Chapter 5

EVALUATION OF PUBLIC INVESTMENT IN R&D – TOWARDS A CONTINGENCY ANALYSIS

by

Amir Piric and Neville Reeve
Ministry of Research, Science and Technology, Wellington, New Zealand

Introduction

The purpose of this chapter is to examine and categorise the different dimensions of evaluation within the context of public research investment and examine how this is applied in the case of a real evaluation exercise. We begin by a short discussion of the context for evaluation followed by a review of the major approaches. This is a followed by a discussion of work in progress by the Ministry of Research, Science and Technology in New Zealand to assess research in the context of a series of derived top-level goals reflecting the needs and aspirations of society.

Along with taxes, regulations, tariffs, quotas and licences, public investment represents yet another instrument which governments may apply to secure their policy goals and to manage economic activity. A government invests in R&D in order to realise economic, social, environmental and cultural benefits for the community it represents. As such, the justification for public investment in R&D should be subject to scrutiny and review as with all other areas of public decision making.

This, however, was largely not the case until relatively recently. Post-World War expansion in the scope of the public research effort in most countries was built on a shaky and incomplete model of the innovation process which tended to stress inputs rather than process, and as a consequence treated innovation as a black box. The rationales for public investment in scientific research incorporated a tangle of issues ranging from its strategic and prestige value through to more practical aims in which governments focused on large-scale research and technology initiatives. Little thought was given, however, to such obviously crucial issues as the best means to identify priorities, efficient allocation mechanisms and the importance of measuring the outputs and wider effects of the research effort. Research in the 1950s on the role of technology as a factor input and other work which began in the 1960s on the nature of technical change itself and the key drivers of this process began to open the lid on the box. Funding pressures which began to take effect in the 1970s and the more recent radical change to the role and position of the public sector which have been a feature in many counties have reinforced the need for a better understanding of why governments should invest in research, how they should do it and what the public gets in return. For the allocation of public funding the overriding key objectives are efficiency, transparency and accountability. These represent the policy goals which evaluation will seek to support.

Evaluation of this investment aims to determine both the costs and benefits of publicly financed projects in R&D and can be used to justify public investment in R&D and improve the efficiency and effectiveness of that investment. It is crucial, however, to recognise that R&D investments should reflect a wide set of issues and not be concerned solely with the more narrow interpretations of "economic". In principle all relevant factors affecting national welfare should be considered, since the object is welfare maximisation and not maximisation of purely economic gains (United Nations, 1972).

In selecting the most appropriate evaluation instrument a number of key decisions are required. Fundamentally, it is necessary to determine when the evaluation should take place and to establish a relevant evaluation framework appropriate to the type and quality of information to be obtained. The evaluation framework will reflect the nature of what is being evaluated as well as the information requirements of those initiating the evaluation. The establishment of a relevant evaluation framework is the most crucial point of the entire process, and is linked with goals/objectives against whom the results/outputs and outcomes are measured.

There are basically three types of evaluation according to the stages at which it is applied:

◊ retrospective *ex-post* evaluation;

◊ monitoring interim evaluation;

◊ prospective *ex-ante* evaluation;

Added to this are the two broad separations between qualitative and quantitative techniques. While these are not hard-and-fast separations, and indeed elements of each approach are likely to be mixed to produce some form of composite, it is useful to briefly explore some of the key aspects in each case.

Ex-ante evaluation is based on the stratification and management of subjective analysis. Methods are often complicated, can be time-consuming to implement and depend for their accuracy on key assumptions on the likely outcome of a diverse set of variables including the development of scientific knowledge and, according to the circumstances, the change in markets.

Ex-post methods are based on hard data and are considered more reliable, but are less useful when making assessment for future projects.

For example, the results of *ex-post* evaluation can be used as a supporting argument and means to identify trends in *ex-ante* evaluation. In some cases, quantitative methods reflect better the outcomes of certain sectors, while qualitative methods are more applicable in sectors where the ultimate outcomes are intangible. Both qualitative and quantitative approaches, or at least some of their elements, can be applied at the same time and as part of the same evaluation. The combination of different methods depends, of course, on circumstances, available data sources, characteristics of the examined sector, etc. For instance, it is quite a common practice to combine both peer review with case studies and cost-benefit analysis.

Many of the most important outcomes of R&D investment, *e.g.* new knowledge, skills and experience, are intangible and unquantifiable, their benefits may not be realised for some years and their impact may be felt in entirely unrelated areas (Swann, 1996; and SPRU, 1996). This illustrates the complexity of the evaluation issue, but it does not mean that we cannot experiment with different methods and be innovative in attempts to create a method that reflect our needs and priorities.

A survey and synthesis of evaluation methods is provided in Tables 1 and 2 of the Appendix.

Qualitative methods

Let us briefly examine three of the most widely applied types of qualitative approach; peer review, case studies and technological forecasting.

