Returns: Hybrid Corn and Related Innovation. *Journal of Political Economy,* 66(5): 419–31.

Griliches, Z., 1992. The Search for R&D Spillovers. *Scandinavian Journal of Economics,* 94(supp.): 29–47.

Hanney, S., Gonzalez-Block, M., Buxton, M. and Kogan, M., 2003a. The Utilisation of Health Research in Policy-making: Concepts, Examples and Methods of Assessment. *Health Research Policy and Systems,* 1(2): 1-28.

Hanney, S., Frame, I., Grant, J., Green, P. and Buxton, M., 2003b. From Bench to Bedside: Tracing the Payback Forwards from Basic or Early Clinical Research – A Preliminary Exercise and Proposals for a Future Study. *HERG Research Report*. Uxbridge: Health Economics Research Group, Brunel University.

Hanney, S., Grant, J., Wooding, S. and Buxton, M., 2004. Proposed Methods for Reviewing the Outcomes of Research: The Impact of Funding by the UK's 'Arthritis Research Campaign'. *Health Research Policy and Systems*, 2(4):1-11

Hanney, S., Frame, I., Grant, J., Buxton, M., Young, T. and Lewison, G., 2005. Using Categorisations of Citations when Assessing the Outcomes from Health Research. *Scientometrics,* 65(3): 357–79.

Hanney, S., Buxton, M., Green, C., Coulson, D. and Raftery, J., 2007. An Assessment of the Impact of the NHS Health Technology Assessment Programme. *Health Technology Assessment,* 11(53)

Herbertz, H. and Müller-Hill, B., 1995. Quality and Efficiency of Basic Research in Molecular Biology: A Bibliometric Analysis of Thirteen Excellent Research Institutes. *Research Policy,* 24(6): 959–79.

Hermens, A., 2001. Knowledge Exchange in Strategic Alliances: Learning in Tension. *Creativity and Innovation Management,* 10(3): 189–91.

Hertzfeld, H., 1992. Measuring Returns to Space Research and Development. in J.S. Greenberg and H. Hertzfeld (eds.), *Space Economics*, Washington, DC: American Institute of Aeronautics and Astronautics

Illinois Insitute of Technology, 1968. *Technology in Retrospect and Critical Events in Science (TRACES)*. Washington, DC: National Science Foundation.

Irvine, J. and Martin, B.R., 1984. *Foresight in Science: Picking the Winners*. London: Frances Pinter.

Jacob, R. and McGregor, M., 1997. Assessing the Impact of Health Technology Assessment. *International Journal of Technology Assessment in Health Care*, 13(1): 68–80.

Jewkes, J., Sawers, D. and Stillerman, R., 1958. *The Sources of Invention*. London: Macmillan.

Jewkes, J., Sawers, D. and Stillerman, R., 1969. *The Sources of Invention* (2nd edn). London: Macmillan.

Kalucy, L., McIntyre, E. and Bowers, E.J., 2007. *Primary Health Care Research Impact Project: Final Report Stage 1*. Adelaide: Primary Health Care Research and Information Service, Flinders University.

Knapp, P., 1999. Evolution, Complex Systems and the Dialectic. *Journal of World Systems Research,* 5: 74–103.

Kogan, M. and Henkel, M., 1983. *Government and Research: The Rothschild Experiment in a Government Department*. London: Heinemann.

Kogan, M., Henkel, M. and Hanney, S., 2006. *Government and Research: Thirty Years of Evolution* (2nd edn). Dortrecht: Springer.

Kreilkamp, K., 1971. Hindsight and the Real World of Science Policy. *Science Studies,* 1: 43–66.

Kwan, P., Johnston, J., Fung, A.Y.K., Chong, D.S.Y., Collins, R.A. and Lo, S.V., 2007. A Systematic Evaluation of Payback of Publicly Funded Health and Health Services Research in Hong Kong. *BMC Health Services Research*, 7(121)

Lavis, J., Robertson, D., Woodside, J., McLeod, C., Abelson, J. and The Knowledge Transfer Study Group, 2003. How Can Research Organizations More Effectively Transfer Research Knowledge to Decision Makers? *Milbank Quarterly*, 81(2): 221–48.

