technopolis |group|

17 June 2009

Evaluating the National Innovation Strategy for Competitiveness

Erik Arnold
Daniel Malkin
Barbara Good
John Clark
Miriam Ruiz Yaniz

www.technopolis-group.com

technopolis [group]

technopolis |group|

Table of Contents

1. Introduction 1

2. The National Innovation Council for Competitiveness (CNIC) and the Innovation Governance System in Chile 4

2.1 Institutional Setting 4
2.2 The CNIC 6
2.3 The Strategy 8
2.4 A Logical Framework 10
2.5 The Main Agencies 12

3. Evaluation Principles 15

3.1 What Is Evaluation? 15
3.2 What Can Be Evaluated? 16
3.3 Evaluation Models 17
3.4 Why Evaluate? 18
3.5 Key Evaluation Concepts 19
3.6 Limits to Evaluation 20
3.7 I-O-O-I – A Generic Model for R&D Evaluation 22
3.8 Summary of Evaluation Methods 24
3.9 Authors' Views and Assessment of Methods 26

4. Evaluation at Three Levels in Chile 29

4.1 The I-O-O-I Model at Three Levels 29
4.2 The State of the Art in Evaluating Research and Innovation in Chile 31
4.3 Statistics and Indicators 35

5. Evaluation at the Programme Level 40

5.1 Making the I-O-O-I Logic Specific to Programmes 40
5.2 Use of Methods at Programme Level 43
5.3 Use of Indicators in Evaluation at Programme Level 57

6. Evaluation at the Organisational Level 57

6.1 Making the I-O-O-I Logic Specific to Organisations 57
6.2 Use of Methods at Institutional Level 57

7. Evaluation at the Strategic Level 57

7.1 Making the I-O-O-I Logic Specific to the Strategic Level 57

7.2 Use of Methods at Strategic Level 57
7.3 Indicators at the Strategic and Systemic Levels 57
8. Managing an Evaluation System 57
8.1 Process and Timetable 57
8.2 Guidelines on Evaluation 57
8.3 Evaluation Capacity 57
8.4 Quis Custodet? 57
Appendix A - Evaluation Toolbox 57
A.1. Introduction 57
A.2. Macroeconomic and Econometric Modelling 57
A.3. Micro-level Economic Analysis 57
A.4. Cost-Benefit Analysis 57
A.5. Useful Statistical Applications 57
A.6. Analysis of 'Spillovers' 57
A.7. Surveys of Beneficiaries 57
A.8. Peer and Panel Reviews 57
A.9. Case studies 57
A.10. Historical tracing 57
A.11. Network Analysis 57
A.12. Prospective studies 57
A.13. Bibliometrics and Patent Analysis 57
A.14. Webometrics 57
A.15. Meta-Evaluation 57
A.16. Organisational Review 57
A.17. Benchmarking/Comparative Study 57
A.18. History 57
A.19. Summary of Evaluation Methods 57
A.20. Authors' Views and Assessment of Methods 57
Appendix B - Summary of the Innovation Strategy for Competitiveness 57

Table of Figures

Figure 1 The national Innovation System Heuristic 3
Figure 2 Governance of State Innovation Funders in Chile 7
Figure 3 Main Elements of the Strategy 9
Figure 4 A Generic Logical Framework 11
Figure 5 Innovation Strategy Intervention Logic 12
Figure 6 Innovation Strategy Implementation Monitoring 12
Figure 7 Other Processes Similar to Evaluation 15
Figure 8 Hierarchy of Performance Contracts and Evaluation 17
Figure 9 Comparison of formative and summative evaluation 18
Figure 10 Timing of Evaluations 19
Figure 11 Key Evaluation Concepts 20
Figure 12 The Inputs-Outputs-Outcomes-Impacts Model 23
Figure 13 Overview of evaluation methods 24
Figure 14 General relevance of evaluation techniques 26
Figure 15 The I-O-O-R Model Across Three Levels 30
Figure 16 Generic Evaluation Issues at Three Levels 30
Figure 17 Methods Used in Ten Evaluations 33
Figure 18 Example of a Simple Intervention Logic 40
Figure 19 Examples of Indicators Relevant to Programme Evaluation 42
Figure 20 The RACER Framework for Programme Indicators 43
Figure 21 Categories of Innovation Intervention 45
Figure 22 Shared intervention logic of innovation policy measures 46
Figure 23 Modern 'Coupling' Model of Innovation 47
Figure 24 Creating New, Innovative Firms: Evaluation Methods 49
Figure 25 Funding and Credits to Reduce Innovation Risks: Evaluation Methods 50
Figure 26 Increasing Absorptive Capacity: Evaluation Methods 51
Figure 27 Internationalisation: Evaluation Methods 53
Figure 28 Technology, R&D Infrastructure: Evaluation Methods 53
Figure 29 Technology Support Largely Based on Existing Knowledge: Evaluation Methods 54
Figure 30 Technology Support Largely based on New Knowledge: Evaluation Methods 55
Figure 31 R&D Support from the Public Research System: Evaluation Methods 56
Figure 32 Researcher-Driven Research: Evaluation Methods 57
Figure 33 Science Communications: Evaluation Methods 57

Figure 34 'Horizontal' Initiatives: Evaluation Methods 57

Figure 35 Evaluation Cycle for Innovation Policy 57

Figure 36 Impacts of R&D support programmes 57

Figure 37 Factors of interest in micro-economic models 57

Figure 38 Costs and benefits of publicly-supported R&D, by stakeholder 57

Figure 39 Principal component analysis 57

Figure 40 Factors influencing spillover potential 57

Figure 41 Innovation Agency Process Model 57

Figure 42 Overview of evaluation methods 57

Figure 43 General relevance of evaluation techniques 57

Figure 44 The Strategy 57

1. Introduction

In line with the high priority given to the promotion of innovation to foster its economic and social development in a sustainable way, Chile has developed an ambitious strategy for innovation and competitiveness.[1] This is a handbook intended to support the National Innovation Council for Competitiveness (*Consejo Nacional de Innovación para la Competitividad* – CNIC) in its work to monitor and evaluate the progress of its strategy.

The CNIC was formed to advise the Government on how to develop a holistic innovation policy, with the overarching aim of increasing GDP per head quickly enough to repeat the doubling of the previous fifteen years. A key instrument for this was the newly created Innovation for Competitiveness Fund (FIC), including an additional stream of revenue, derived from a levy on mining, for supporting innovation, and part of the CNIC's initial role has been to advise the Government how to use the FIC. More recently, it has advised the government on strategy, commented on the government's proposed research and innovation budgets and cooperated with various implementation agencies in developing specific policy measures or instruments and considering how to monitor and evaluate their effects.

The CNIC was originally set up by presidential decree in 2005 as an interim body that had a mandate to develop an initial national strategy for innovation and competitiveness. When it reported in 2006, the President expanded its membership a little and made it into a permanent body. The Council then deepened its strategy work, producing Volume 1 of its White Paper ("*Towards a National Strategy*") early in 2007 and Volume 2 at the start of 2008. A key feature of the strategy has been that – contrary to previous doctrine – it introduced 'selectivity' into innovation policy, primarily by prioritising eleven industrial clusters, of which eight were initially implemented.

The Inter-ministerial Committee on Innovation has endorsed this strategy, which includes a number of strategic orientations supported by detailed policy recommendations to be implemented by various agencies such as Corfo, Conicyt, Fia operating under the authority of sectoral Ministries. Over the recent years and up to 2009, substantial increases of budgetary appropriations to these agencies as well as those channelled through the innovation Fund for Competitiveness (FIC) have provided significant resources to support the various priority programmes and policies and deemed a priori necessary to ensure their success.

To a large extent, policy recommendations made by the CNIC take into account those made by international organisations such as the OECD[2] and the World Bank[3] that have reviewed Chile's innovation system highlighting its specific weaknesses, adapting these organisations' recommendations to the Chilean institutional context. However, in their reviews both organisations emphasise that in a context of the consolidation of Chile's innovation system characterised by evolving institutional capabilities, clearer priority setting and increased resources appropriation but still impaired by problems of policy fragmentation and coordination among implementation agencies, the monitoring and assessment of innovation policy as well as that of the performance of the innovation system as a whole called for an increased attention on the part of Chilean authorities responsible for the governance of the system and policy

[1] Consejo Nacional de Innovación (2008), *Hacia una Estrategía Nacional de Innovación para la Competitividad*, Santiago de Chile.

[2] See OECD (2007), *Chile - OECD Reviews of Innovation Policy*, Paris.

[3] World Bank (2008), *Chile: Towards a Cohesive and Well Governed National Innovation System*, Washington, D.C.

implementation. It is argued that in Chile, which is a relative latecomer in the development of S&T indicators,[4] strengthening monitoring, benchmarking and assessment capabilities throughout the system is a prerequisite to steer it towards institutional and policy reforms, adaptive policy mixes that, in the last instance, condition an efficient translation of strategic orientations into a portfolio of support programmes and policies aiming at strengthening the performance of that system in terms of socio-economic outcomes. As a latecomer Chile is in a better position to "leapfrog" in the development of its indicators development and evaluation capabilities along recognised best practices, taking advantage of conceptual and methodological advances made in more developed countries to adapt their system of indicators to evolving policy needs.[5]

Implementing the strategy involves three levels of activity, each requiring separate evaluation. At the highest level, the strategy itself should be evaluated, in order to understand the effectiveness and efficiency with which it is moving Chile towards its strategic goals. Second, the role of the implementing agencies should be evaluated, primarily in order to identify opportunities to improve their practices and effectiveness. Third, the programmes and other instruments used to bring the concrete measures proposed in the strategy to their beneficiaries and to create spillovers should be evaluated.

Enhanced assessment capabilities should be developed at the following levels: policy and support programmes or instruments; policy delivery and performance of implementation agencies; public institutions, such as universities and research and technological centres, that perform an active role in knowledge generation, diffusion and commercialisation; overall performance of the innovation system that includes the monitoring of the determinants of innovation, the evaluation of outcomes of policy mixes and the consistency between and/or complementarity across policy areas.[6]

The idea of a National Innovation System is central to this report – and of course to what the CNIC is trying to achieve. The original discussions of innovation systems used a narrow definition. When Freeman[7], who coined the term, referred to national systems of innovation he meant the state funding and research-performing apparatus. Soon, however, as it became clear in the discussion of innovation systems by people like Richard Nelson[8] and Bengt-Åke Lundvall[9] that innovation performance depends upon a much greater set of institutions and actors across the whole of society. We follow this broader tradition. In our discussion, an innovation system or a research and innovation system is roughly that which is illustrated in Figure 1.

[4] The first exercise in compiling S&T related statistics and indicators was undertaken in 1995 with the launching of the first innovation survey by the Ministry of Economy.

[5] In the framework of its Group of national Experts in S&T Indicators (NESTI) created in 1963 the OECD has continuously worked towards the an improvement of the scope and the measurement of S&T indicators trough regular updating of dedicated manuals that take into account evolving policy needs and academic advances in the understanding of the determinants of innovation performance (e.g. Frascati, Oslo, Canberra and patents manuals cited in references).

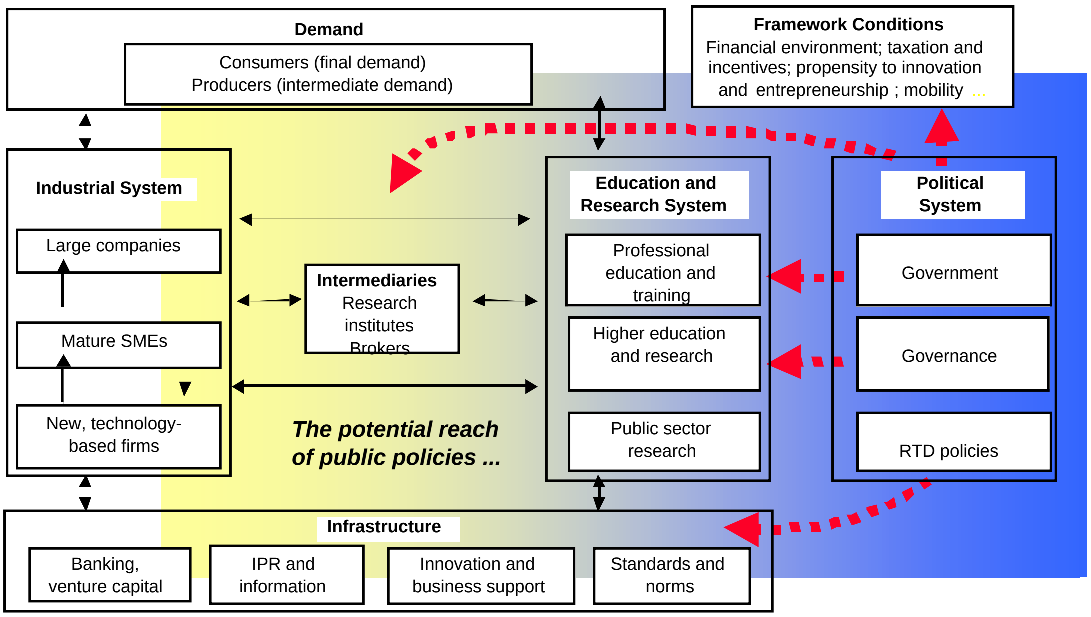
[6] See OECD (1997), *Policy Evaluation in Innovation and Technology: Towards Best Practices*, Proceedings, OECD, Paris

[7] Christopher Freeman (1987), *Technology Policy and Economic Performance: Lessons from Japan*, London: Frances Pinter

[8] RR Nelson (1993), *National Innovation Systems*, New York: Oxford University Press

[9] Bengt-Åke Lundvall (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London: Pinter

Figure 1 The national Innovation System Heuristic



Source: Erik Arnold and Stefan Kuhlman (2002), *RCN in the Norwegian Research and Innovation System*, Background Report No 12 in the Evaluation of the Research Council of Norway, Oslo: Royal Norwegian Ministry for Education, Research and Church Affairs. Also available at www.technopolis-group.com

In this handbook, we begin (in Chapter 2) by discussing the institutional set-up within which the CNIC and its strategy operate, taking special interest in how relevant aspects of this set-up are evaluated today. In Chapter 3, we set out some broad principles of evaluation that provide coherence to an overall approach to evaluation. Chapter 4 considers in turn how to evaluate the three levels: programmes; organisation (agencies); and the overall strategy. It looks at the current state of the art in evaluating research and innovation in Chile and goes on to discuss the state of existing indicators. The following three Chapters (5–7) look respectively at the levels of programmes, organisations and the overall strategy for innovation. Each describes how to use the I-O-O-I idea to develop evaluation criteria relevant to its respective level, makes suggestions about which evaluation methods to use and discusses the role of indicators in evaluation. We then discuss in Chapter 8 how to manage the overall system of evaluation, in particular how to coordinate evaluation across the various actors involved, proposing a cycle of evaluation and strategic review.

The handbook builds upon a repertoire of evaluation tools used internationally. Rather than interrupt the flow of the handbook, we have put this toolbox at the end of the report, so that the reader can dip into it as needed. The toolbox can also be read as a self-standing document. It describes and exemplifies various evaluation techniques, discussing their strengths and weaknesses.