Although sometimes being used in an *ex-post* sense, *peer review* is typically used as one of the major techniques in *ex-ante* funding allocation decisions. The central principle is of a panel of experts in respective fields evaluating proposed projects or programmes based on their own judgements and expertise. The relative simplicity and economy of this method makes it popular, but on the other hand it remains limited in several aspects. There are several forms of peer review (Gibbons and Georghiou, 1987):

- ◊ Direct peer review
- ◊ Modified direct peer review
- ◊ Ancillary peer review
- ◊ Traditional peer review
- ◊ Indirect peer review
- ◊ Pre-emptive peer review

The crucial starting point of the peer review is the selection of the expert panel. The panel should be as objective as possible, with clear formulation of related questionnaires and data-collection procedures. There are basically two approaches in selecting a panel: internal and external. The internal panel is composed of local experts in the field, *i.e.* academics, professionals, policy makers, etc. This type of panel selection is useful in large countries where the possibilities for selection of experts are more numerous due to well-developed S&T systems, and where the possibility of subjective evaluation is minimal. The external panel is mainly composed of foreign and internationally recognised experts in the respective field. This type of panel selection is used in some European Union Member States as well as in the United States, Australia and New Zealand.

Both approaches can be criticised, the first for its inability to cope with local lobbying of interested groups within the scientific community, and the second for lack of knowledge of external evaluators on certain particularities of the respective country and the possibility that the expertise might be misleading due to the different scientific environments prevailing in different countries. In addition, peer review sometimes ignores wider social and economic effects due to its self-oriented and highly scientific approach. Certain authors (*e.g.* Bozeman and Melkers, 1993) underline that peer review is quite conservative in analysis and therefore not able to acknowledge other fields' scientific achievements. The problem that occurs in peer review procedures is the identification of relevant research performance indicators, and in such circumstances the method usually employs indicators based upon the reward system and publications, *e.g.* bibliometrics. The results of peer review differ among many countries that have applied this method. In some countries, peer review has failed in evaluating some proposed projects, while in other countries peer review has partly justified its application. Perhaps an alternative solution for peer review is to combine both internal and external panel selection, along with the specification of performance indicators that are closely linked with the desired outcomes.

Another major type of qualitative method is the *ex-post case study*. Quite often, the case study is also considered as a quantitative method, or both. The major advantage of the case study is that it is based in hard data, *i.e.* based on already completed projects where it is much easier to identify performance indicators since the outcomes are "visible". The case study method usually improves judgements in support of science and provides a mechanism for accountability to the public. It is a proven and recognised technique, and it has been extensively used in various fields. The case study technique usually has a clearly defined agenda and the results are well-based, but the major problem is in the interpretation of the results due to the possible subjectiveness of evaluators. Comparability between sectors remains a major problem as is also the wider applicability of conclusions. To overcome these problems, case studies are often used in conjunction with other techniques. Sometimes the case study technique is used as a part of other methods, *e.g. ex-post* cost-benefit analysis based on specific case studies.

Technological forecasting methods are applied as a part of larger priority-setting process. The aim is to determine technological developments, based on assumptions of future events. They provide certain directions with regard to which fields of research should be fostered or remodified. There are three basic methods (Capron, 1992): scenario methods; cross-impact matrices (or interdependency matrices); and morphological analysis. The *scenario method* is based on the creation of several scenarios of future developments and is quite similar to Delphi-based approaches. *Cross-impact matrices* are used to identify interactions among different research fields after certain events have taken place or as part of a larger foresight exercise that forecasts the probabilities of interaction between different research fields. *Morphological analysis* is a method of technological forecasting that combines assessment methods and creativity techniques. The weaknesses of these methods are their inability to forecast certain events that might change assumptions and the subjectivity of the experts that have performed the exercise.

Quantitative methods

Turning now to quantitative approaches we will review eight of the main approaches.

Financial methods are resource-intensive and can provide accurate calculations of the efficiency and distributional effects of research. The results are usually calculated over a long period using Internal Rate of Return (IRR) and Net Present Value (NPV). These methods are widely used in private sector R&D investment evaluation processes. One of the advantages of these methods is the optional application via software that makes this process much faster and easier. The major obstacle in applying these methods in public investment procedure is that public investment in R&D has completely different connotations from conventional investment.

For instance, financial methods do not provide significant insight into the strategic agenda and they are focused only on economic factors. They are also focused on a few goals and therefore are quite inflexible and limited. The push to use these methods usually comes from governmental departments that are by nature quite conservative and are looking for calculation of returns on the public investment in dollar values. This represents a significant obstacle to the reasonable explanation of all benefits that come out of the investment in R&D. Therefore, their use has to be carefully considered, and perhaps limited to evaluations of particular projects in which the outcomes are much easier to express in economic terms. The main financial evaluation methods are (Capron 1992; United Nations 1978; United Nations 1980):

- ◊ Cost-benefit/cost-effectiveness analysis
- ◊ Programming models
- ◊ Risk profiles
- ◊ Ratio methods
- ◊ Portfolio models

Quantitative methods are relatively well-developed, and their application is substantial in various sectors. For the purpose of more focused analysis, we will be discussing the cost-benefit analysis because it seems that use of and interest in it is rather substantial.