Mansfield, E., 1991. Academic Research and Industrial Innovation. *Research Policy,* 20(1): 1–12.

Marjanovic, S., 2009 (forthcoming). The Performance of Global Health R&D Alliances and the Need for Interdisciplinary Research Approaches Towards its Understanding. In Konrad, M. (Ed) *Global Science, International Health and the Ethics of Collaborative Practice*, Cambridge: Cambridge University Press.

Martin, B.R. and Irvine, J., 1989. *Research Foresight: Priority Setting in Science*. London: Pinter.

Martin, B. and Salter, A., 1996. *The Relationship between Publicly Funded Basic Research and Economic Performance: A SPRU Review*. Brighton: Science Policy Research Unit, University of Sussex.

May, R.M., 1997. Scientific Wealth of Nations. *Science,* 275(5301): 793–6.

Miflin, B., 1997. Institutes Top UK's Scientific League Table. *Nature,* 390(6655):12.

Molas-Gallart, J., Tang, P. and Morrow, S., 2000. Assessing the Non-academic Impact of Grant-funded Socio-economic Research: Results from a Pilot Study. *Research Evaluation,* 9(3): 171–82.

Murphy, K.M. and Topel, R.H., 2003. The Economic Value of Medical Research. In K.M. Murphey and R. H. Topel (eds) *Measuring the Gains from Medical Research: An Economic Approach*. Chicago, IL: University of Chicago Press.

Mushkin, S., 1979. *Biomedical Research: Costs and Benefits.* Cambridge, MA: Ballinger Publishing.

Narin, F., 1989. The Impact of Different Modes of Research Funding. In D. Evered and S. Harnett (eds). *The Evaluation of Scientific Research*. Chichester: John Wiley and Sons.

Naval Ordnance Laboratory (White Oak), 1966. *Research and Exploratory-development Origins of the Naval Mines Mark 56 and 57: Research and Exploratory-development Event Descriptions for Project Hindsight.* Washington, DC: Office of the Director of Defense Research Engineering.

Nonaka, I., 1994. A Dynamic Theory of Organizational Knowledge. *Organization Science* 5(1): 14–37.

Norris, K. and Vaizey, J., 1973. *The Economics of Research and Technology*. London: George Allen & Unwin.

Office of Science and Technology and PREST, 1993. *Returns to Research and Development Spending*. London: HMSO.

Office of Technology Assessment, 1986. *Research Funding as an Investment: Can We Measure the Returns? A Technical Memorandum*. Washington, DC: United States Congress.

Oortwijn, W.J., Hanney, S.R., Ligtvoet, A., Hoorens, S., Wooding, S. Grant, J., Martin, J. Buxton, L. and Bouter. M., 2008. Assessing the Impact of Health Technology Assessment in the Netherlands. *International Journal of Technology Assessment in Health Care*, 24(3): 259–69.

Pang, T., Sadana, R., Hanney, S., Bhutta, Z.A., Hyder, A.A. and Simon, J., 2003. Knowledge for Better Health: A Conceptual Framework and Foundation for Health Research Systems. *Bulletin of the World Health Organization*, 81(11): 815–20.

Peckham, S. and Willmott, M., 2007. The Impact of the NHS Service Delivery and Organisation Research and Delivery Programme 2001–2006. London: (National Co-ordinating Centre for NHS Service Delivery and Organisation Research and Development).

Powell, W.W., Koput, K.W. and Smith-Doerr, L., 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly*, 41(1): 116–45.

Roessner, D., Carr, R., Feller, I., McGeary, M. and Newman, N., 1998. *The Role of the NSF's Support of Engineering in Enabling Technological Innovation: Phase II.* Arlington, VA: SRI International.