The CNIC's full strategy is published elsewhere. A second Appendix tabulates the goals of the strategy, something we suggest in the report should be the basis of the CNIC's ongoing monitoring and reporting of progress to the public and the government

2. The National Innovation Council for Competitiveness (CNIC) and the Innovation Governance System in Chile

This Chapter discusses the institutional setting, within which research and innovation policy exists in Chile, the CNIC, its strategy, places the strategy within a logical framework, and briefly describes the main agencies responsible for implementing the strategy.

2.1 Institutional Setting

In line with the concerns expressed by international organisations and largely shared by the CNIC, Chile has initiated an ambitious series of studies that should ultimately result in the implementation of a comprehensive system of monitoring and assessment of its innovation strategy. The relevance and efficiency of this system is predicated upon

- The development of a regularly updated base of science and technology (S&T) statistics and indicators compiled in accordance with international standards[10]
- The ability to extend this base to address issues more specific to Chile's innovation system
- The ability to launch specific ad hoc qualitative and/or quantitative evaluations of support programmes and develop specific performance accounting procedures that can affect resources allocation among institutions and/or programmes
- The ability to detect positive or negative externalities induced by, or arising among, support programmes within the framework of a systemic view of the innovation system
- The attention paid to the resources required to maintain the information system and the burdens this may pose to the beneficiaries of the support programmes
- An appropriate institutional setting for the steering of the information base necessary not only to monitor and assess the performance of the innovation system, but also to disseminate intelligence among stakeholders and draw the lessons of monitoring and assessments on evolving policy priorities, adapted policy mixes, and policy delivery, including their possible termination

As emphasised in OECD (1997)[11], beyond the variety of methodological approaches that can be used, "the institutional set-up within which programmes and policies are evaluated in effect determines the nature, quality, relevance and effectiveness of evaluation practices". This institutional set-up shows a wide variety among OECD countries, from more centralised systems such as the ones in place in France, Japan or the United Kingdom to more decentralised ones as in Netherlands and Germany.[12]

In Chile when the innovation system is still in the process of consolidation it seems that a clear distinction between the CNIC and other institutions, in particular the Ministry of Economy, should be made as regards their respective responsibilities in

---

[10] Chile has committed itself to meet this requirement as quickly as possible in the context of its application to OECD membership. Necessary resources would be appropriated by the Ministry of Economy

[11] Papaconstantinou, G. and W. Polt (1998), "Policy Evaluation in Innovation and Technology: An Overview", Chapter 1 in OECD (1997)

[12] Notwithstanding this variety, in most countries the Ministry of Finance can or must request or conduct an ex-ante or ex-post cost-benefit analysis of support programmes. This is typically the case in Chile where the DIPRES performs that function

the process of monitoring and evaluation of innovation policies, and the development of S&T and innovation related indicators.

Evaluations can indeed have different purposes. Policy makers are mainly interested in those that highlight the efficiency of programmes, their additionality and impact on innovation behaviour of beneficiaries and the implications on evolving policy mix and policy complementarity issues. They are also concerned by the management and delivery performance of implementing agencies. Such agencies are generally more interested by the leverage effects of the support or incentive programmes they manage and their effectiveness with respect to their stated objectives. Additionally, they are more and more concerned internal monitoring management performance based on quantitative and qualitative indicators of policy delivery procedures and beneficiaries' satisfaction.

As the body responsible for defining the innovation strategy proposed to government approval and detailing this strategy into policy to be implemented by specialised agencies the CNIC plays a pre-eminent role in the steering of the innovation system, notably as regards the evolution of the policy mix, the improved coordination among agencies and the eventual streamlining of policies or support programmes. In fulfilling this role, the CNIC should be one of the main users of evaluations that have systemic implications and help identify better practices that cut across policy areas and, ultimately, have implications on the allocation of resources to support programmes, financial instruments and public institutions involved in research and technological development. The CNIC should have the capability to undertake or commission such evaluation exercises, to entrust them to implementing agencies or to contract them out to independent external experts. This view is in line with the recommendation made by the OECD in its review of Chile's innovation policy.[13]

This however leaves open two important questions. On the one hand that of assessment of specific individual policies or support programmes; on the other that of the development and management of an S&T and innovation database of statistics and indicators, in particular the internationally comparable ones requested by international organisations.

In the present institutional setting the Division of Innovation of the Ministry of Economy and which effectively also acts as the Executive Secretariat of the Inter-Ministerial Committee on Innovation chaired by the Minister of Economy could in fact be entrusted with the responsibilities of coordinating and overseeing the functions related to:

- The development and management of the databases of S&T and innovation statistics and indicators (including innovation surveys), in compliance with international standards as set out in relevant OECD manuals
- The provision of statistical information requested by the OECD for the purpose of benchmarking exercises and international comparisons of innovation policies
- The development of ad hoc indicators and surveys as required for evaluations requested by the Inter-Ministerial Committee, the CNIC or the Ministry of Economy
- The monitoring and evaluation of specific innovation policies implemented by agencies under the Ministry of Economy or that cooperate with these in the context of jointly administered programmes

[13] "Until now, there has been no official permanent organisation in charge of monitoring and evaluating Chile's innovation policy, which is able to provide an overview of the system and assess progress towards its overall consistency. Today, however, there is a consensus that such an evaluation body should be attached to the new established National Council for Innovation for Competitiveness", in OECD (2007) pp. 185et seq.

These functions would imply close relationships – involving the possibility of joint design of indicators and surveys – with institutions such as the National Institute of Statistics (INE) and agencies in charge of implementing R&D and innovation policies, such as primarily CORFO (including InnovaChile), CONICYT, the Ministry of Higher Education and sectoral ministries responsible for Public Technological Institutes.

It is to be noted that such an institutional setting is in accordance with the one presented in the Innovation for Competitiveness project agreed between Chile and the World Bank in 2008 with a loan from that institution and to be implemented in the next five years.[14]

2.2 The CNIC

The National Innovation Council for Competitiveness (CNIC) was originally set up by presidential decree in 2005 as an interim body to develop an initial national strategy for innovation and competitiveness. In response to the report of the interim Council, President Michelle Bachelet re-formed the CNIC in May 2006 and made it into a permanent body, with a mandate to advise the President of the Republic on all aspects related to policies in the area of innovation including the education of specialised human resources and the development, transfer and diffusion of technology. The mandate of the Council set out in Decree No 505 is to

- Publish strategic proposals
- Establish mechanisms to consult and enter dialogue with relevant actors, especially the regions
- Propose a national strategy for innovation for competitiveness
- Propose how to allocate the FIC tax on mining companies
- Undertake studies
- Make proposals for institutional redesign

Based on this mandate, the CNIC produced Volume 1 of 'Towards a National Innovation Strategy for competitiveness' early in 2007 and Volume 2 at the start of 2008. In parallel, it launched its' flagship 'clusters' project.

Figure 2 shows (in simplified form) the governance relationships in the Chilean research and innovation funding system. In practice, this involves the 'two pillar' system used in most countries, with an industry ministry running an innovation agency[15] and an education ministry running a research council. (In this report, we use 'research council' to mean an agency that funds research, primarily in the higher education and/or research institute sectors. Such agencies normally have a majority of academics in their governing bodies. In the past, many research councils 'owned' research-performing organisations but today most are pure funding bodies, not research performers. Some may give advice to government about research or science policy but for our purposes here that is not a necessary part of their job. In Europe, the research councils are networked via the European Union Research Organisations Heads of Research Councils – EUROHORCS – organisation.) Other ministries (such as Agriculture, with its FIA agency) are also involved in research and innovation funding but to a lesser degree and their innovation activities are mixed with other sectoral policy concerns.

---

[14] In effect, this project includes a component that "would support the establishment of a Monitoring and Evaluation Unit within the Ministry of Economy directly reporting to the Minister. The unit will be responsible for (i) providing both technically sound information on the status and trends in key indicators of innovation, in particular, and related competitiveness areas; (ii) performing international benchmarking of the aforementioned information or indicators; and (iii) evaluating the impact and effectiveness of policies and publicly funded instruments to stimulate innovation and competitiveness."

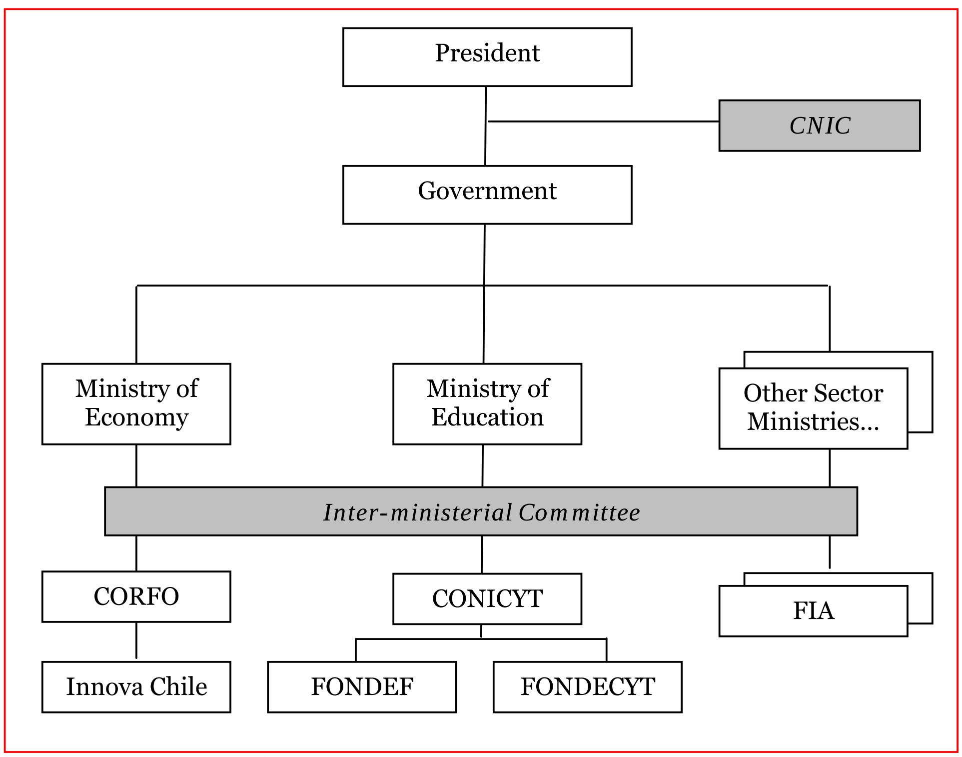
[15] Strictly, CORFO is not an agency of the Ministry of Economy but a ministry-like organisation whose head is appointed by the President. In practice, however, CORFO plays the societal role of an innovation and business development agency

The CNIC, whose design was influenced by Finland's Science and Technology Policy Council, is intended to solve the problems of coordinating national research and innovation strategy that are in fact encountered in all countries. It provides advice on research and innovation strategy to government (largely, but not only, in the form of a proposed strategy). An inter-ministerial committee of ministers concerned with innovation decides what parts of this advice to take. The Ministry of Economy provides a secretariat to the Committee but all affected ministries and their agencies are expected to adopt and implement policies consistent with the advice accepted by the inter-ministerial committee. This means in practice that the actors shown in Figure 2 are expected collectively to implement the strategy while at the same time no single actor is able to instruct all the others what to do. Both the strategy and the clusters project therefore make significant demands on the ability of actors to work together 'vertically' (between ministries and their agencies) and on 'horizontal' projects across administrative boundaries and the regional and national levels.

The composition of the CNIC has varied a little over time but reflects the fact that it can coordinate but not instruct the government research and innovation actors by involving the key actors in the Council itself. The CNIC comprises

- Two experts in public policy, one of whom chairs the Council
- Fourteen other members, comprising a mixture of academics, industrialists and heads of foundations. These include the head of the Chilean Academy of Sciences and a trades union representative
- The Ministers of Economy, Education, Finance and Agriculture
- The Executive Vice-President of CORFO, the President of CONICYT and the head of CNIC's secretariat attend as observers

Figure 2 Governance of State Innovation Funders in Chile



Note: Shaded boxes refer to strategy coordination structures

Figure 2 in effect shows that there is a mechanism for making strategy (the CNIC), one for implementing it (the inter-ministerial Committee) but no feedback or monitoring mechanism that allows CNIC to monitor and report on the progress being made towards the goals of the strategy. This report contributes to the construction of this monitoring and feedback mechanism.

2.3 The Strategy

The overarching goal of the strategy is to double Chilean GDP per capita over the next fifteen years to 2021, repeating the doubling of the previous fifteen. To achieve this, the first volume of the strategy states that

- Total Factor Productivity must rise, driven by greater knowledge-intensity, technological change, human capital and innovation
- Average years of schooling must rise to 12 by 2010 and 14 by 2021, while the proportion of the cadre of 18-24 year olds entering higher education must rise from 43% today to nearly 80% by 2021; Chile should make significant progress in the results achieved in OECD's international PISA evaluations
- R&D investment should rise from 0.68% of GDP in 2004 to some 2.3% by 2021 and the business share of this R&D investment should rise from 37% to 50% of that total
- Dependence upon a small number of economic sectors must be reduced. At the time, the 25 main items comprised 76% of exports. This should fall to below 50%
- The country's position as measured by international innovation and economic indicators should improve[16]

The CNIC published a second volume of its strategy[17] in January 2008. It stresses the need to move towards a knowledge economy, in line with developments in much of the rest of the world. It also points out that in a time of increasing global specialisation, Chile cannot afford to try to be good at everything and must therefore specialise. It argues that Chile must build its innovation strategy on three major pillars

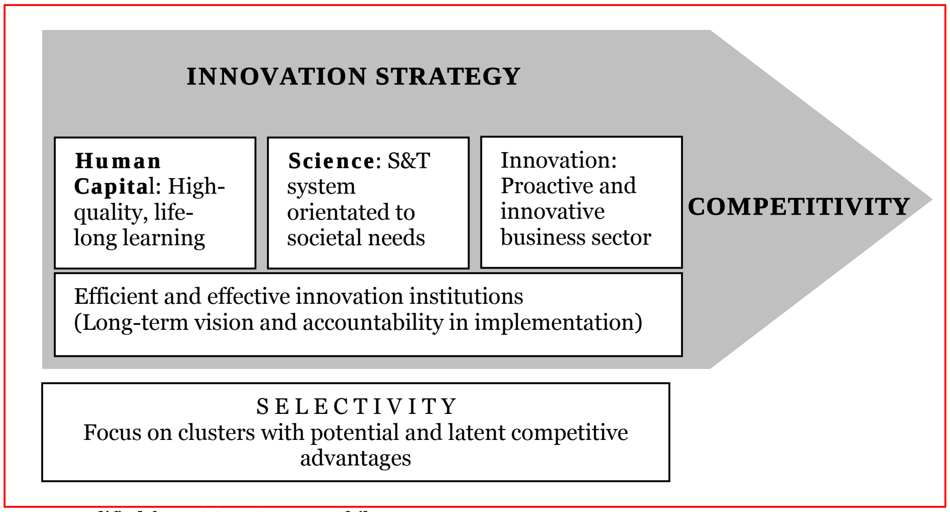
- High-quality, life-long learning
- A science and technology system orientated towards social needs
- A proactive and innovative business enterprise sector

For each pillar, the CNIC proposes concrete measures intended to secure the goals it suggests (see Appendix).