The most prominent method, due to its particular characteristics, is *cost-benefit analysis* (CBA). The purpose of the analysis is to justify and explain the social benefits and costs of certain economic projects in terms of a common monetary unit. The aims of CBA are (IRMD, 1995):

- ◊ to ensure that investments in certain projects are prudent and cost effective;
- ◊ to methodically document a project's costs and benefits;
- ◊ to demonstrate the costs and benefits of a project over the estimated life cycle of the resource;
- ◊ to address viable alternatives and to aid in the selection of the best solution.

CBA is based on relatively sound economic principles, *i.e.* conventional NPV and IRR (Levy and Sarnat, 1990). The benefits are expressed in a form of cost-benefit ratio (CBR), and recognise that capital investments produce social benefits over a certain period of time. The increase of net benefits should be parallel to the increase of social utility or social welfare. CBA has found its application in all public expenditure, but only because this area lacks clear output measures due to the long-term nature of many public projects in which benefits occur after many years.

CBA frequently faces the problem of "measuring the immeasurable". In particular CBA has problems dealing with externalities that have been produced by R&D and requires identifiable projects for evaluation, and since most R&D projects are characterised by a high degree of sophistication and following externalities, this is a problem for the science and technology (S&T) community. Even though CBA is primarily focused on *ex-ante* evaluation, the data for analysis come from *ex-post* analysis of more or less well-defined and completed projects, and therefore it has limited accuracy. In addition, CBA does not provide significant insight into strategic objectives since it focuses only on economic factors. Apart from several drawbacks, CBA could be applied in certain sectors and on certain levels that are able to produce significant data and consistent actualisation rates along with tangible outputs. CBA is not able to calculate spin-off or external effects of R&D activities. The major problem with CBA is the fact that it requires measurable factors in financial terms. It is also quite difficult to implement CBA since it requires specialised skills and demands significant resources.

An additional quantitative *ex-ante* method is *technometrics* . This is a relatively new method that was specifically designed to analyse the industrial development phase of technological change (Meyer-Krahmer and Reiss, 1992). It employs the output indicators from technical specifications of products and processes, and is a useful tool in observing the evolution of technological change. The main indicators for technometrics are collected from the technical specifications of the products or processes that are the subject of evaluation. The selection of the indicators is conducted by consulting scientific literature and experts in the relevant field.

Technometrics are quite useful for practical problem solving, *e.g.* the analysis of technological gaps and identification of technological processes that should be modernised. Technometrics are also useful for designing specific topics of public research programmes that are created to target specific technological issues. Technometrics could be employed both as the strategic and operational elements of *ex-ante* evaluation. The main deficiencies of technometrics are this method's limitation to just a few sectors and its exclusion of technological disparities and domestic competition within a national economy in analysis. Therefore, it is impossible to use it across various sectors and circumstances. It also has difficulties in identifying data for certain international comparisons due to the differences in technical standards around the world and disparities in technical measures. Nor does it provide significant input to the overall evaluation of public investment in R&D. It could be a useful tool for the detailed and relevant explanation of the situation before a programme is started, yet it has to be developed, and in particular the identification of indicators needs improvement. Technometrics should not be used alone in identification of technological gaps, rather it should be used along with other methods in order to avoid potential misidentification. This method finds some applications in industries that are facing technological gaps in production or trying to foster technological improvement of already existing products.

Optional pricing (OP) is applied, by and large, in financial markets, and most recently in evaluation of different investment opportunities that are under certain restrictive conditions. Some authors have experimented with the application of OP to evaluation of R&D projects, and have indicated the two most common types of OP used in evaluation of R&D (*e.g.* Newton, 1992):

◊ provision of an analogy which will help in persuading investors of the value of R&D projects;

◊ provision of OP with numerical data as an alternative evaluation method.

OP has been applied in finance, while the empirical demonstration in R&D evaluation is still meagre due to the relative paucity of data. Compared with the financial market information, the analogous R&D information is less quantitative and frequently not expressed in financial terms. The alternative for those financial terms is a type of substitute in the form of different qualitative outcomes, *e.g.* "reasonable", "optimistic" and "pessimistic" merits in assessment of outcomes. This is similar to a decision-tree approach, but the major difference is that OP uses an appropriate discount rate rather than an arbitrarily chosen discount rate. The crucial point is that the value ascribed to an option evolves with the time that is analogous to the R&D project implementation. OP usually employs the statistical assumptions that are linked with "random walk" and "Brownian motion", *i.e.* fluctuations in future stock prices should follow geometric "Brownian wave" and the future outcomes are distributed log-normally.