Rothwell, R., 1977. The Characteristics of Successful Innovators and Technically Progressive Firms (with Some Comments on Innovation Research). *R&D Management,* 7(3): 191–206.

Rothwell, R., Freeman, C., Horlsey, A., Jervis, V.T.P., Robertson, A.B. and Townsend, J., 1974. SAPPHO Updated: Project SAPPHO, Phase II. *Research Policy,* 3(3): 258–91.

Royal Netherlands Academy of Arts and Sciences, 2002. *The Societal Impact of Applied Research towards a Quality Assessment System*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Ruegg, R. and Feller, I., 2003. *A Toolkit for Evaluating Public R&D Investment: Models, Methods and Findings from ATP's First Decade.* Accessed 19 August 2009: http://www.atp.nist.gov/eao/gcr03-857/contents.htm]

Runde, J. and de Rond, M., 2007. Evaluating Causal Explanations of Samples of One. Paper presented at Organization Studies as Applied Science: The Generation and Use of Academic Knowledge about Organizations, Third Organization Studies Summer Workshop, Crete, 7–9 June.

Shah, S. and Ward, J.E., 2001. Outcomes from NHMRC Public Health Research Project Grants Awarded in 1993. *Australian and New Zealand Journal of Public Health,* 25(6): 556–60.

Sherwin, C.W. and Isenson, R.S., 1967. Project Hindsight. A Defense Department study on the Utility of Research. *Science*, 156(3782): 1571–7.

Smith, K., 1997. Economic Infrastructures and Innovation Systems. In C. Edquist (ed.) *Systems of Innovation Technologies, Institutions and Organisations*, pp.86-107. London and Washington, DC: Pinter/Cassell Academic.

Smith, R., 1987a. Comroe and Dripps Revisited. *British Medical Journal* (clin. res. edn), 295(6610): 1404–7.

Smith, R., 1987b. The Roots of Innovation. *British Medical Journal* (clin. res. edn), 295(6609): 1335–8.

Spangenberg, J., Starmans, R., Bally, Y.W., Breemhaar, B. and Nijhuis, F.J.N., 1990. Predictions of Scientific Performance in Clinical Medicine. *Research Policy,* 19(3): 239–55.

Stryer, D., Tunis, S., Hubbard, H. and Clancy, C., 2000. The Outcomes of Outcomes and Effectiveness Research: Impacts and Lessons from the First Decade. *Health Services Research*, 35(5): 977–93.

Trochim, W.M., 2006. *The Research Methods Knowledge Base* (2nd edn). [Accessed 10 October 2008: http://www.socialresearchmethods.net/kb/]

Van den Bosch, F.A.J, Volberda, H.W. and de Boer, M., 1999. Coevolution of Firm Absorptive Capacity and Knowledge Environment: Organisational Forms and Combinative Capabilities. *Organization Science*, 10(5): 551–68.

Walsh, J., 1973. Technological Innovation: New Study Sponsored by NSF Take Socioeconomic, Managerial Factors into Account. *Science,* 180(4088): 846–7.

Weisbrod, B.A., 1971. Costs and Benefits of Medical Research: A Case Study of Poliomyelitis. *Journal of Political Economy,* 79(3): 527–44.

Wooding, S., Hanney, S., Buxton, M. and Grant, J., 2004a. *The Returns from Arthritis Research: Volume 1: Approach, Analysis and Recommendations*. Santa Monica, CA: RAND Corporation.

Wooding, S., Anton, S., Grant, J., Hanney, S., Hoorens, S., Lierens, A., Shergold, M. and Venema, M., 2004b. *The Returns from Arthritis Research: Volume 2: Case Studies*. Santa Monica, CA: RAND Corporation.