[16] CNIC (2007), *Hacia un Estrategia de Innovación para la Competitividad*, Volumen I, Consejo Nacional de Innovación para la Competitividad, Santiago

[17] CNIC, (2008) *Hacia un Estrategia de Innovación para la Competitividad*, Volumen II, Consejo Nacional de Innovación para la Competitividad, Santiago

Figure 3 Main Elements of the Strategy



Source: Modified from CORFO InnovaChile

The report argues that human capital development is a major challenge, to which the education and training system must respond through growth with social equity. Chile needs to progress by

- Establishing a system of flexible life-long learning covering not only initial education but also subsequent re-skilling and re-training
- Developing an integrated system of higher education and qualifications, in order to allow people not only to acquire high-level skills and qualifications but also to navigate non-traditional courses through the education system and enable mobility
- Develop a system of technical and vocational education based on skills relevant to work
- Modernise university curricula so that they are relevant to national and international needs
- Strengthen and increase the throughput capacity of the system of higher education and associated qualifications

The CNIC proposes to tackle these issues through the creation of a system of life-long learning, securing the quality and relevance of education and qualifications and increasing the capacity of the education system, especially to tackle technology-intensive areas and those parts of the economy where incomes are low.

Science for development is the second challenge, which must be tackled by

- Finding a new balance between curiosity-driven and "relevant" research
- Deciding on which great questions to focus publicly funded research
- Designing a funding system to meet the challenges – tackling the full costs of research, assuring quality through peer review and respecting national priorities while being internationally networked and multidisciplinary
- Strengthening research capacity

To tackle these needs, the CNIC proposes to strengthen and increase the activities of the science base, providing incentives to encourage research in strategically relevant areas; reform the science funding system, which will still be focused on CONICYT; and to provide the grant funding needed to meet the national challenges through training researchers, strengthening institutions, quality-assuring postgraduate training, increasing doctoral training, ensuring the system can absorb the increases in manpower and encouraging immigration by relevant researchers.

Improving innovation in business is the third challenge, and addressing it has to involve an element of selectivity. The main needs identified in the strategy are

- Incorporating new knowledge into production by raising the level of R&D
- Encouraging the diffusion of new technologies
- Supporting innovative entrepreneurship

To do this the CNIC proposes to strengthen the development of clusters through road mapping, increasing relevant research capacity and attracting foreign direct investment. It aims at

- Incorporating new knowledge into production processes by forming R&D consortia, increasing support funding, increasing innovation skills among company managements, improving the tax treatment of business R&D and increased development and protection of intellectual property rights
- Strengthening technology diffusion by establishing a network of technology institutes, creating a network of technology brokers, setting up world class technology diffusion centres for ICT and other technologies needed by innovating firms and creating incentives for innovation among companies currently operating at low technological levels
- Increasing innovative entrepreneurship through busies incubators, promoting spin-offs from existing companies, encouraging greater availability of seed and venture capital, developing instruments to encourage business angels and reducing the way bankruptcy currently tends to prevent entrepreneurs from trying again with new business ideas

Implementing these three pillars requires institutional changes

- Establishing institutional roles based on a clear division of labour
- Improving the governance of state institutions, especially by making a clear distinction between policymaking in the executive and implementation by agencies
- Strengthening the roles of the CNIC, management within the Executive branch and the implementing agencies
- Reforming and strengthening the sub-systems responsible for science and for business innovation
- Empowering and strengthening the capacity of the regions

The main points of the strategy are summarised at the Appendix.

The principle of **selectivity** – choosing to support certain clusters and not others or prioritising strategic over curiosity-driven research – is a significant departure in Chilean policy. In the past, selectivity has been associated with 'capture' of policy by interest groups, with the aim of diverting state subsidies to themselves. Selectivity also runs counter to the traditional position of the economics profession, which holds that firms themselves are the more efficient mechanisms for allocating resources and therefore that the state should not in any sense try to 'pick winners'. Internationally, the pendulum swung away from selective innovation policies in the 1980s and has now begun to swing back. These swings seem more associated with changes in political perspective than evidence about the effectiveness of policy. There is no a priori way of settling which policy is more effective – though there is growing agreement on the need for specialisation (especially in small economies) in the face of globalisation, and this necessarily entails selectivity.

2.4 A Logical Framework

Logical Framework Analysis (LFA) is an increasingly widely used tool for planning projects and small programmes. In principle, it can also be applied to more complex entities such as the Innovation Strategy. It consists of a chain-link logic (following the arrow at the left of Figure 4) that essentially says, "If we do certain activities, they will produce outputs that trigger outcomes and eventually contribute to achieving our overall objective". Expressing this 'intervention logic' in a logical framework makes it easier to consider whether all the necessary steps in the intervention logic are present. We can then think about what indicators can be used of progress at each level and how to verify these. In the case of the Innovation Strategy, that will involve a lot of hard and soft indicators, while for a project it may be simpler to find such indicators. Finally, the logical framework encourages us to ask what assumptions about the context or others' actions we are depending upon in order to move from one level to the next above it. For example, what do we assume about the activities when we expect them to lead to outputs?

Figure 4 A Generic Logical Framework

Intervention Logic	**Verifiable Indicators**	**Means of Verification**	**Assumptions**
Overall Objective	Impact Indicators	Monitoring and Evaluation Systems	
Outcomes	Outcome Indicators		
Results (Outputs)	Output Indicators		
Activities	Process Indicators		

Normally, when we draw intervention logics we expand the left-hand column and of the framework and do the remainders of the analysis elsewhere. A complete intervention logic for the Innovation Strategy would be massive – as can be inferred from the length of the summary in the Appendix to this document. But the structure of the Strategy is effectively that shown in Figure 5.

Figure 5 Innovation Strategy Intervention Logic

Intervention Logic				
Overall Objective	Improve the performance of the Chilean economy so as to redouble GDP per head within 15 years			
Outcomes	Improved human capital	Better, more relevant science	More innovative businesses	Modernised Institutions
Results (Outputs)	See Human Capital Chapter …	See Science Chapter …	See Business Innovation Chapter …	See Chapter on Institutional Reform …
Activities				

It is possible to 'nest' logical frameworks. For example, we could take the "Improved human capital" Outcome and draw a new logical framework where that was the Overall Objective. That would force us to be more detailed about the activities. In practice, the Innovation Strategy provides enough detail about the activities to allow them to be handed over to others (sometimes policymakers, sometimes agencies) for implementation. This use of agency is consistent with the CNIC's role as advisor to the government rather than itself being an operational actor. At the highest level, therefore, CNIC has the potential transparently to monitor (for example on its Web Site) the implementation of its strategy by expanding on the summary of its strategy provided in the Appendix to this report as shown in Figure 6. Together with a review of macro indicators of economic and innovation performance, and an account of annual changes in the strategy, this would also provide a clear basis for the annual reporting with which the CNIC has tasked itself.

Figure 6 Innovation Strategy Implementation Monitoring

Challenge	Actions	Agent	Baseline	Progress	Effects on the Challenge
For example. 'Strengthening research capacity'	For example, "Ensure sufficient researchers are trained'	Who is responsible for the action? In this case, CONICYT	What baseline data, indicators or studies are available?	In qualitative terms, what has so far been done with this Action?	Output information from monitoring data; evaluations; a judgement about whether the Challenge has been addressed

2.5 The Main Agencies

This section briefly summarises the characteristics of the main agencies involved in implementing or evaluating aspects of the Innovation Strategy.

2.5.1 CORFO

The Corporación de Fomento de la Produccíon (CORFO) was established in 1939 to foster industrialisation in Chile. Over time its role as owner of nationalised industries has diminished and it is now primarily concerned with private sector development, but still obtains 85% of its income from its investments. Its mission is: To promote the establishment and growth in Chile of innovative, dynamic and accountable firms with access to global markets by supporting high-impact projects that contribute to making Chileans' aspirations for growth and prosperity a reality. Its main objectives are

1. Business innovation
2. Innovation culture
3. Regional development
4. Enabling infrastructure for innovation

CORFO acts in three main areas related to SMEs, Innovation and Regional development. It supports Chilean firms to improve their competitiveness in international markets. Its scope of action ranges from individual companies and networked firms to full production chains, including clusters or geographic groups of companies working in a particular industry. The Agency also promotes the formation of new businesses that renew and diversify opportunities for growth. For this reason, it supports innovative entrepreneurship and investments, especially those that allow Chile to access highly competitive international production networks.

CORFO allocates resources through two main types of measures

- **Credit:** through the private financing system (banks and others). In 2006 CORFO allocated $ 72.47 billion (66%) to 22.815 enterprises (99% micro and SMEs) and $ 38.135 billion (34%) to students below doctoral level
- **Subsidies:** In 2006 CORFO allocated $ 57.355 billion, of which 67% went directly to enterprises and the rest to other beneficiaries (regional programmes, universities, incubators, technological diffusion or promotion of investment)

By means of matching funds and credits, CORFO helps Chilean firms particularly smaller companies face their main challenges

- Innovation: Through Innova Chile CORFO provides matching funds to meet the needs of individuals, companies or groups of companies, universities and technology-related institutions for innovation activities
- Improving quality and Productivity, CORFO provides matching funds to improve the quality and productivity of Chilean small and medium-sized enterprises, thus strengthening both their own competitiveness and that of the overall domestic production system
- Financing: By means of bank and non-bank agents, CORFO provides credits for productive investments and working capital for micro, small and medium-sized enterprises, as well as credits for tertiary education
- Investment: it supports Chilean and foreign investors wishing to undertake projects in Chile by offering multiple services and incentives to facilitate project assessment, location and start-up

2.5.2 CONICYT

CONICYT functions in practice as an agency of the Education Ministry, though the President of the Republic appoints its head. Its responsibilities cover those of research councils in many other countries but also extend beyond into academic-industrial cooperation, which is more usually (but not always) the responsibility of innovation agencies.

CONICYT's mission is to: Promote, strengthen and diffuse scientific and technological research and innovation in Chile in order to contribute to economic, social and cultural development.

To pursue its mission, CONICYT has two specific objectives

- "To visualize and anticipate the country's Advanced Human Capital needs; to coordinate and promote a national and integral training and financing policy, aimed at increasing the number of professionals with graduate degrees in the country and abroad, promoting the attraction of experts to the country and adequate labor insertion into universities, companies and the government" via
  - EXPLORA programme in science communications and pubic understanding of science
  - Advanced Human Capital Programme which provides scholarships ad fellowships to postgraduate and higher-level researchers
- "To consolidate an articulated public support system for fundamental research (basic and applied), that is gradual with regards to: the amount of resources granted, the duration of the initiatives supported and the degree of collaboration and number of researchers involved " via
  - National Fund for Scientific and Technological Development (Fondecyt)
  - Astronomy Programme
  - Fund for the Promotion of Scientific and Technological Development (Fondef)
  - Collaborative Research Programme (PIA)
  - Regional Scientific and Technological Development Program
  - Scientific and Technological Information Programme
  - International Cooperation Programme

FONDECYT was set up within CONICYT in 1982 and provides individual grants for researcher-initiated research. It is the main instrument for 'bottom-up' funding that corresponds to the traditional role of research councils in other countries.

FONDAP was set up in 1999 to fund collaborative centres of research excellence,. Three of the seven existing centres are purely scientific in nature, while the other four have additional 'Basal' money to include industry in the consortia, corresponding to he 'competence centres' movement internationally.

2.5.3 Fundacion para la Innovación Agraria (FIA)

FIA is a private foundation created by the Ministry of Agriculture. It supports the development of regional and sectoral strategies that aim to encourage efficient agriculture that exploits local competitive advantages. It supports innovation, primarily at the planning stage – implementation funds typically come from CORFO. It spreads information about the opportunities provided by new agricultural techniques.

3. Evaluation Principles

This Chapter starts by discussing what we mean by 'evaluation', as the concept is used in different ways in different contexts (and, indeed, in different languages). Id asks 'What can be evaluated?' and explores reasons why it is useful to do evaluation. It is important to touch upon some key evaluation concepts, so we briefly introduce these then go on to discuss some of the limits to what evaluation can do. We go on to introduce the Input-Output-Outcome-Impact model of state intervention, which forms the basis of much modern evaluation and which we use as a structuring device for the sections of this report that discuss evaluation criteria and what methods to use in evaluation. Finally, we summarise our views on available evaluation methods, via two tables that summarise the Evaluation Toolbox, which is shown at the Appendix.

3.1 What Is Evaluation?

There are many formal definitions of evaluation, one being

> *"Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object."*[18]

This report is concerned with the context of research and innovation policy and their implementation. For this purpose, a more public policy-orientated definition is appropriate.

> *"Evaluation examines the outturn of a policy, programme or project against what was expected and is designed to ensure that the lessons are fed back into the decision-making process. This ensures government action is continually refined to reflect what best achieves objectives and promotes public interest."*[19]

Evaluation is an activity distinct from appraisal, monitoring or audit.

Figure 7 Other Processes Similar to Evaluation

Activity	Description
Design Studies	At least in Europe, the process of doing a study to support or test programme design is increasingly referred to as 'Impact Assessment' or 'Ex ante Evaluation'. There are important links between the thinking that programme designers and evaluators need to do, but in our definition evaluation is essentially an activity based on evidence about what has happened rather than an exercise in modelling potential impacts of a future intervention.
Appraisal	The process whereby decisions are made about which projects or programmes will be funded. Although appraising interventions requires some evaluative skill (e.g. judging the relevance of a particular programme), it differs from evaluation, as it is not concerned with the impacts and successes of interventions.
Monitoring	This is the day-to-day collection of information to assist in programme management. The key aim of monitoring is ultimately to ensure programme/project delivery. Monitoring should be systematic and regular whereas evaluation is not necessarily systematic and may occur periodically. Crucially the difference lies in the fact that evaluation involves judging. You can have monitoring without evaluation, but you cannot have evaluation without monitoring. Evaluation therefore relies on effective monitoring procedures.
Audit	The process of auditing is one which ensures that public funds have been spent honestly. It is important that stakeholders appreciate the difference between auditing and evaluation from a buy-in perspective. Stakeholders may feel defensive if they feel they are being financially inspected and scrutinised.