The advantages of OP are that no decision-tree analysis is required and a more comprehensive set of future options is covered, while the only key number that is required to delineate the set is the volatility. Volatility, in this context, is the expected standard fluctuation of stock prices, which is based on previous experiences in the respective field. The most used technique in estimating volatility is a time series linked to recent historic data. The option price can be calculated by using several factors: exercise price, stock price, constant-time at expiry, variable-time, risk-free interest rate and volatility. The risk-free interest rate is the rate on government bonds over the respective period, and since the public investment in R&D is committed by a government, the same rate should be applied. In the evaluation of R&D projects, the data should include the aggregates and timing of cash inputs and outputs and certain estimates for each project's extra value which is generated for the respective organisation. Data should be collected for a set of projects which at the beginning of each project were considered to be equally likely to produce any one of a range of possible outcomes. The number of projects in this set should be large enough for a statistically useful curve of number of outcomes *vs.* profit/loss to be obtained, generating estimates for the return and standard deviation in the usual way that is applied in financial analysis. The drawbacks of this method are lack of relevant information and a quite modest empirical record in R&D evaluation.

The *scoring method* involves marking the list of projects against a list of prepared criteria which have predetermined weighting. Part of the information may be collected through published sources, although mainly the information descends from scientists. They have the ability to rank a long list of projects on both qualitative and quantitative grounds. The method is good for systematising and simplifying the decision making involved at any level of the research system. The results are relatively transparent, which facilitates their understanding by researchers and administrators. They can be used to incorporate multiple goals and objectives. The more sophisticated scoring methods allow for sensitivity analysis of the more important criteria and also iteration procedures. In general, the method remains subjective, both in terms of the criteria and the weighting attached to criteria, lacks flexibility, tends to rely heavily on statistical information and can be subject to abuse. Individuals involved are often asked to give opinions on subjects in which they have no expert knowledge. As with peer review, this approach is most useful for comparing similar proposals and is limited to evaluating the economic returns. These methods do not incorporate any form of discounting or research spillovers. Scoring methods are repeatedly applied in project selection work, but they provide only a very crude estimate of efficiency or distributional results. The scoring methods are usually based on matrix or systemic approaches.

Bibliometrics, *S&T indicators* and *patent data* are also *ex-post* evaluation methods. Their application is widely used for identification of the national S&T system. *Bibliometrics* is the study of publication-based data and it is quite accurate for comparative analysis. It includes publication counts, citation counts, co-citation analysis, co-word analysis, scientific "mapping" and citations in patents, etc. Bibliometrics seems to be most useful in basic research for comparative reasons, but beyond that its use is quite limited and inflexible. Therefore, it is usually necessary to combine bibliometrics with other methods in order to gain some conclusion and it has extremely limited utility for R&D evaluation. Bibliometrics is frequently used as a supporting tool in describing the stage of development of the national S&T system. It does not provide the full coverage of publications and double-counting is quite frequent. On the other hand, *patent data* could be a good indicator of innovation activity, but the criteria applied in various countries for the registration of patents differ significantly. The other question is how many of the registered patents have been commercialised. Not all patents get a chance to be marketed. Economists and science policy analysts are not interested in the count of patents as such; they are interested in patents only if they provide a measurable yardstick of a much wider phenomenon – inventive and innovative activities. Therefore patent data does not necessarily reflect the capacity to innovate of the national S&T systems.

So far, *econometric methods* are the only ones which are able, even partially, to meet the requirement of measuring R&D's contribution to economic growth and its direct and indirect effects at the macroeconomic level (Capron, 1992). Econometric methods are statistical, regression-based methods which incorporate economic theory. Using these methods demands the collection of historical data on production, inputs, prices, past research expenditure, and so on, to statistically assess the relationship between research and some measure of output, usually productivity growth.

These models have been widely adopted for *ex-post* analysis and they have been used to justify an increased expenditure on R&D. Econometric methods have produced significant results across various sectors and were adopted in several countries. Contemporary versions of econometric methods are based by-and-large on the Cobb-Douglass production function with a Solow-neutral exogenous embodied technical change. There are several versions of this basic concept, all of them with certain commonalities and outcomes. Along with the improvement of the method, *i.e.* insertion of new economic factors into the calculation, interpretation of results has developed as well. Owing to the disadvantages of other methods, *i.e.* both qualitative and quantitative, econometric methods have been widely accepted, in particular because they cover certain elements and produce certain parameters that cannot otherwise be inserted or produced. Even then, there are several studies that indicate difficulties in estimating the main impact parameters and the value of results, due to (Capron, 1992):

- ◊ theoretical and methodological problems and the availability of statistics;
- ◊ relevance of past production and technical progress to future tendencies;
- ◊ aggregation bias and omission of some variables.

In addition, the production function ignores non-measurable factors, such as transfer of technology at both the micro and macro level, social and cultural dimensions and the stage of development of the national S&T system.