Appendix

A summary of key studies and their insights

STUDY	YEAR	KEY FINDINGS AND INSIGHTS
The Sources of Invention (Jewkes et al.)	1958	No single model of funding or organising research is better than another: variety should be supported.
Project Hindsight (Sherwin and Isenson)	1967	Supporting applied research is key for innovation. Industrial and government laboratory settings are key loci for applied research leading to innovations.
TRACES (Illinois Institute of Technology)	1968	Supporting basic research is key for innovation. Universities are a key locus of research feeding into innovations.
Battelle Laboratories	1973	Both basic and applied research are important for innovation. Recognition of an unmet need or technological opportunity, internal management characteristics and technical entrepreneurs are important for innovation. Not all events contributing to an innovation have equal impact (i.e. a move beyond the 'zero order' assumption).
Comroe and Dripps	1977	Basic (non-mission-oriented research) makes important contributions to innovation. Universities, medical schools and hospitals are key loci of innovation.
SAPPHO project (Rothwell et al.)	1974	Determinants of innovation success in the commercial sector include innovator understanding of user needs, efficiency in product development, manager characteristics, efficiency of communications and the nature of marketing and sales efforts.
Rothwell	1977	In addition to factors identified in SAPPHO, the following are important success determinants: embracing innovation as a corporate-wide task, efficient planning and management techniques, after-sales service and user education, willingness to learn from others and adapt, government help, firm size.
Irvine and Martin	1984	Innovation is underpinned by the cross-fertilisation of knowledge from diverse disciplines and stages of research (e.g. basic and applied). More research on the themes of knowledge exchange, interaction and cross-synthesis is needed.
Evered et al.	1987	Undergraduate education and training influences the careers and research success of medically qualified professionals.

Smith	1987a	There is a lack of evidence on how innovation occurs and evidence tends to be too anecodotal. There is a need for more methodologically rigorous research on research.
Smith	1987b	There is a void of integrative conceptual and methodological frameworks to underpin and guide empirical investigations of innovation processes. The relationship between basic science and innovation is unclear and merits further research. The interactions of basic and applied research and the confluence of various disciplines in innovation are also themes that need more research.
Narin	1989	No particular research settings (e.g. large versus small, university versus medical school) is more conducive to innovation than another. Project grants and intramural programmes are most effective means of funding support.
Spangenberg et al.	1990	Organisational factors are important for research performance. These include time spent on research, the volume of human resources, the degrees of craftmanship of the research manager and the frequency of international communication.
Mansfield	1991	The average rate of return on basic academic research is 28%, but varies significantly by sector. Efforts to explore how socioeconomic impacts of research funding can be quantified are exceedingly challenging: they are based on many assumptions rather than facts, and it is difficult to integrate the complexity of research translation processes and impacts into simple calculations.
Australian National Board of Employment, Education and Training	1992	An examination of how research organisation influences outputs. It suggests that groups of between three and five researchers (including postgraduates and technical staff) have a positive effect on research outputs, but no further economies of scale are realised above that critical mass. However, somewhat in contradiction to this finding, the study also suggests that large, well-funded research groups with committed and involved leadership produce more publications of higher impact and international recognition than smaller groups.
Herbertz and Muller-Hill	1995	A contribution to bibliometric measurement of outputs: evidence for the possibility to include measures such as ‘citations per publication’ as a measure of research quality and ‘research costs per citation’ as a measure of research efficiency.
Buxton and Hanney	1996	A research impact evaluation framework and logic process model. This is a framework for evaluating a comprehensive range of potential outputs from research, as well as a way of conceptualising the process through which outputs are created.
PACE Report (Arundel et al.)	1995	Dissemination processes are important for innovation. Conferences, recruitment processes, informal contacts and temporary human resource exchanges complement publications as important channels for knowledge exchange and scientific advancement.