[18] William M.K. Trochim (1999), *Introduction to Evaluation, Research Methods Knowledge Base*, 2nd edition,

[19] HM Treasury (2003), *The Green Book on Appraisal and Evaluation in Central Government*, London, p.45

3.2 What Can Be Evaluated?

Historically, evaluation in a policy context grew up with the development of major social policy programmes. For the purpose if this report, we can think of evaluation as happening at four levels

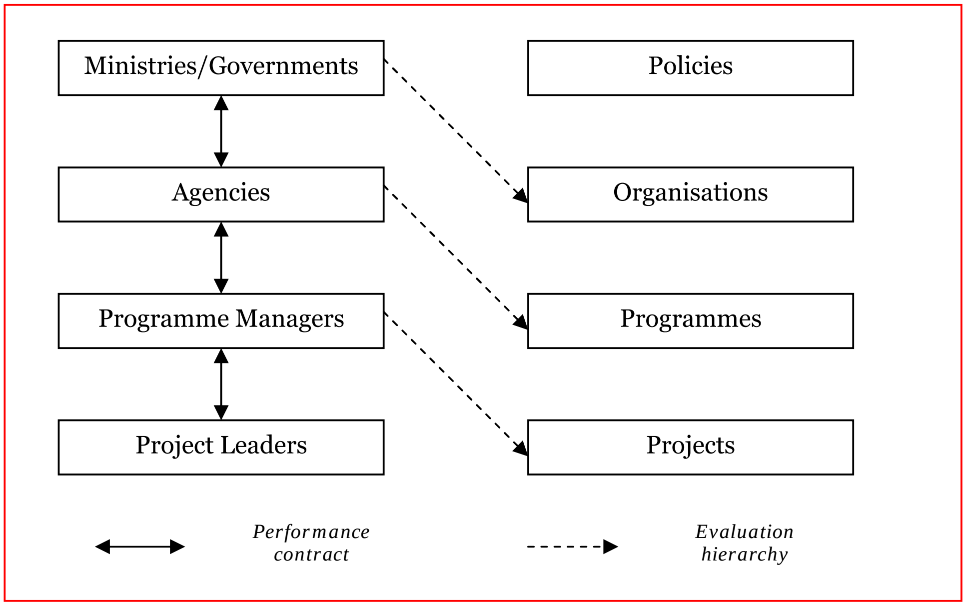
- Policy – the level of the overall set of decisions made by an arm of government or by the government as a collective
- Organisations – the level of the agents that implement policies, typically agencies such as research councils and innovation agencies. Such organisations typically have missions and goals, which are at least partly set by policymakers
- Programmes – normally collections of projects (often technical in nature) that together are intended to achieve a social purpose. In modern practice, programmes tend to address one or more of the goals assigned to agencies
- Projects – individual sets of activities planned to provide a set of technical outputs

Evaluation can also take place at a thematic or portfolio level – for example, a research council may want to evaluate its activities in physics – grouping together relevant programmes or projects.

The distinction between programmes and projects is slippery, since projects can contain sub-projects and can themselves be more or less complex. For this report, the key distinction is **purpose**: programmes address social goals; projects address technical ones directly and achieve social goals only indirectly or in combination with other projects.

Figure 8 shows the hierarchy of explicit or implicit performance contracts between the different levels of the policy system in countries that follow the principles of the new public management. In former times, many countries did not make clear distinctions between these levels and as a result governance was unclear. Performance contracts involve explicit allocation of tasks and goals, improving clarity. In most places, evaluation follows a 'waterfall principle' where actors at each level evaluate actions at the level below them. This tends to make policy an evaluation-free zone. A key goal of the CNIC in establishing a system of evaluation is to bring policy – in the form of its own strategy – into the sphere of evaluation.

Figure 8 Hierarchy of Performance Contracts and Evaluation



3.3 Evaluation Models

Evaluation models are conceptually distinct ways of thinking about, designing and conducting evaluation efforts. We find Hansen's typology useful[20], which shows the range of evaluation models available. She lists

- Result models, which explore goal attainment and try to identify all intended and unintended effects
- Explanatory process models, which focus on implementation processes
- Systems models, which try to assess how inputs, outputs, structures and outcomes change
- Economic models – cost-efficiency, cost-effectiveness and cost-benefit
- Actor models, which focus on the perspectives of those involved in an intervention, such as its beneficiaries
- Programme theory models, which try to (re)construct the logic of the intervention (logic model) and use this as a hypothesis about expected effects, which the evaluation then tests

Most R&D evaluations have in the past used (implicit) results models: essentially trying to understand what happened as a result of interventions and then connecting this back to programme goals. This can lad to an unfocused search for possible effects to explore but has the advantage that unexpected effects may well be discovered. There is a growing interest in programme theory models because these provide an explicit overlap with programme design by using the intervention logic as an hypothesis about possible effects. This has the corresponding risk that by focusing on the intended effects, the unintended ones may not be identified. In this report we use an Input-Output-Outcome-Impact model (discussed in more detail below) that follows intervention logic and is therefore a variety of programme theory model.

[20] Hanne Foss Hansen (2005), "Choosing evaluation models: A discussion on evaluation design," *Evaluation*, Vol 11 No 4, pp 447-462

3.4 Why Evaluate?

We evaluate in order to

- Provide accountability – essentially explaining to the taxpayer what happened as a result of tax money being used and providing information to support performance contracting
- Generate evidence about the effectiveness of a policy intervention that can provide information about future decisions
- Learn about how to improve the way current and future interventions are designed and managed

In effect, there are two main types of evaluation, which can be used depending on the purpose. These are

- Formative evaluation
- Summative evaluation

Formative evaluation asks how, why, and under what conditions a policy intervention works, or fails to work? In essence, formative evaluations are geared towards learning and programme or policy improvement. Formative evaluations are important for assisting the effective implementation and delivery of policies, programmes or projects. Hence, often mid-term evaluations are intended to be formative.

Summative evaluation (sometimes called impact evaluation) asks questions about the impact of a policy, programme or intervention on specific outcomes and for different groups of people. It looks back at achievements and is aimed at accountability.

The distinction between summative and formative evaluations is not always as rigid as the above characterisation might suggest. The following table (Figure 9) gives a quick comparison of formative and summative evaluation.

Figure 9 Comparison of formative and summative evaluation

	Formative	Summative
Target audience	Programme managers/practitioners	Policy-makers, funders, the public
Focus of data collection	Clarification of goals, nature of implementation, identifying outcomes	Implementation issues, outcome measures
Role of evaluation	Interactive	Independent
Methodology	Quantitative (hard data) and qualitative (soft data) – emphasis on the latter	Emphasis on quantitative data
Frequency of data collection	Continuous monitoring	Limited
Reporting procedures	May be informal via discussion groups and meetings	Formal reports
Frequency of reporting	Throughout study	On completion of evaluation

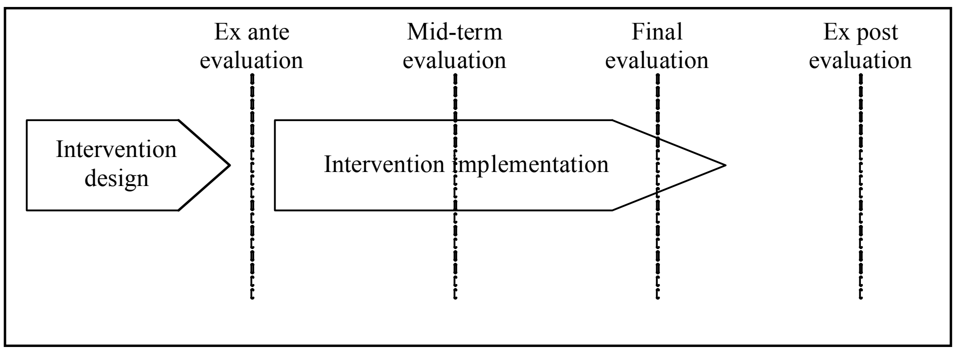
Adapted from Joan Herman, Lynn Lyons Morris and Carol Taylor Fitz-Gibbons (1987), *Evaluator's Handbook*, London: Sage, p. 26

Evaluation has different uses at different parts of the policy or programming cycle (see Figure 10).

- Ex-ante evaluations: In our sense these are not strictly evaluations but planning studies, which occur at the beginning of a programme, in the planning or developmental stages. It is at this stage that the logic and rationale for the programme are considered, indicators developed and baseline data collected. Robust planning at this stage will make all subsequent stages of the evaluation easier. This type of study can have two forms: one, where it aims to diagnose problems and propose and plan an intervention to solve the; another where such a plan has already been prepared and the 'ex ante evaluation' is in fact a design review, aiming to test the robustness of an intended intervention through systematic criticism.

- Mid-term evaluations: These types of evaluations occur at some point during programme implementation and are designed to measure and report on performance to date. They allow for adjustments and refinements of programme operational procedures and can help to identify whether the original objectives are still relevant. They tend to be less focused on outcomes and more focused on processes.
- Ex-post evaluations (or final programme evaluation): These take place at (or towards) the end of an intervention. They are focused more on the outcomes, impacts and results of an intervention. The final evaluation report should largely be based on the accumulation of monitoring data and mid-term evaluation reports. Evaluation that is saved to the end of a programme is evaluation too late to improve and sustain the programme.
- Longitudinal evaluations: These cover all evaluation stages and commonly involve follow-up at some point down-stream once an intervention has finished.

Figure 10 Timing of Evaluations



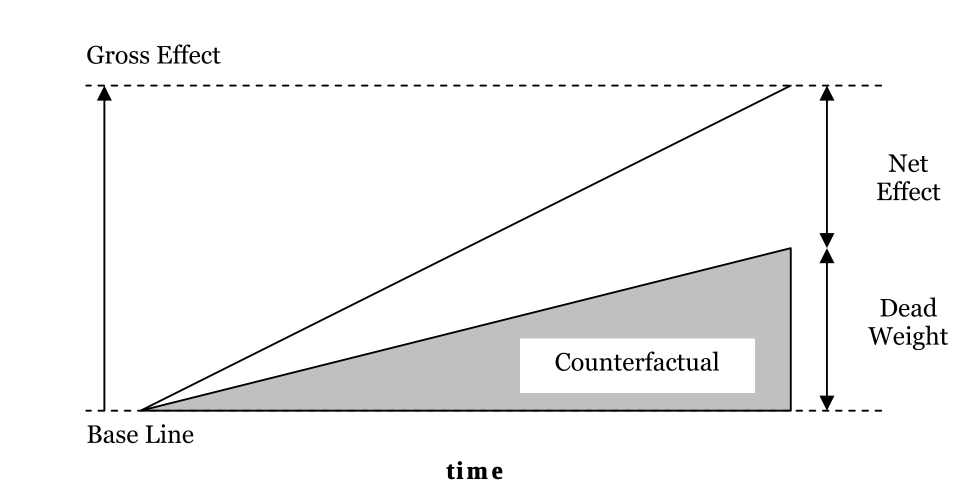
It is in fact very common to evaluate programmes before they finish. This is because of the need to decide whether they should continue. Such a decision normally has to be taken before the programme end, otherwise it will not be possible to reserve a budget line for the continuation and the intervention will be lost in the budgeting process.

3.5 Key Evaluation Concepts

The central problem in evaluation is to connect an intervention with its effects: that is, to establish what it **caused**. Solving this involves identifying its **net effect** and solving the **attribution problem**.

In the ideal case, at the start of an intervention a **baseline** measurement exists or is taken – so that there is an adequate description of the state of affairs that the intervention is intended to change. Over time, the intervention appears to have an effect and if the baseline measurement is repeated the change in the baseline (represented by the arrow at the left of Figure 11) is the **gross effect**. Many evaluations report this as the effect of the intervention. However, in many cases, there are changes in the value of the variable(s) measured by the baseline that happen quite independently of the intervention- for example, an intervention to increase economic growth will be only one of many factors influencing such growth. The **attribution problem** is to decide of much of the change in the variable is caused by the intervention. The change that would have happened anyway is called **deadweight** – it is the change that would have happened in the so-called **counterfactual situation**, ie the situation without the intervention. The gross effect minus the deadweight is the **net effect** that we can attribute to the intervention[21]. **Displacement** or **substitution** effects can also complicate the measurement, where the intervention causes activity to move between categories rather than to increase.

Figure 11 Key Evaluation Concepts



While these concepts are simple in theory, they are very difficult capture in practice – especially in complex systems like national research and innovation systems where many different causes often have to come together to effect change.

3.6 Limits to Evaluation

In the following Chapters we explore a range of different evaluation methods. It is important to recognise that while they have strengths they also have weaknesses. **No single technique is on its own sufficient to obtain a robust evaluation result.** R&D evaluators therefore tend to take care to use several methods in combination and to look for convergence among the results they provide. It is especially important to take care with economic methods, whose quantitative nature makes the look 'hard' but which are often based on questionable assumptions and unreliable data.

Limitations of technique mean that we do not quite know what we are measuring or that an instrument only captures one perspective of the phenomenon under study. For example, we can ask industrial beneficiaries of R&D programmes for estimates of the cash flow benefits they obtained, but it is rarely clear (except in cases of failure) how good their estimates are or how much of the total benefits they represent. Many of the available techniques allow us to demonstrate with a fair degree of confidence that there **are** effects, and sometimes to say that these effects may be quite large in comparison with the state's investments in the intervention being evaluated.

However, naturally enough, a key concern of many policymakers is with the **relative** effects of different types of intervention, since they want to optimise the allocation of scarce resources. Our study for the Dutch Ministry of Economic Affairs on methods

[21] Deadweight can also be negative, as in the case of an intervention intended to increase economic growth that is launched in a period of economic decline. In such a case, the net effect could be a reduction in the rate of decline rather than having a positive slope

and practices for relative performance measurement of innovation programmes[22] identified a number of problems with trying to do this. Many of these problems arise from the fact that, if you do not know what proportion of the total effects you are measuring when looking at two different interventions, there is no logical basis for saying that one is bigger than the other.

There are a number of further problems associated with R&D evaluations, which make R&D evaluations particularly difficult. These are

- The 'project fallacy' – the fact that what is administratively defined and funded as a 'project' may bear little relation to the beneficiary's 'real' project or agenda, complicating any attempt to understand and disentangle the effects of the administratively defined project
- 'Skew' – the fact that commercially successful R&D projects normally only make up a very small proportion of any portfolio and the fact that the majority of subsidy may therefore be 'wasted'. (As with buying the winning lottery ticket, however, techniques for predicting which projects will succeed are not well developed.)