An additional problem with econometric methods is the inability of these methods to measure the entire stock of knowledge, the accumulation of which is the key outcome of public investment in R&D. There have been some attempts to measure the stock of knowledge, *e.g.* indicators such as share of S&T staff in the labour force, number of new graduates and post-graduates from universities, etc., and this information is useful, but it could not possibly cover the entire stock of knowledge of a nation.

Evaluation of S&T in New Zealand

Not including the normal processes of research monitoring and auditing, there have been several attempts in New Zealand to evaluate in a more long-term sense the benefits of public investment in R&D. The most prominent have been the evaluations of meat research, by the Foundation for Research, Science and Technology in 1996, and wool research, by the Wool Research Organisation of New Zealand in 1994. Despite their methodological complexity, in both cases a combination of case studies and cost-benefit analysis, these evaluations have generated wider debate among stakeholders in New Zealand on this issue, as well as providing a solid background for further work.

Since the beginning of 1996, the Ministry of Research, Science and Technology has taken the initiative in evaluating public investment in R&D. The aim of this evaluation initiative is to establish a process which should generate useful information and guidelines for the allocation and management of public investment, as well as feed into the decisions on priority-setting. The approach adopted by the Ministry can be described as a “logical paradigm” consisting of open-ended and long-term scientific, economic, environmental and social goals, followed by research-area-specific short- and medium-term objectives, using both qualitative and quantitative performance indicators which measure progress towards the top-level goals.

In developing the evaluation approach a major concern was to reflect the outcome focus of recent policy initiatives adopted by the Ministry. In recent years New Zealand has undergone some radical changes to the structure of the public sector involving the dismantling of many of the existing institutions and relationships. A key concern has been the separation between policy making, funding and the provision of services. For the science community this has meant a restructuring of the previous Department of Scientific and Industrial Research (DSIR) which has been replaced with nine so-called State Owned Entities, in this case called Crown Research Institutes (CRIs). Accompanying these structural changes have been measures to improve accountability and the focusing of research efforts. One, perhaps inevitable, consequence of this process has been the tendency to monitor achievement according to specified and, in terms of their impact, limited outputs. Whilst such measures are indicators of successful activity, it is abundantly clear that public support to research in science and technology cannot be justified simply on the basis of achievement audits. Thus the emphasis for performance measurement and hence evaluation has been shifted towards assessing "real" beneficial change. Trial evaluations of two research areas – forestry and eco-systems – have commenced recently, and the aim is to expand evaluation across all research areas covered by the Public Good Science Fund – the largest component of R&D investment in New Zealand. At its heart the approach has combined both questionnaires and structured interviews. Other components have included an analysis of the research area to gain an impression of the relationship between the New Zealand effort and international activity, in addition to the collection of data for a bibliometric exercise.

The scope of the evaluation exercise is wide-ranging (see Appendix), reflecting both the fact that the aim is to trial different techniques, as well as the requirement that the impacts and benefits of research are assessed at a level which provides real added-value to New Zealand. In keeping with the aims of all evaluation exercises it has been important to develop an approach which is transparent, which provides clear evidence on efficiency and beneficial gain and which enhances the accountability.

Concluding comment

By way of a summary, let us review some of the main points to have emerged from this brief discussion.

We have shown how evaluation of research can be classified according to two major criteria: the timing of the evaluation, essentially whether it occurs before, during or after the research; and the type of information to be collected. We have seen also that evaluation is a complex process reflecting the complexity and uncertainty of the processes of research and innovation and the demand for useful information to inform decision making in this area. Indeed, it has been noted that in many cases a variety of different evaluation methods are combined as a means to address these uncertainties and deliver a richer set of information.

The composition of respective evaluation approaches will depend amongst other things on the nature of the research sector, how big the research portfolio is and on the respective objectives of the research. In certain sectors, *e.g.* agriculture and dairy, CBA has been widely used, while in, for instance, environment and eco-systems, peer review and case studies have been applied (the fields of application are summarised in Table 2 of the Appendix). Finding useful indicators of research performance is perhaps the central difficulty facing all evaluation approaches for there is often a tension between what is measurable and what is interesting. Although this applies in relation to both *ex-ante* and *ex-post* assessments it is however the latter where the most interesting questions are found since these are often crucial to the rationale for publicly supported research. To list just a few of these questions: what has been the impact as measured in financial returns; the quality and scale of knowledge produced; human resource development; or the standing and competence of the research performing system.

In this discussion we have made a key distinction between “outputs” and “outcomes”. Others sometimes refer to this as the difference between effects and impacts, although the issues are the same. By “outputs” we mean the process developments from research and technology development that are routine and do not in themselves represent anything more than achievement at the level of the project or programme. Such measures will include, of course, papers and other forms of codified knowledge dissemination, artefacts such as prototype products, processes, new techniques and possibly services. None of these taken alone, however, constitute a benefit that would provide justification for the spending of public money on research. For this we have to consider a broader set of factors which, according to the terminology set-up, are referred to as “outcomes”. These are the results which are likely to provide material benefit towards quality of life through a variety of mechanisms including greater prosperity, more effective management of the environment and improvements to the structure and operation of society. Thus, we might include increased industrial competitiveness, as measured perhaps in terms of market shares and market size, the quality of the environment, and more effective social regulation. Neither should we forget the benefits to skills and learning and the ability to have an informed and capable scientific infrastructure.