Martin and Salter	1996	Suggest a positive relationship between investment in basic research and economic benefit, but cannot quantify the size of the benefit. Suggest categories of 'intermediary' benefits that researchers can consider in impact studies, which lead to economic benefits in the long-term (e.g. new information, new instrumentation and methodologies, technological skills, access to expert networks, problem-solving skills, spin-off companies, company attitudes to basic research). The study also emphasises the importance of considering both success and failure rates (and case studies thereof), in efforts to draw realistic macro-level estimates of social returns on research investment.
Cronin and Normand	1998	Criteria for evaluating R&D programmes need to integrate both qualitative and quantitative information bases. Single indicators are misleading.
Dawson et al.	1998	Insights on comparative UK and international biomedical research performance: the UK share of biomedical publications remained relatively static in the 1988–95 period, but shares of publications in the cardiology, genetics, nursing and ophthalmology subfields increased. National and international collaboration also increased. The government remained the largest funder of biomedical research at the time, but the investment of private foundations and industry grew over the years.
Roessner et al.	1998	Nurturing human capital and networks of knowledge exchange is key for innovation success. Institutions should support research through involvement in activities such as education, direct research funding, contributions to the knowledge base, research infrastructure, technology organisational leadership and the facilitation of interaction and communication.
Bozeman et al.	1999	Suggest methods to quantify case-study data by scoring the inputs and outputs of research. Look at how input factors (e.g. size of team, collaboration, basic versus applied focus) influence outputs (e.g. publications, patents). Suggest that basic research produces more patents than applied research. Identify four stages in evolution of case-study based evaluation research: (1) historical descriptions; (2) research 'event' based case-studies; (3) matched comparisons of successful and unsuccessful innovations; and (4) conventional case studies (using action research, interviews, surveys, narrative descriptions) complemented with economic and bibliometric techniques (in an attempt to increase methodological rigour and objectivity).
Buxton et al.	1999	Suggestions on concrete actions for collecting data which can inform the payback categories: for example, sending an annual (electronic) payback survey to researchers; collecting supplementary information from bibliographic databases and also from 'grey' literature; collecting samples of research examined using social science approaches for deeper understandings of impact and for triangulation against other streams of evidence. The study highlighted the challenges of dealing with benefit attribution issues and time-lags in impact assessments.
Molas-Gallart et al.	2000	A research impact evaluation framework. It considers the types of outputs that can be expected from specific research, the channels for diffusion and dissemination of outputs to relevant audiences and the forms of impact.
Stryer et al.	2000	A research evaluation framework. Suggest the need to consider multiple levels of impact: research findings (level 1); impact on policy (level 2); impact on practice (level 3); impact on healthcare outcomes (level 4).

Christie and Tansey	2001	Suggest that witness seminars bringing together people associated with a theme or research of interest are a useful way to build a history of research, its evolution and impacts.
Royal Netherlands Academy of Arts and Sciences	2002	A research impact evaluation framework. It focuses on healthcare research and emphasises two key evaluation criteria: (1) research relevance for healthcare providers and healthcare delivery processes; and (2) relevance for policymakers and policymaking.
Australian Society for Health Research (Access Economics)	2003	Suggests a method for quantifying economic returns on health research, using estimates of wellness and longevity gains. The study finds that investments in R&D had surpassed all other sources of raising living standards, and that cardiovascular research had the highest rates of return on investment. However, the authors admit that that there are no solid estimates but only intelligible guesses of wellness and longevity gains.
Hanney et al.	2003a, 2003b	Explore the possibility of looking at basic and applied clinical research done in the past and following it forward to the present day. Demonstrate that a combination of bibliometric and qualitative approaches can be used successfully to gather data on the evolution and range of impacts from research conducted in the past. Contribute to refining analyses of impact: for example by categorising citations in various ways and interpreting bibliometric data in the context of evidence from narratives.
Lavis et al.	2003	A research impact evaluation framework. Suggests the following: that different target audiences merit different measures of impact and that impact assessment processes should include steps of identifying the target group; selecting tailored measures depending on who is promoting research use among the target audiences; selecting measures according to resource availability; identifying data sources; and collecting assessment data.