An important sub-category of R&D evaluations are impact assessments – evaluations that are concerned with the effectiveness of a policy measure. Typical problems associated with (economic) impact assessment are

- The time lag between intervention and effects differs from programme to programme. There is no single moment in time when the ultimate effects can be compared between more than one instrument
- Given the attribution problem, there is a risk of double counting the effects of various policy measures used simultaneously by companies. The more instruments are included in the comparison the more difficult it will be to define control groups (and in some countries, such as Norway, it can be coherently argued that there are so few companies not involved in the state's R&D subsidy schemes that the idea of a 'control group' is incoherent)
- The variation in risk levels of different policy instruments where some low risk programmes have many incremental and short term effects whereas high risk programmes have fewer but potentially more radical effects in the longer term. Comparing the two in the medium term would always favour the low risk programme and therefore lead to a certain risk averseness of public action, whereas the 'market failure' justification assumes that government acts when risks are too high for the private sector
- The failure to quantify the 'softer' effects that governments want to achieve in changing the behaviour of the target groups in the current cost benefit analyses
- Possible changes in the context of the 'problem' that a government action wants to address. Even though an instrument performed perfectly well in terms of cost-benefit analysis, it could be that the context of the firms and the innovation system has changed drastically in the meantime, making the same instrument ineffective in the future. This would argue against using ex-post cost benefit analysis as the sole input for policy decisions
- Effectiveness measurements typically measure only some of the effects of programmes. In particular, cost-benefit analyses tend to focus on the private returns to intervention rather than the social returns or 'externalities,' which

[22] Patries Boekholt, Maureen Lankhuizen, Erik Arnold, John Clark, Jari Kuusisto, Bas de Laat, Paul Simmonds, Susan Cozzens, Gordon Kingsley and Ron Johnston, *An international review of methods to measure relative effectiveness of technology policy instruments*, report to EZ, Technopolis: Amsterdam, 2001

normally justified the intervention in the first place, but which are harder to measure. Because the ratio between the measured and unmeasured effects of individual programmes is not known, different cost-benefit measures are typically incommensurable and cannot validly be compared

- Policy instruments should be seen in the context of their role in the innovation system and the specific objectives and target groups they address. 'Relative effectiveness' in economic terms assumes that interventions are mutually substitutable and can lead to allocation of resources away from vital innovation system functions, such as postgraduate education. The idea that a modern economy can run without postgraduate technologists is about as plausible as the notion that you can make your car go faster by buying a bigger engine and throwing away the wheels

3.7 I-O-O-I – A Generic Model for R&D Evaluation

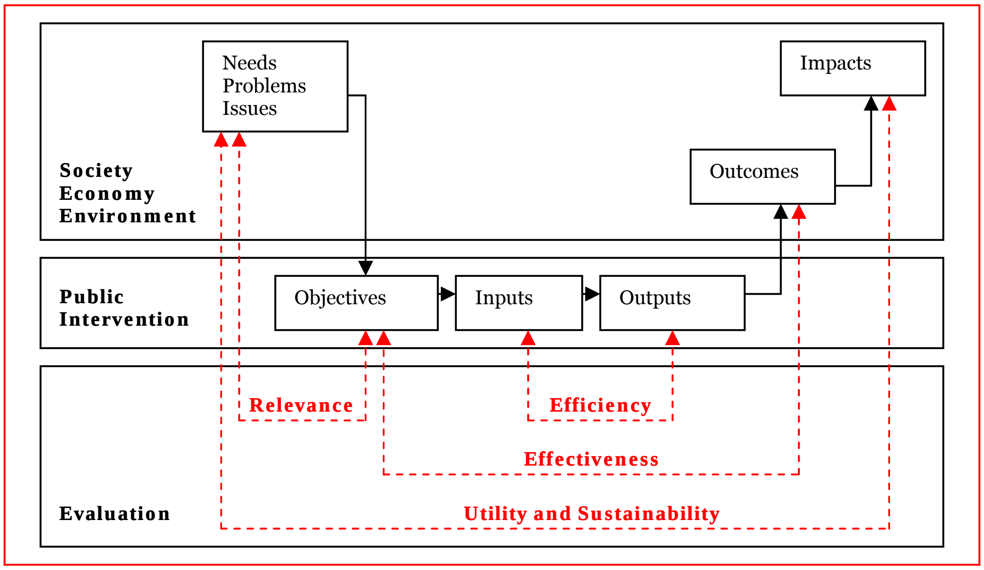
As discussed in the section on LFA above, intervention logics tend to have a common structure. The I-O-O-I model shown in Figure 12 boils this down to its simplest form. The generic intervention logic is made up of the following steps

- An analysis of **problems, needs or issues** that need intervention, in that markets and other normal social processes will not correct them
- This analysis implies a set of **objectives** – essentially to fix the problems
- An intervention therefore provides **inputs** – typically money and other resources, normally in the world of research and innovation in the form of project funding
- These enable activities that are expected to lead to **outputs** – direct results of the work enabled by the inputs, which can normally be specified in a project contract
- The outputs enable wider effects to be created. Usually, however, in R&D and innovation funding, these **outcomes** primarily affect the beneficiaries of the projects. In this sense, wider society has not yet received a payback on its investment
- The outcomes enable wider social **impacts**. For example, the increased competitiveness of participants in the intervention may flow through into increased GDP and national wealth

Evaluators sometimes treat outputs, outcomes and impacts as first, second and third order effects. While this a useful approach in that it emphasises the chain-link logic' that leads from the intervention to the eventual social effects, it is misleading in that the number of links in the chain from intervention to social effects can vary. From the policy perspective, the number pf links (or 'orders') is not important. What matters is that the intended effects start with outputs that belong to, or are close to, the beneficiaries and that the outcomes that these enable tend also so be connected to the beneficiaries. In economic terms, these are the result of redistributing social resources to the beneficiaries. Society gets a payback on this investment only when the results spill over to society itself. It is often useful, therefore, to distinguish between outcomes that relate to beneficiaries and impacts that provide spillovers or externalities to society.

It is comparatively straightforward to articulate these questions in the form of a logic model, such as a logical framework. That involves choosing to use a 'programme theory model' for evaluation, focusing attention on mapping and trying to identify the intended effects of the intervention. This is useful in the context of public management, making the evaluation a servant of the public management system. But where evaluations have other purposes, other evaluation models should also be considered.

Figure 12 The Inputs-Outputs-Outcomes-Impacts Model



As Figure 12 suggests, the I-O-O-I relationships give rise to a number of generic evaluation issues

- **Relevance**. The issue of relevance consists of examining whether the objectives of an activity correspond with the needs, problems and issues it is intended to address
- **Effectiveness**. The issue of effectiveness is especially pertinent in the context of mid-term and ex post evaluation. It consists of asking whether results and impacts generated by the activities supported meet the objectives
- **Efficiency**. The issue of efficiency consists of examining the level of resource use (inputs) required to produce outputs and generate effects. In other words, optimisation of resource utilisation is concerned. An activity that is assessed as having an effect, may not necessarily be efficient: the same effect could have been reached with less resource
- **Utility**. The issue of utility consists of looking for expected and unexpected effects (i.e. those that were respectively identified and not identified at the design phase as objectives) and whether these, when they are positive, correspond with needs, problems and issues of different groups in society and the economy
- **Sustainability**. The issue of sustainability consists of examining whether the positive impacts on critical clients and beyond would continue into the future, even after the ending of an activity

A less precise but simpler way of posing these questions is to ask

- Are we doing the right thing?
- Are we doing it well?
- What happens as a result?
- What should we do next?

These questions go to the heart of what we are trying to do in evaluating public policy

3.8 Summary of Evaluation Methods

Figure 13 gives an overview of the strengths and weaknesses of the evaluation methods discussed in the previous sections, while Figure 14 discusses their general applicability. We discuss these methods in some detail in the Appendix.

Figure 13 Overview of evaluation methods

Method	Strengths	Weaknesses
Macro-Economic modelling	• Based on established principles • Susceptible to rigorous testing • Provide consistent accounts of the importance of R&D for the macroeconomy	• Dependent on large quantities of high-quality data • Resource-intensive, skill-intensive and time consuming • Difficult for non-specialists to understand
Micro-economic modelling	• Provide a structured, formal framework for project/programme evaluation • Susceptible to rigorous testing	• Dependent on large quantities of high-quality data • Resource-intensive, skill-intensive and time consuming • Difficult for non-specialists to understand • Risk that important externalities will be excluded
Cost-benefit analysis	• Offers a systematic and comprehensive evaluation procedure • Assumptions can be transparently presented, little specialist knowledge required for understanding	• Quantification of some costs and benefits can be very difficult, and 'guesses' can yield misleading results • CBA models are often very sensitive to realistic changes in the assumptions used
Spillover analysis	• The idea of spillover is central to most justifications for state intervention • Spillover studies have demonstrated the huge social value of R&D • They have also helped identify the mechanisms of spillover, allowing policy to focus in improving these	• Ambiguous definition of 'spillover variable' • Frequent need to resort to proxies • Some studies have been found to produce results that are unstable
Survey of beneficiaries	• Economical • Provides understandable and credible results • Can accommodate comparison groups and counterfactuals • Can be representative • Supplies high numbers necessary for statistical analysis	• Subjective • Positive bias • Does not capture richness of individual projects and experiences
Peer review	• Low-cost, fast-to-apply • Well-known, widely accepted • Versatile	• Peers' independence • Problematic when highly innovative or interdisciplinary work is to be assessed
Panel review	• 'Method of last resort' when budget, time and methods constraint are presents	• Selection of suitable peers/composition of panels
Case studies	• Comprehensive view of phenomenon under study/richness of detail • Can be used for theory-building • Easy for policy makers to read and understand	• Qualitative evidence generally considered less persuasive than quantitative evidence • Not representative, not generalisable to population

Method	Strengths	Weaknesses
Historical tracing	• Produces interesting and credible study • Sheds lights on process dynamics • Suited for the tracing of social impacts	• Attribution and establishment of cause and effect difficult • Time-consuming and costly
Network analysis	• Different and complementary perspective • Can bring surprising results • High comparability	• Theoretical paucity • Largely descriptive • Data collection time-consuming
Prospective studies (foresight and technology assessments)	• Broadens scope of evaluation studies	• Time-consuming and costly • Attribution and establishment of cause and effect difficult
Bibliometrics and patent analysis	• Objective, credible • Applicable to variety of evaluation topics • Easily scalable • Results relatively easy to understand • Places no burden on people being evaluated	• Narrow view of outputs, limited to publications and patents • Narrow view of impacts, limited to citations • Publishing and citation patterns differ across fields • Weak in the social sciences and humanities • English language bias
Webometrics	• More timely than bibliometrics • Covers wide range of scholarly produced artefacts • Free access to web data, affordable	• Web not quality-controlled, hence results indicative rather than robust evidence • Difficult to separate out different types of publications/ sources of different value processed in webometric results • Web-data incomplete and arbitrary/ majority of academic articles not freely available online
Meta-evaluation	• Inexpensive, comprehensive • Increases the power of individual study • Can guide an improve future evaluations	• Limited to answering questions already addressed by existing studies • Hampered by quality of existing studies • Only as current as existing studies
Organisational review	• Useful tool for exploring the causes of good or bad performance in a=management and administration • Provides opportunities to study organisational options for performance improvement	• No basis in science or evidence-based study • Highly dependent on the evaluator's experience and understanding of the specific context of research and innovation funding
Benchmarking	• Useful tool for identifying performance standards where these cannot be deduced from first principles • Can trigger organisational learning and process improvement	• Poor availability of data in the field of research and innovation funding
History	• Identifies causes and behaviour drivers not necessarily obvious from a static analysis • Makes 'sense' of performance in ways that make sense to evaluators and evaluation readers	• Requires respect for historical methods and standards of evidence • Sources can be had to find

Figure 14 General relevance of evaluation techniques

Method	Ex ante	Process	Ex post		
			Outputs	Outcomes	Impacts
Macroeconomic modelling	Y				Y
Microeconomic modelling			Y	Y	
Cost-benefit analysis	Y			Y	Y
Statistical approaches			Y	Y	Y
Spillovers				Y	Y
Survey of beneficiaries	-	Y	Y	Y	-
Peer review	Y	Y	Y	-	-
Panel review	Y	Y	(Y)	(Y)	(Y)
Case studies	-	Y	Y	Y	Y
Historical tracing	-		Y	Y	Y
Network analysis	-		Y	(Y)	(Y)
Prospective studies (foresight and technology assessments)	Y	(Y)	-	-	-
Bibliometrics and patent analysis	Y		Y	-	(Y)
Webometrics	Y		Y	-	(Y)
Meta-evaluation	-	(Y)	Y	Y	Y
Organisational review		Y			
Benchmarking	Y	Y	Y		
History		Y			

3.9 Authors' Views and Assessment of Methods

The repertoire of techniques used in R&D evaluations has been evolving over the past 20-30 years. Scientometrics and social network analysis are important extensions to the repertoire. It is increasingly argued[23] that at least some of the techniques routinely used and the associated styles of evaluation have reached diminishing returns in terms of generating new knowledge. For example, repeated surveys of beneficiaries in pre-competitive, collaborative R&D programmes produce similar results.[24]

Evaluating R&D policies and interventions is complex – especially when objectives are pitched at the level of changing important characteristics of research and innovation systems (as is the case for the EU R&D Framework Programmes). The overall repertoire of indicators and evaluation methods used needs to be applicable at the systems level as well as informing us about lower-level phenomena. Indeed, there appear to be especially large opportunities to improve methods for evaluating effects at the meso-level of policies and systemic bottlenecks and at the macro level of the whole systems. More specifically, areas of potential innovation and improvement include

- Further articulation of social network analysis, with better linkage to research and innovation processes. At present, social network analysis in R&D – while interesting – is mainly descriptive. Linking meso- and macro-network topologies

[23] Luke Georghiou (1997), 'Issues in evaluation practices in OECD countries', in OECD, 1997; Erik Arnold (2004), 'Evaluating research and innovation policy: a systems world needs systems evaluations,' *Research Evaluation*, Volume 13 Number 1, April

[24] Erik Arnold, John Clark and Alessandro Muscio (2005), 'What the evaluation record tells us about Framework Programme performance', *Science and Public Policy*, Vol 32, No 5, pp385-397

to micro-level behaviour and strategy is needed, if social network analysis is to become more explanatory and potentially predictive

- Better connection between macro- and micro-economic modelling and simulation, feeding back results from micro models to the macro level
- Longitudinal studies of institutions, 'knowledge value collectives'[25] actors and actor groups, as distinct from interventions, in order better to understand the implications of the 'project fallacy', to understand the choices faced by intended beneficiaries and how they behave in relation to programmes and policies
- Techniques of meta-evaluation and systematic review appear to be an area where genuine innovation is needed in R&D evaluation practice, rather than one where a great deal of 'good practice' exists to be imitated
- Methods to trace social, as well as economic, effects of policies and interventions. The Finnish VINDI project as well as the 'tracking-back case studies' we conducted in the humanities could show a way

Against the background of the discussion conducted above, we can formulate the following general recommendations:

- Meta-evaluation and other systematic reviews of evidence should be used to provide inputs to policymaking, especially at systems level.
- Since there is significant scope for policies and programmes to have unexpected outcomes, evaluation models should find a balance between focusing only on testing the programme logic set out by the policy or programme designers, on the one hand, and a results-oriented approach that tries to uncover all possible outcomes.
- Some methods are more appropriate at certain stages of the policy or programme life cycle than at others, so methods should be selected in part for their appropriateness. For example, cost-benefit analysis is a very useful prospective technique for thinking in a structured way about the potential impact of an intervention, but is difficult to operationalise ex post in order to obtain a reliable estimate of actual impacts. These difficulties also underlie problems in making comparative estimates of the effects of different interventions ex post. On the other hand, broad economic impact estimates – while partial – can be useful in explaining the effects of policies and interventions.
- Panel reviews – where respected people are asked to make judgements about things in which they are not necessarily expert – need to be carefully managed where they cannot be avoided, as
  - There is a difficulty of assembling panels that exclude people from major beneficiary organisations
  - They are liable to being exploited to promote disconnected political or policy agendas.
  - In contrast, peer review by scientists of scientific aspects of programmes and policies still has an important role to play in both ex ante and ex post evaluation.
- More generally, R&D evaluation methods in use are often individually not fully reliable, and should generally be used in combination so that the evaluator can 'triangulate' between different methods and look for convergence

[25] Barry Bozeman and Juan Rogers (2002), 'A churn model of scientific knowledge: Internet researchers as a knowledge value collective,' *Research Policy*, Vol 31, , pp 769 - 794

- R&D funders should aim to incorporate opportunities for methodological experimentation and innovation in their evaluation strategies and allow for experimental funding in order to improve techniques because the opportunities for experimentation will be limited and the rate of innovation will continue to be slow, as long as evaluation budgets remain tied closely to the performance of 'operational' evaluations

4. Evaluation at Three Levels in Chile

This chapter sets out in broad terms how to apply the I-O-O-I model at the three levels of programmes, organisations (agencies) and the strategy for innovation as a whole. It discusses the strengths and weaknesses of current evaluation practice in Chile and of the national state of the art in collecting and using relevant indicators.