The question arises how best should we measure the broader outcomes from research. This has been the central issue in the evaluation framework now proposed by New Zealand’s Ministry of Research, Science and Technology (MoRST). As a general principle we may say that the structure of a national economy and national strategic priorities determine the character of public investment in R&D. The top-level goal of public investment in R&D is the generation of new knowledge in areas of strategic importance for a country. To measure the value of this new knowledge the approach put forward by MoRST is to make assessments in four areas; economic, scientific, environmental and social outcomes. It follows that these outcomes are *de facto* the returns on public research investments and may be used to build a case in support of such investment. The scale of outcomes ranges from different innovations within respective programmes at the micro level, to accumulated and extended knowledge in some industries at the meso level, up to increased competitiveness on international markets of a national economy and improved environmental management at the macro level, etc. Regardless of the method or combination of methods used to evaluate public investment in R&D, the eventual aim must be to explore the link between the generation of new knowledge and the contribution to a wide set of possible outcomes at the level of the economy, science, environment and society.

A-state-of-the-art assessment of current evaluation practices around the world shows that there are a great many methods available, although it also fair to say that the literature on the assessment of benefits from public research is rather modest by comparison, as also is the empirical evidence based on methods that have been applied. The approach of New Zealand’s Ministry of Research, Science and Technology should therefore be seen in the context of a vanguard initiative to extend the reach of existing evaluation methods and identify the bridge which links the government’s goals for science and technology with long-term outcomes which are the aspirations of society.

Appendix

Table 1. Evaluation methods

Methods	Qualitative and semi-qualitative		Quantitative	
ex-ante	**Peer reviews** **Questionnaires** **Interviews**	**Techn. forecast. methods** Scenario method Cross-impact matrices Morphological analysis	**Cost-benefit analysis** **Technometrics** **Optional pricing**	**Scoring methods** *Matrix approaches* - analysis matrices, -decision-making matrices, -multicriteria analysis, -relevance trees *Systemic approaches* -system analysis, -dynamic modelling
ex-post	**Case study** **Histographical tracing of accomplishments** **Critical scientific events**		**Quantitative indicators** Bibliometrics S&T indicators Patent data **Econometric models**	**Financial methods** *Ratio methods* *Risk profiles* *Programming models* *Portfolio models*

Table 2. Synthesis of evaluation methods

Adapted and expanded from Capron (1992)

Method	Relevance	Drawbacks	Field of application
Peer reviews **Questionnaires** **Interviews** **Technometrics** **Case studies**	- evaluation by experts in field - screens of projects and research orientations - relative simplicity - widely used method - significant experience	- subjectivity of experts - partial forecasts - lack of independence of experts - does not allow to measure the global economic impact - expensive and time consuming	- selection and technical evaluation - technological forecasting - identification of technological gaps - improving judgements in support of science and accountability to the public
Scoring methods matrix approaches: - analysis matrices - decision-making matrices - multicriteria analysis - relevance trees	- rich information - decision-making process - rationalise and simplify choices - profiles projects and R&D planning - provide lots of information	– difficult, even impossible to collect the required information - subjectivity - lack of flexibility - the number of statistics required is substantial - requires constituting a specialised group - subjective choice of criteria and weighting - strongly empirical -subjectivity in the allocation of quantitative values	– evaluation of the industrial impact of R&D expenditure - multicriteria interpretation - project selection - emphasizing the links between different research projects, technology and economy
Systemic approaches: - systemic analysis - dynamic modelling	– R&D strategies - appropriate to select projects - takes the evolutionary character of the economy into account - includes social, historical and ecological structures - Takes feedback phenomena into account	– not really suitable for evaluating as such - very difficult to implement	– selection and control - analysis of the evolution of a system and its adaptability
Financial methods: -cost-benefit/cost-effectiveness analyses - ratio methods - risk profiles - programming models - portfolio models	– measure marketable outputs and commercial resources - simple instruments	– difficult to collect the information - some factors cannot be measured or financially assessed - the actualisation rate is difficult to choose - do not allow to take R&D externalities into account - difficult to estimate time-lag between R&D - highly variable results - subjectivity in the choice of the success probability and of criteria - purely financial aspects	– financial evaluation of a project - measurement of the *ex-post* return - financial evaluation - project selection - determining the financial lump sum to invest in R&D
Optional pricing	- provides optional values of R&D	- difficulties with adequate data	- experimentally applied in few sect.
Technological forecasting methods: - Scenario method - Cross impact matrices - Morphological analysis	– allows to reverse the causality chain - takes social transformations into account - overcome the problem of interdependencies between questions - discontinuous character	– subjectivity always present	– selection and technical evaluation - technological forecasting
Quantitative indicators: - S&T indicators - Bibliometrics - Patent data	– easy measurement - measure technical resources - builds up fundamental research indicators	– purely descriptive - does not take the indirect effects into account - micro-macro cross-cutting - not well suited for evaluating development - partial information	– measuring how efficient the R&D input is at the macro level - analysis of the evolution of a system and its adaptability
Econometrics	– the only general quantitative method available for evaluating the economic impact of R&D expenditure	– theoretical and methodological background - availability of statistical material - aggregation bias - not well-suited for forecasting	– evaluation of the impact of R&D expenditure upon the economy