4.1 The I-O-O-I Model at Three Levels

4.1.1 The I-O-O-I Model at the Levels of Strategy, Organisations and Programmes

The I-O-O-I model gives us a coherent way to think about what to evaluate at the different levels of the CNIC's strategy, the implementing agencies and their programmes.

Figure 15 provides a general summary. It distinguished between the rationales for action at each level. The strategy is concerned with improving the performance of the innovation system as a whole. There is a division of labour, whereby agencies take on the task of improving performance of sub-systems – for example, CORFO works with the industrial sub-system – while within each agency there is a further division of labour so that individual programmes tackle components of the sub-systems. The red arrows represent this. In principle it is possible to produce a (large and complex) set of nested logical frameworks that articulate this division of labour.

At the bottom level, programmes put resources (mostly, but not only, money) into projects that are intended to generate both private and public knowledge goods. These outputs should in turn improve the performance of the sub-groups addressed by the programme and benefits are expected to spill over to impacts on the wider society.

The organisations or agencies work at a higher level of aggregation. Their unit of activity is the programme, whose intended output is improved performance by a specific sub-group of beneficiaries. This should lead to better sub-system performance that contributes to the systemic performance of the NIS.

At the level of the CNIC strategy, the activities are reports, monitoring and influence whose intended outputs are increased agency ability to improve the performance of the large sub-systems with which they deal. These should cause better systemic performance of the NIS that leads to the greater overall welfare reflected in the CNIC's overall objective of redoubling Chile's GDP per head.

The black arrows are intended to indicate two things. One is that the outcomes and impacts at lower levels contribute to reaching those intended at higher levels. The other is that the 'nested' division of labour means that the outcomes at the bottom level become the outputs at the middle level.

It follows that the generic evaluation questions of Figure 12 can be articulated differently for each of the three levels of intervention: strategy; organisation / agency; and programme (Figure 16). This involves applying a programme theory model, which is inherently blind to unintended effects. In practice, therefore, questions about unintended and perverse effects need also to be included with the issues of effectiveness and efficiency when designing evaluations. A second dander with using a programme theory model is that it is implicitly static. Needs can change during the course of interventions, so in testing utility it is also important to try to understand whether the problems addressed by the intervention have changed in nature.

Figure 15 The I-O-O-R Model Across Three Levels

Objectives	Inputs	Activities	Outputs	Outcomes	Impacts
Strategy Improve system performance by increasing and balancing sub-system performance	Strategic intelligence Governance/steering	Studies, monitoring, evaluation Influence on budget	Innovation strategy Other policy documents	Improved performance by agencies Improved performance by NIS sub-systems	Improved NIS systemic performance Contributions to overall welfare, quality of life
Organisation Improve performance of sub-system(s)	Strategic intelligence Management Budget	Programmes	Improved performance by specific beneficiary sub-groups	Improved sub-system performance	Contributions to improved NIS systemic performance
Programme Improve performance of sub-system components	Programme management Money	Projects	Knowledge • For beneficiaries • Public goods	Improved performance by specific beneficiary sub-groups	Contributions to improved NIS sub-system performance

Note: NIS means national innovation system in the sense of all the components that are involved in innovation processes – not just the state organisations but also industry, education, banking and finance systems, libraries, patent systems etc etc

Figure 16 Generic Evaluation Issues at Three Levels

	Relevance	Efficiency	Effectiveness	Utility	Sustainability
Strategy	Do the objectives of the strategy reflect national needs?	Has the strategy been developed in an efficient way and with high quality?	Is the strategy deployed efficiently and effectively? Are sub-goals being achieved?	Have increases in the performance of the NIS satisfied the national needs originally identified?	Are improvements in the performance of the NIS based on structural changes so that they are likely to be permanent?
Organisation	Do the objectives of the organisation correspond to the needs of the sector with which it deals?	Does the organisation design and implement programmes that work in efficient ways? Does it spend the right amount in administration?	Do programmes reach their goals and increase sub-system performance? What is the overall effect of the agency, over and above programme goals?	What are the effects of the organisation on the overall performance of the sub-system (sector) that it addresses?	Are improvements structural in nature? Have needs changed?
Programme	Do programme goals match identified needs?	Does the programme meet its objectives in a cost-efficient manner?	To what extent does the programme meet its goals, especially in relation to beneficiaries?	Does the programme solve the problem it was intended to address?	Is this a permanent solution? Have needs changed?

4.2 The State of the Art in Evaluating Research and Innovation in Chile

4.2.1 Organisations and Systems Involved in Evaluation

The key organisations that commission evaluations of research and innovation in Chile today are DIPRES, CONICYT and CORFO.

DIPRES, the Budget Office of the Ministry of finance, has been a driving force for the increased use of evaluation in Chile during the past decade that has accompanied the growing use of the principles of the New Public Management. It introduced the principle of routine evaluation of all government programmes in 1997, according to a timetable to be agreed yearly with the interdepartmental committee responsible for evaluation that comprises representatives of the Presidential cabinet and the Ministries of Finance and Planning and Cooperation. These evaluations of government programmes (EPG) use logical framework analysis as their central tool to evaluate the consistency between objectives, the way programmes are designed and implemented and, where relevant, their results. They therefore tackle

- Design
- Organisation and management
- Effectiveness and quality
- Efficiency and economy (financial performance)
- Justification

The main emphasis of the EPG evaluations is on quality assuring programme designs. These can be desk-based exercises or involve some degree of fieldwork. A panel of three external experts, supported by DIPRES, performs each.

DIPRES introduced Impact Evaluations (EI) in 2001. These are intended to complement the logical framework focused EPG evaluations by, as their name suggests, evaluating programme impacts and cover

- Effectiveness of the programme (results at short, medium and long term)
- Use of resources (economy and efficiency)
- Global performance (programme design and management of internal processes)

Spending reviews or evaluations of organisations ('Institutional Evaluations' – evaluación comprehensiva del gasto – ECG) were introduced in 2002. They explore

- Consistency between mission, strategic objectives, offered goods and services and beneficiaries.
- Use and result of resources (performance indicators, effectiveness, quality, efficiency and economy)
- Institutional capacity, mechanisms or management processes

Both the impact Evaluations and the Institutional Evaluations are done by external contractors – academics or consultants.

All DIPRES evaluations are reported to the National Congress and are published online. This is an important good practice, not only because it enables citizens to understand how their tax money is being used but also because it provides a source of learning for policymakers and others and because it exposes the evaluations to the potential for external criticism, thereby tending to keep quality high.

A second important driver for evaluation has been funding from international agencies, especially the Inter-American Development Bank and the World Bank, both of which require evaluation of the interventions they fund. These evaluations are reported to the funders and do not necessarily feed back into the Chilean system.

A third, new driver is an ongoing World Bank funded project to develop evaluation capacity and systems in CORFO, CONICYT and the Ministry of Economy.

CORFO has been commissioning impact and other kinds of evaluations since before the new drive to evaluation by DIPRES and undertook a total of 23 programme evaluations in the period 1997-2006, all of them producing positive conclusions about the usefulness of the programmes. However, the number of evaluations has risen dramatically in the past few years. CORFO started work on an internal Evaluation and Monitoring System (SEM) project in 2006, with sub-systems for design, operation and impact evaluation. CORFO is also engaged in the World Bank evaluation project.

Like other research councils, CONICYT has long-standing processes for appraisal of proposals and projects are also subject to technical evaluations on completion. However, impact evaluation is at an early stage. Some studies have been done (eg FONDAP and FONDEF) but there is no overall impact evaluation framework or policy. This is likely to change once the World Bank evaluation project is complete.

4.2.2 Methods

We have reviewed a sample of ten evaluations of innovation and research carried out in recent years. These include two DIPRES EPG logical framework studies, two DIPRES EI impact studies, three studies carried out for the Inter-American Development Bank, two carried out for CORFO and one for FIA. They are listed, together with a summary of the methods they use, in Figure 17. The last row in the Figure indicates whether the evaluation is significantly formative, which is to say that it devotes a lot of attention to learning and therefore process improvement.

In general, the evaluations could have been presented in more useful ways. Tables of contents were either missing or buried some tens of pages into the report. There was rarely a statement of the methods used (or any discussion of their limitations). They were sometimes very long. In most cases, summaries were also very long, making it difficult for the reader to absorb the key messages and making it rather unlikely that policymakers would actually read the summaries.

There is one clear evaluation tradition, primarily pursued in teams led by Jose Miguel Benavente, which focuses on quantitative (normally econometric) investigations using control groups. These are applied to rather simple situations, where the beneficiary is a (probably small) firm or a person, and focus on establishing statistical differences between treatment and non-treatment groups. They focus on outputs and outcomes, ie essentially changes associated with the beneficiaries. They tend not to be formative but summative: that is, they tell us whether an intervention works but not much about how to improve it.

The use of logical frameworks is focused on DIPRES EPG evaluations, which are essentially about quality controlling programme designs. More widely, the formative evaluations tend to use simpler surveys and more qualitative methods, focused on obtaining experience about processes.

4.2.3 Levels

Evaluations in the Chilean research and innovation system to date have almost entirely been focused at the programme level. The first substantive evaluations of research and innovation funding organisations came onto the agenda only in 2009.

The situation is very similar abroad. Less than a handful of evaluations have been done internationally of research or innovation funding organisations. While countries produce endless studies and analyses diagnosing national competitiveness in fields that include research and innovation, there is no established practice of evaluating strategies. In this sense, Chile is already at the frontier.

Figure 17 Methods Used in Ten Evaluations

Subject	PNIN[26]	FONDEF[27]	Regional STRPs[28]	MECESUP[29]	CORFO FDI[30]	Postgrad. Grants[31]	FIA[32]	FONDECYT[33]	FONTEC[34]	PDIT/Chile Innova[35]
Evaluator	Gerens	Addere	Panel	Panel	Benavente	STATCOM	GPI	Benavente	Benavente	GPI
Client	CORFO	CONICYT	DIPRES EPG	DIPRES EPG	DIPRES EI	DIPRES EI	FIA	IADB	IADB	IADB
Control group				X	X			X	X	
Econometric / statistical					X	X			X	
Bibliometric				X				X		
Beneficiary survey		X			X	X	X		X	
Benchmark	X									
Logical framework			X	X						X
Interviews		X	X	X			X			X
Case studies										X
Database analysis				X	X	X	X	X		X
Formative	X	X	X	X		X	X			X

[26] Gerens SA (2006), *Evaluacíon de Incubadores-InnovaChile*, Informe Finale, Providencia-Santiago

[27] ADDERE Consultores Ltda (2006), *Evaluacíon de Casos Exitosos de Proyectos Financiados por FONDEF*, September

[28] Lucy Winchester (2006), Verónica Loewe and Roberto Prado, *Programas Regionales de Investigación Científica y Tecnologica*, Santiago: DPRES

[29] Víctor Salas, Ornella Yacometi and Javier Corvalán (2004), *Informa Final de Evaluación* – MECESUP, Santiago: DIPRES

[30] Facultad De Ciencias Económicas Y Administrativas, Universidad De Chile (2005), *Informe Final De Evaluacion De Impacto Fondo De Desarrollo E Innovación (FDI)*, CORFO: Santiago

[31] STATCOM Estadisticos Consultores (2007), *Evaluación en Profundidad Programa Becas de Postgrado*, Santiago: STATCOM

[32] GPI Consultores (2005), *Evaluación de Impacto Fundación para la Innovación Agraria*, Santiago: GPI Consultores

[33] José Miguel Benavente, Gustavo Crespi and Alessandro Maffioli (2007), *The Impact of the National Research Funds: An Evaluation of the Chilean FONDECYT*, Working Paper OVE/WP-03/07, Washington: IADB

[34] José Miguel Benavente, Gustavo Crespi and Alessandro Mafioli (2007), *Public Support to Firm-Level Innovation: An Evaluation of the FONTEC Programme*, OVE/WP-05/07, Washington: IADB,

[35] GPI Consultores (2004), *Evaluación de Medio Termino del Programa de Desarollo e Innovación Tecnologica*, Santiago: GPI Consultores

In our sample, there is no use of theory as an evaluation criterion – indeed the evaluations are completely atheoretical, making no reference into the research literatures and very rarely using international experience. There is no use of systematic review or meta-evaluation, to draw more general lessons from evaluation experiences. There is no longitudinal study – for example, to understand how interaction with the support system or the knowledge infrastructure affects company development or looking at carers over time. The evaluations with a longer time perspective are impact evaluations, but they nonetheless look at single events rather than interactions over time. There is no use of domain knowledge or policy experience. Overall, the evaluations are aimed firmly at the management of individual instruments rather than contributing to wider policy discussions.

It is difficult (and somewhat unfair) to try to sum up the evaluation work of a whole country in such a way, but the effort to use 'clear', quantitative methods is striking and positive. The limitations of any evaluation method used on its own are also clear from these studies, which would have benefited from using a broader range of methods in order to address more formative questions. The comparative absence of a 'research on research and innovation' community in Chile is strikingly obvious from the evaluations and this undermines their contributions to policy.

4.3 Statistics and Indicators

To date, in terms of coverage, methods of compilation and reliability, Chile's system of S&T and innovation indicators is not yet deemed to be up to the international standards required by the OECD and complied with by its Member countries. Moreover, although Chile is an observer member of the OECD Committee for Scientific and Technological Policy (CSTP) and its Working Party on S&T indicators (NESTI), it does not figure among the non member countries for which a selected number of indicators are regularly published by this organisation.[36] What are the main shortcomings of Chile's system?

4.3.1 Institutional Setting and Coordination

In Chile as in other countries statistics and indicators related to S&T and innovations activities (inputs, outputs and flows of knowledge) are collected by a variety of public institutions, the most prominent ones being policy implementing agencies such as CONICYT, and CORFO, the Ministries of Economy, of Agriculture and of Education, the National Institute of Industrial Property (INAPI), and the National Institute of Statistics (INE). The Central Bank and the Ministry of Finance also compile and process information that can be of use to monitor and map the performance of the innovation system.[37]

Also in Chile, as in many countries with a much longer history of S&T statistics, the collection and compilation of these statistics and indicators have been undertaken over time in an uncoordinated manner by these institutions for their own monitoring needs without explicit or even implicit reference to their contribution to a comprehensive information base aiming at a better understanding of the determinants and performance of the research and innovation systems.

However in most OECD countries an earlier recognition of the importance of the contribution of R&D and innovation to economic growth and, later, the diffusion of

---

[36] The bi-annual publication "Main S&T Indicators" (OECD, 2008) includes a special section on non member countries that compiles 68 indicators for all countries that are observer to the OECD Committte for S&T Policy except Chile and Brazil (i.e. China, Israel, Russia, Slovenia and South Africa), as well as for selected other non member countries (Argentina, Romania, Singapore, and Chinese Taipei). It is to be noted however that all these countries do not provide data for all 68 indicators. The area less well covered is that related to human resources in S&T.

[37] Such as input/output tables that can be used to map inter-sectoral flows of technology embodied in capital goods or tax information related to R&D fiscal incentives.

the innovation system conceptual framework into policy making has facilitated the emergence of a more comprehensive and integrative view of the S&T and innovation quantitative (and qualitative) information base. In this regard the OECD NESTI Group of national experts and Eurostat have played key roles. Whether in centralised or decentralised information systems, efforts have been made to ensure to the extent possible an homogeneity of definitions and measurement practices as well as a compatibility – or matching possibilities – between databases or statistical information based on different measurement units or types of coverage of surveyed populations (e.g. administrative registries, censuses or surveys). To a large extent this has facilitated not only the development of more robust information bases at the national level, but also the international comparability of S&T statistics and indicators.

In Chile, this coordination process is just beginning and the steering of the S&T statistics and indicators system is still incipient. As highlighted above, it is to be hoped that, as recommended by the OECD and the World Bank, a specific institution be entrusted with an oversight responsibility over the consolidation of the system, ensuring a maximum compatibility and complementarity between the information bases developed by various agencies, as well as, if need be, their compliance with international comparability requirements. In this regard it is highly recommended that the responsible institution be supported by a Steering Committee on S&T statistics and indicators composed of representatives of the main agencies or ministries undertaking data collection in this area. The CNIC would obviously have to be represented to this Committee.

4.3.2 Main Gaps in Chilean S&T Statistics and Indicators

In the compilation of the main S&T statistics and indicators such as those collected in the OECD regular publications,[38] Chile has progressively achieved a good level of compliance with measurement methodologies defined in Frascati and Oslo manuals[39]. However there are a number of important gaps that need to be overcome.

4.3.2.1 R&D Expenditures.

The main sources of information on these expenditures are CONICYT surveys and administrative data sources for R&D performed in the higher education sector (HERD), the Innovation Survey periodically conducted by the Ministry of Economy in collaboration with the National Statistical Institute for business enterprises' R&D (BERD) and from budgetary data for R&D performed in public research institutions (other than universities) and public technological institutes (GOVERD). Among the various aspects that require improvement one can highlight the following

- An improved compliance of CONICYT with the survey methodologies detailed in the Frascati manual so as to deal with the information distortions coming from administrative date about the resources devoted to research activities; the CONICYT survey which already covers private non profit organisations should also include all public research institutions[40]
- The measurement of business R&D through the innovation survey introduces biases and reduces the scope of collected information as compared with dedicated R&D surveys conducted in the majority of OECD countries. It is in particular likely that in Chile the magnitude of BERD be overestimated by the innovation survey due to the use of sampling rather than a census approach, industry coverage

[38] Such as MSTI; see OECD (2008), *Main Science and Technology Indicators, Volume 2008/2*, Paris

[39] See OECD (2002), *Frascati Manual: Proposed Standard Practices for Surveys on Research and Development*, Paris and OECD/Eurostat (2005), *Oslo Manual for Collecting and Interpreting Innovation Data*, 3rd edition, Paris and Luxembourg

[40] CONICYT has already taken initiatives to improve the quality and coverage of its surveys, which should fully comply with the Frascati methodology and include categories related to the fields of science.

limitations and positive biases introduced by the nature of innovation survey respondents. Business R&D surveys should better cover service activities and provide information by size of enterprises in line with criteria that facilitate international comparisons [41]

- The sources of funding of R&D activities performed by diverse types of public and private institutions constitute an important characteristic of a country's S&T and innovation system. In the case of Chile the methods of collecting information about R&D performance and financing are making difficult to construct reliable matrices of sources of funds vs. institutions of performance[42]

4.3.2.2 Human Resources in S&T (HRST)

This is an area that is not satisfactorily covered in Chilean statistics at present. The main sources of information are labour surveys, CONICYT registries and surveys for human resources in public research institutions, and the innovation survey. Discrepancies in definitions and other biases account for the weak reliability of the indicators that can track the evolution of human capital in S&T, its distribution among various types of institutions, sectors of the economy and scientific disciplines. Moreover, to date there are no reliable systematic way to track the mobility of HRST among different institutions and to follow the career paths of highly skilled personnel (such as Ph.D. holders). Among the areas that are in need of improvement are the following

- More robust methods to account for full time equivalent (FTE) for higher education researchers. The estimation of the number of full time researchers employed in the public sector is deemed to be unreliable.[43] Disaggregation by field of research and/or discipline and by gender is either incomplete or just incipient in the new surveys contemplated by CONICYT to be undertaken in 2009. These surveys should also cover non-academic public research institutions and technological institutes
- The information on business HRST is essentially provided by innovation surveys. This information suffers from the same sampling biases as those indicated above concerning R&D expenditures, which would point to a possible overestimation of business researchers, but this interpretation has to be nuanced due to a definition problem: in the innovation survey the measured unit is not the researcher but the research personnel. Careful definition of sub-categories is needed, in order to ensure that the data collected about researchers are compatible with those used elsewhere. This emphasises the importance and need to ensure comparability of measured variables across surveys so that becomes possible to monitor the evolving share of researchers and other types of S&T personnel between the public and the private sectors and link this evolution to supply/demand determinants and inter-institutional mobility
- Methodological work and surveys to better track the evolution of supply and demand of HRST across sectors and institutions, as well as inter-institutional mobility of researchers and career paths of cohorts of doctorate holders trained in Chile or abroad

---

[41] It is to be hoped that Chile will follow the better practice to undertake separate innovation and R&D surveys for which sampling size and criteria and questionnaire content may be rather distinct. On the other hand R&D related questions may be usefully posed in enterprises censuses that are usually undertaken with a lower frequency

[42] An added difficulty is the poor, or absence of, identification of foreign sources of R&D funding, apart from the funding of large scientific projects such as in astronomy

[43] On the one hand it could be overstated as all researchers participating in CONICYT funded projects are counted as full time. On the other it could be understated as only researchers participating in these projects are counted

4.3.2.3 Innovation Surveys

It is in the area of innovation surveys that Chile's S&T statistical system has the longest experience in the collection of S&T statistics and the five surveys that have been carried out to date have benefited from accumulated experience as regards sampling, coverage and scope of collected information. However, as highlighted above the innovation survey still suffer from shortcomings that could be overcome by a stricter adhesion to the recommendations made by the latest version of the Oslo Manual.[44] This concerns in particular

- The broadening of the sectoral coverage (in particular as concerns the service sector), a finer classification of enterprises (e.g. by size, type of ownership), and a lowering of the minimal size in the sample
- The broadening of the concept of innovation to non-technological innovations, as well as non R&D based technological innovations
- A greater and more detailed focus on sources of knowledge and modalities of collaboration in the innovation process
- Information providing feedback on relevance and use of innovation support policies.

The sixth survey, which is currently being designed, should in principle take these improvements on board.

4.3.2.4 Indicators Cutting Across Policy Areas

The strategy proposed by the CNIC and endorsed by the Inter-ministerial Committee on Innovation rests on three major complementary pillars: science for development, human capital and business innovation. While statistical work need to be developed to better track the progress made in each of these individual areas and the outcomes of dedicated policies, it is equally important for the CNIC to devote a particular attention to indicators that cut across policy areas and reflect the dynamics of the innovation system, notably as regards the interactions and knowledge flows among agents and institutions. Among these dynamic elements figure prominently

- Industry/Science relationships captured through various forms technological transfers (including licensing and selling of patents by public research institutions), the provision of S&T services, cooperation agreements on projects, longer-term research and innovation partnerships, co-patenting and citations of scientific papers in industrial patents, spin-offs creation from public research institutions[45]
- Flows of knowledge embodied in human resources as measured by evolving skills distribution in the business sector disaggregated by sector (or priority clusters), evolving proportion of S&T Ph.D. in the business sector, and inter-institutional mobility of HRST
- Flows of knowledge embodied in R&D intensive capital goods through input/output matrices that can track the inter-sectoral (at national and international levels) diffusion of embodied technologies

[44] OECD/Eurostat (2005)

[45] Policy and measurement issues related to this area are reviewed in OECD (2002), *Benchmarking Industry-Science Relationships*, Paris

- Impact of innovation support policies on the "behavioural additionality" of innovative firms through the development of qualitative indicators that highlight the complementarity among policy instruments that foster this additionality[46]

[46] Policy and measurement issues related to this area are reviewed in OECD (2006), *Government R&D Funding and Company Behaviour – Measuring Behavioural Additionality*, Paris

5. Evaluation at the Programme Level

This Chapter reviews the way the I-O-O-I model and the evaluation criteria it implies can be applied at the programme level. It talks first about evaluation criteria, second about which methods to use and third about he use of indicators in relation to programme-level evaluation.

5.1 Making the I-O-O-I Logic Specific to Programmes

At the level of the intervention logic, programme evaluations have a common core. At the level of the individual programme, every one is different. The programme evaluator should

- Establish and test the intervention logic
- Explore and analyse key processes involved in the programme
- Consider what baselines and indicators can be used or created in order to help assess the actual progress of the programme
- Identify and consider the stakeholders involved – not only beneficiaries but others who have an interest in the programme and its success – and include them in the evaluation as sources of data and/or recipients of information from the evaluation

5.1.1 Intervention Logic

An evaluator must describe the programme being evaluated and a comprehensive programme description clarifies all the components of a programme, including its intended outcomes and impacts so that the evaluation can be focused on the most important questions. A key task for the evaluation is to examine the validity of the programme's intervention logic (its **relevance**). This should not involve redoing all the analysis that was needed in order to design the programme but it should involve testing during the course of the evaluation that the needs addressed are indeed (still) need – for example by including questions about them in questionnaires and in interviews.

Figure 18 Example of a Simple Intervention Logic

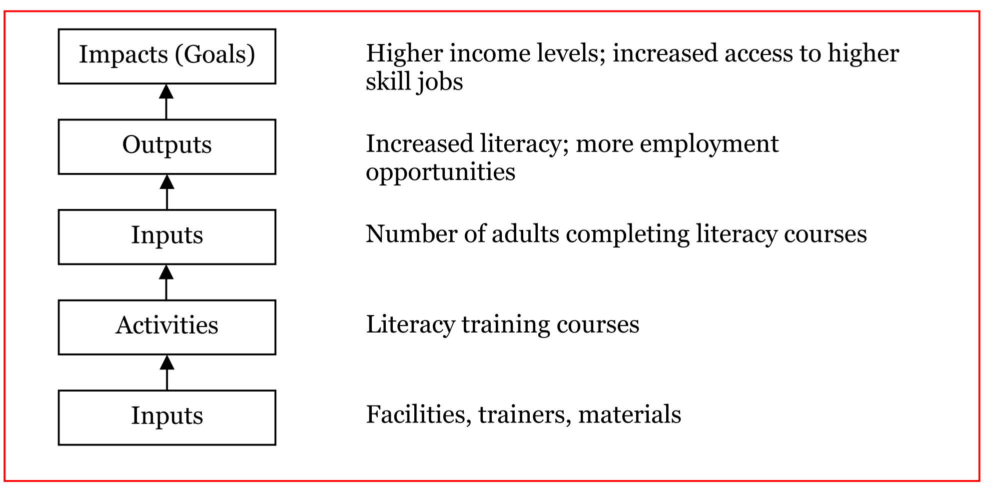


Figure 18 offers a very simple intervention logic – in this case for an adult literacy programme – by way of illustration. The intervention (or programme) logic can be thought of in the following sequential way

- A public action is undertaken for a **reason**

- It has **objectives** which address **needs**
- It provides **inputs** which lead to **activities**
- It achieves **outputs**
- Which lead to **outcomes**
- Leading to **impacts**

Taking these steps in turn; a programme is always undertaken for a **reason** and is conceived with a given set of **needs** in mind. These needs might be the socio-economic problem the programme is seeking to address for example. These **needs** are usually documented in a range of policy and strategy documents. They should ideally be described and analyses in a **programming document** that makes the case for funding the programme. In the ideal case, too, the programming document contains a **baseline**: a measurement of the value of key variable before the intervention, so that it is possible to compare the situation at the outset with the situation later on, when an evaluation is done.

In order to address the **needs**, programmes pursue a number of **objectives** (desired effects). To address these programme objectives it is necessary to have a set of **inputs**, which can include the following

- The programme budget which can include funding from different sources, matched funding, in-kind funding
- Time such as operational and management staff time
- Resources such as equipment (computers, laboratories etc…)
- Physical space
- Staff experience, expertise and skills

We need inputs in order to run the programme **activities** which are the specific tasks (in research and innovation, normally projects) undertaken in order to deliver **outputs**.

Outputs are the programme delivery targets - concrete goods or services, which are produced in order to fulfil the **operational objectives**. Outputs may take a wide range of forms, such as documented knowledge, facilities, services or information.

The **outputs** from the activity lead to **outcomes**[47]**. Outcomes** are advantages (or disadvantages), which direct beneficiaries obtain at the end of their participation in a public intervention. **Impacts** are the wider effects on society. Outcomes and impacts are sometimes together referred to as the **effects** of an intervention. Impacts can be observed some time after the **results** and **outcomes**. Certain impacts (specific impacts) can be observed among direct beneficiaries after a few months and others only in the longer-term (for example through the monitoring of assisted firms). In some cases the **impacts** of interventions will therefore not be known without long term follow up of beneficiaries or analysis of the effects of outcomes.

As we illustrated in Figure 12, this provides a framework that defines many of the key evaluation questions. However, it systematically steers us away from **unintended effects**, so a search for such effects should form an element within the wider evaluation.

5.1.2 Processes

Most evaluations, especially those guided by an intervention logic, tend to focus on outputs and outcomes – what was achieved by the intervention and whether this represents success or failure. It is important that evaluation also covers the **process** and identifies the learning points from the overall approach. Process evaluation may include looking at how any partnership operated, and whether this cooperation was

[47] In some terminologies there are called **results**

useful and beneficial to achieving objectives, and to the beneficiaries. It may also look at how the programme was managed and how the stakeholders worked together. If processes are examined alongside the programme (rather than just at the end), changes can be implemented and programme operations improved while the programme is still live.