Table 3. Social and private rates of return in 37 innovations

Mansfield (1981)

Innovation	Rate of return (%) Social	Private	Innovation	Rate of return (%) Social	Private
Primary metals innovation	17	18	Industrial product A	62	31
Machine tool innovation	83	35	Industrial product B	negative	negative
Compo. for control system	29	7	Industrial product C	11	55
Construction material	96	9	Industrial product D	23	0
Drilling material	54	16	Industrial product E	37	9
Drafting innovation	92	47	Industrial product F	161	40
Paper innovation	82	42	Industrial product G	123	24
Thread innovation	307	27	Industrial product H	104	negative
Door-control innovation	27	37	Industrial product I	113	12
New electronic data	negative	negative	Industrial product J	95	40
Chemical product	71	9	Industrial product K	472	127
Chemical process A	32	25	Industrial product L	negative	13
Chemical process B	13	4	Industrial product M	28	23
Major chemical process	56	31	Industrial product O	62	41
Household cleaning device	209	214	Industrial product P	178	148
Stain remover	116	4	Industrial product Q	144	29
Dishwashing liquid	45	46	Industrial product R (sic)	103	55
			Industrial product S	29	25
			Industrial product T	198	69
			Industrial product U	20	20

Table 4. Estimates of private rates of return to R&D

Edwards (1996)

Author	Year	Industry	Rate of return (%)
Griliches	1964	Agriculture	53
Mansfield	1965	Petroleum	40
		Chemicals	30
Evenson	1968	Agriculture	57
Minasian	1969	Chemicals	50
Terleckyj	1974	33 mfg and non-mfg firms	28
Mansfield	1977	16 significant innovations	25
Nathan	1978	20 significant innovations	35
Foster	1978	17 significant innovations	24
Griliches	1979	80% of US industrial R&D	32-40
		Chemicals	93
		Metals	23
		Machinery	24
		Motor vehicles	25
		Electrical equipment	5
		Aircraft	3
Mansfield	1982	20 mfg innovations	25
Link	1982	Petroleum	21
Odagri	1983	370 Japanese scientific firms	26
Clark, Griliches	1984	924 business units	20
Odagri, Iwata	1986	135 Japanese firms	20
Suzuki	1989	Drugs (Japan)	42
		Electrical (Japan)	22
Lichtenberg, Siegel	1989	5 240 firms	13
Goto, Suzuki	1989	50 industries (Japan)	26
Griliches, Mairesse	1990	525 firms	25-41
Hall, Mairesse	1992	196 firms (France)	22-34

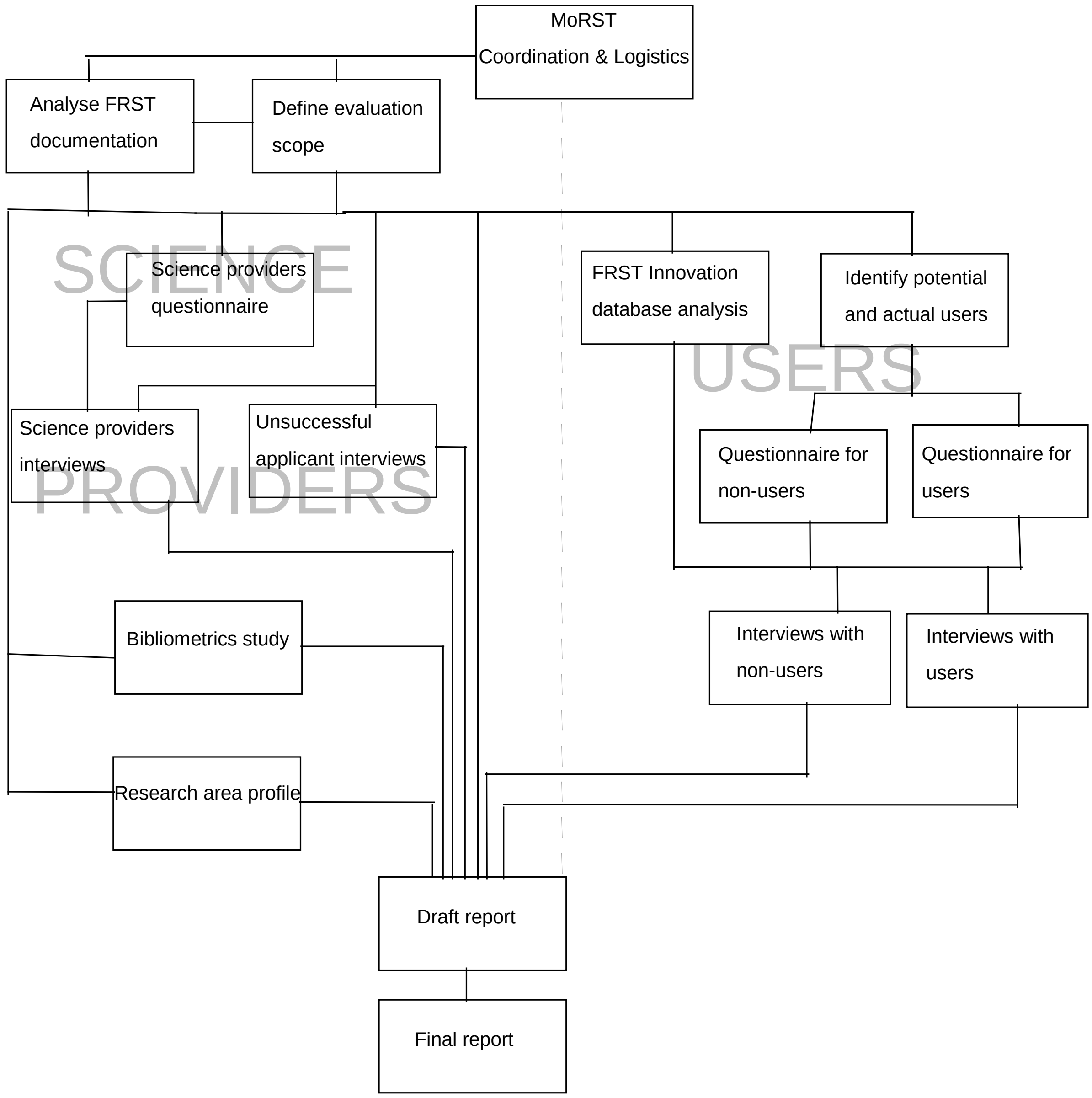
Table 5. Social rates of return to R&D

Edwards (1996)

Author	Year	Industry	Rate of social return (%)
Terleckyj	1974	33 mfg and non-mfg firms	80
Mansfield	1977	17 significant innovations	56
Nathan	1978	20 significant innovations	70
Foster	1978	17 significant innovations	99
Mansfield	1982	20 mfg firms	70

Figure 1. Science providers and users

Science Providers	- Crown Research Institutes, Research Associations, Universities, Private Sector, etc.
Science Users	- Government Agencies, Private Sector, etc.
MoRST	- Ministry of Research, Science and Technology
FRST	- Foundation for Research, Science and Technology



REFERENCES

BOZEMAN, B. and J. MELKERS (1993), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, MA.

CAPRON, H. (1992), *Economic Quantitative Methods for the Evaluation of the Impact of R&D Programmes*, EUR 14864 EN, Monitor – Spear Programme, CEC, SRD DG XII, EU, Brussels.

EDWARDS, F. (1996), in Winsley, P. (1996), "R&D as a 'Socially Efficient' Investment", FRST Discussion Paper No. 4, Foundation for Research, Science and Technology, Wellington.

GIBBONS, M. and L. GEORGHIOU (1987), *Evaluation of Research, A Selection of Current Practices*, OECD, Paris.

IRMD (1995), *Cost-Benefit Study Procedure*, Department of Administrative Services, Salem.

LEVY, H. and M. SARNAT (1990), *Capital Investments and Financial Decisions*, Prentice, Hertfordshire, UK.

MANSFIELD, E. (1981), "How Economists See R&D", *Harvard Business Review* 59(6), Harvard.

MEYER-KRAHMER, F. and T. REISS (1992), *Government Programmes, Ex-ante Evaluation and Technology Assessment – Two Emerging Elements of Technology Policy Evaluation*, Research Evaluation Beech Tree Publishing, Surrey, UK.

NEWTON, D. (1992), "Application of Option Pricing Theory to R&D", Working Paper No. 237, MBS, Manchester, UK.

ROMER, P. (1994), "The Origins of Endogenous Growth", *The Journal of Economic Perspectives*, Volume 8, No. 1.

SPRU (1996), *The Relationship Between Publicly Funded Basic Research and Economic Performance*, Science Policy Research Unit (SPRU), University of Sussex, Brighton, UK.

SWANN, P. (1996), *The Economic Value of Publicly Funded Basic Research*, Programme of Policy Research in Engineering, Science and Technology (PREST), University of Manchester, Manchester, UK.

UNITED NATIONS (1972), *Guidelines for Project Evaluation*, New York.

UNITED NATIONS (1978), *Guide to Practical Project Appraisal, Social Benefit-Cost Analysis*, New York.

UNITED NATIONS (1980), *Manual for Evaluation of Industrial Projects*, New York.