5.1.3 Indicators

Programmes, especially in research and innovation, tend to deal with types and sizes of populations that do not have their own statistical categories. As a result, published statistics are rarely useful and programme evaluation (and design) needs to include the definition of specific indicators relevant to the programme.

Figure 19 Examples of Indicators Relevant to Programme Evaluation

Type of indicator	Related to	Example
Contextual	Needs Reasons	General economic performance – wider than the programme Problems (e.g. admin problems) Structure of the economy E.g. number of start-ups, employment Sectoral issues
Baseline	Needs Reasons Inputs	General performance that the programme is going to address E.g. % awareness % beneficiaries % Potential beneficiaries
Outputs	Inputs Outputs	Internal changes that benefit the beneficiary for example, relating to Human capital Training Factors of production Financet
Outcomes	Objectives Inputs Outputs Outcomes	1. Nominal scales Changes in output (Y/N questions) E g Employment, Profit in business or out of business 2. Ordinal data >x% in the same areas 3. Temporal data An extension of ordinal but with a time element (more difficult to collect)
Impacts/net attribution/ counterfactual	Contextual Baseline Objectives Inputs Outputs Outcomes Results Impacts	How much the outcomes are attributable These use the same types of indicators as outputs and effects but question for attribution to the actual intervention These can be on two levels Associated with effects on beneficiary Associated with effects on the economy.
Inputs relative to impacts/ Performance	Objectives Inputs Outputs Outcomes Results Impacts	Performance relative to impacts This goes one step further and compares outputs or outcomes or impacts

There are other categories of indicators that can be used in any of the above circumstances.

It is sometimes not possible to measure the effects of the programme exactly because that data may not be available, or may be difficult or expensive to collect. In these cases, **proxy indicators** can be used. These are indirect measures – eg number of new televisions as a proxy for increased household income. They may provide a suggestion of programme effects – but need to be viewed as such. They may be less precise but just as efficient in discerning trends.

Sometimes indicators are more qualitative such as changes in attitudes or behaviours. This is because some objectives (such as increased quality of life or social cohesion) cannot necessarily be captured using hard quantitative data alone. Even where hard outcomes are the ultimate objectives **soft indicators** can be used to capture information on progress towards those outcomes.

Difficulties that regularly appear in relation to indicators include

- An absence of indicators (particularly at the level of impacts)
- An excessive number of indicators – some programmes define too many indicators because it is easy. Sometimes they are not necessary and make it difficult to identify the important ones
- Badly defined indicators. Sometimes indicators can be inadequately presented and explained which can make it difficult to understand what has been collected and also to use this indicator again (for example in a similar programme)
- Indicators that are strongly influenced by the context and weakly influenced by the programme. Some indicators may be more influenced by external factors than by the programme itself. Measuring an increase in the whole employment rate of an area will not tell you a great deal about the success of your programme on its own as it is likely to be influenced by wider economic trends rather than the intervention
- Indicators that are difficult/expensive to quantify. Indicators which are difficult to collect can be interesting conceptually but are of no practical value to an evaluation

A simple way to quality-assure the choice of indicators is shown in Figure 20.

Figure 20 The RACER Framework for Programme Indicators

The following can be a useful way of checking the validity and usefulness of indicators.
R= Relevant
Is there a clear link between the indicator and the objective to be reached? Where change is being assessed have baseline data been made available?
A = Accepted
Have specialists/stakeholders been consulted on the construction of indicators? Is the indicator actively used in connection with the intervention itself?
C= Credible
Are there definitions for each indicator including statements of what the indicator shows and how data are collected?
E = Easy
Are indicators updated to reflect changes to interventions?
R= Robust
Are there established criteria or characteristics for assessing the quality?

5.1.4 Stakeholders

Programmes usually involve more stakeholders than just the direct beneficiaries. In many cases, such stakeholders can have effects on the successfulness of the programme or can be affected by it. For example, a postgraduate grant scheme will involve stakeholders that include policymakers, industrial employers, professors, university administrators and perhaps parents of postgraduate students or banks, in addition to the postgraduate students themselves. It is therefore useful systematically to list the stakeholders involved with a programme and to ensure that their perspectives and influence are captured in the evaluation.

5.2 Use of Methods at Programme Level

As we discuss in our Toolkit chapter, R&D evaluation tools are individually unreliable, are liable to various forms of bias and only give partial insight. A robust evaluation therefore needs to use multiple methods. It is desirable to be able to draw statistical inferences wherever possible but it is also necessary to use other techniques in order to understand likely causality and the way processes work. The need to make policy decisions means that evaluators must also be able to deliver the best answer they can provide at a given time, and not wait to collect the perfect data set. There is a point in a (useful) evaluation where the data collection stops and the evaluator has to reach a **judgement**. In this important respect, evaluation is not scientific research, even if we strive to be as scientific in our methods as the circumstances will allow.

Programme evaluation has also to take account of the fact that there is a lot of 'noise' and 'cross-talk' in modern innovation systems. A control group approach to evaluating the effects of programmes on company performance can work well, for example, where there are few relevant programmes and companies are mainly small. In such circumstances, the 'treatment group' is fairly distinct from the 'non-treatment group'. Because the companies 'treated' tend to be simple, it is easy to see changes in performance. Once the treatment group contains larger, multidivisional or multi-activity firms, changes in performance are both harder to measure and harder to attribute.

Growing numbers of support programmes and growing company size and sophistication are both characteristic of more developed economies. Olav Kvitastein[48] at SNF in Norway has convincingly shown that medical-style control-group methods can no longer be applied in the Norwegian economy owing to the lack viable control groups. Most of us who do multiple evaluations in small economies also quickly recognise that there are 'usual suspects' among the beneficiaries, which tend to be successful firms. Give the modest levels of subsidy involved and the often-demanding requirements of government support, this is unlikely to be simple rent-seeking behaviour. Rather, it represents an adjustment of behaviour to an overall incentive system that includes public action – action that ranges from subsidy to the widespread provision of higher education[49]. Programme evaluation has to deal with these changing patterns and the increasing complexity they involve. With development in both the economy and in innovation policy in Chile, the simple control-group approach to impact evaluation will be severely challenged. It will become increasingly difficult to attribute responsibility among measures for changes in company and system performance. This implies, first, a need to move to more qualitative ways of exploring causality – literally following the intervention logic to see whether interventions have desired effects, rather than treating the links from inputs to outputs as a 'black box' and trying to re-establish these links by statistical means. And it suggests, second, the usefulness of treating different parts of the intervention logic with different evaluation and measurement techniques.

Thus, the links from increased R&D to improved economic performance are relatively well established in the macroeconomic literature, even if there is a great deal of work left to do in exploring, for example, what **kind** of increased R&D has the best economic effects. We can thus imagine using evaluation methods that focus on, say, establishing how intervention changes R&D behaviour in tandem with exploiting wider studies to explain the relation between R&D and economic performance. This not only simplifies the methodological problem but also speeds up policymaking. If we can establish, for example, that a particular measure drives up industrial R&D then we can use that fact as a basis for action, without having to wait for a sophisticated evaluation exercise to try (yet again) to show that if you do more R&D you tend to become richer. In fact, the intended outputs and outcomes of many interventions are

---

[48] Olav Kvitastein (2002), *Offentlige evalueringer som styringsinstrumenter: Kravspesifikasjoner og kontrollproblemer*, SNF-Rapport nr. 30/2002, Beren: SNF

[49] Curiously, many economists will regard repeated use of subsidy schemes by firms as rent-seeking, while not regarding a selective hiring policy that favours graduates as rent-seeking – even though the graduates embody massive public investment and their higher levels of performance bring well-documented benefits to the firms

precisely to change variables that we have good, usually research-based reasons to believe drive economic performance such as levels of absorptive capacity, education, innovation activity, etc. Since multiple interventions can affect these variables, it is perhaps in any case more sensible to worry about their aggregate value and effects on the economy at the level of the overall innovation strategy rather than at the programme level.

5.2.1 What Are We Trying to Evaluate?

Chile has as wide a range of research and innovation policies and instruments as most other OECD countries. The institutional implementation is somewhat fragmented and unusual. For example, few other systems involve a research council (CONICYT) as heavily in innovation as is the case in Chile. But we can reduce the complexity by mapping the interventions into groups with similar or connected intervention logics (Figure 21). Then we can discuss what methods are useful in relation to these logics, distinguishing between mid-term and ex-post evaluation needs (so, essentially between formative and summative evaluations).

Figure 21 Categories of Innovation Intervention

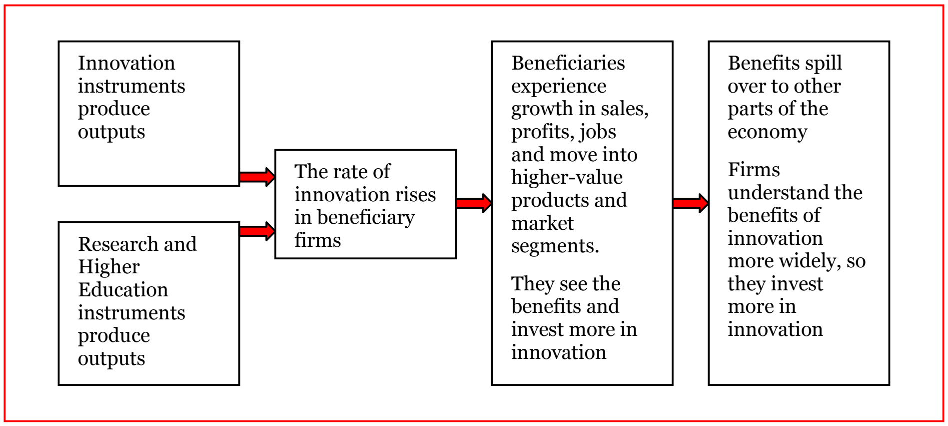
Category	**Programmes**
Creating new, innovative firms	Start-up capital schemes Business Angel networks Company spin-off scheme Start-up support services Building business incubators and training people to work in them
Funding and credits to reduce innovation risks	Innovation project subsidies Subsidies for technology investments by firms Credits to fund innovation projects and investments Support to implement completed innovation projects
Increasing absorptive capacity	Technology management scheme Overseas technology missions Technology internships Training to meet technical skill shortages Support for patenting and IPR protections *Implanting scientists and engineers in industry*
Internationalisation	Business internationalisation platforms Market research in foreign markets
Technology, R&D Infrastructure	*Strengthening national research performers* *Establishing regional research capabilities* Strengthening technology institutes
Technology support largely based on existing knowledge	Subsidy for technical consulting Technology nodes for SMEs
Technology support largely based on new knowledge	Technology consortia (feasibility and R&D stages) Tax credits for R&D projects done with registered technical centres FONDEF
R&D support from the Public Research System	Pre-competitive R&D projects with intended industrial beneficiaries Public goods for innovation (produced in the Research and Higher Education sector) *FONDAP* *Centre of Excellence grants*
Researcher-Driven research	*Principal Investigator grants* *Postgraduate grants* *International research cooperation grants* *Research rings* *Science programmes, eg Astronomy*
Science Communications	*Explora*
'Horizontal' initiatives	Clusters

Note: CONICYT interventions in *italics*

From an innovation perspective, these interventions share a common logic: they all aim to increase the innovativeness of the firm population, in the expectation that this will lead to increased economic performance. We should not forget that innovation and economic performance are not the only motives involved, however. In particular, research is partly performed for reasons of curiosity and culture. It is fortunate that research done for these reasons sometimes also has wider social and economic significance.

From an innovation perspective, the intervention logics of the various measures listed in Figure 21 all share the idea that the intervention should make Chilean industry more innovative and therefore more competitive (Figure 22). The Figure captures that idea that benefits spread outwards from the direct beneficiaries to society more widely and that in many cases this flow of benefits is accompanied by learning that leads to behavioural change (sometimes called 'behavioural additionality').

Figure 22 Shared intervention logic of innovation policy measures



5.2.2 Innovation is not a Linear Process

While it is convenient to represent the shared logic of innovation interventions in a linear way, if we are to design good interventions and evaluations, we need to recognise that the links between research and innovation more generally are **non-**linear. While in the popular imagination science somehow 'causes' wealth creation, the scientific community that researches on research and innovation has comprehensively rejected this 'linear model'. Key elements of this rejection come from

- The fact that trying to make realistic models of the relation between knowledge production, innovation and society quickly leads to much more complex models[50] with many linkages among actors. Innovation processes do not always 'start' at a particular place ('basic' science, or the market) but can be prompted by changes anywhere
- The observation that the innovation system contains many sources of knowledge and many people who produce knowledge – this is not a monopoly of the universities[51]. The Community Innovation Survey and other similar surveys consistently confirm that a very small proportion of innovations have their

[50] David Mowery and Nathan Rosenberg (1978), 'The Influence of Market Demand upon Innovation: A Critical Review of Some Recent Empirical Studies', *Research Policy*, April; SJ Klein and Nathan Rosenberg, "An Overview of Innovation", in R. Landau and N. Rosenberg (eds.) (1986), *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, National Academy Press, Washington, DC

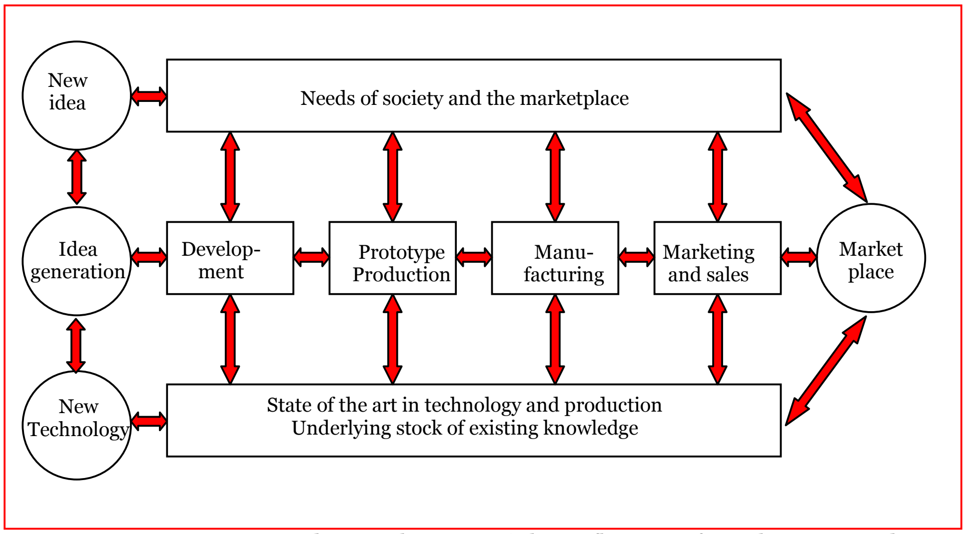
[51] Michael Gibbons, Camilla Limoges, Helga Nowotny, Schwartzman, S., Scott P. and Trow, M. (1994), *The New Production of Knowledge*, London: Sage

antecedents in university research – but also that successful innovators tend to build relationships with the knowledge infrastructure

- The realisation that innovation is essentially done by firms. Much innovation therefore does not involve formal R&D. Indeed, R&D is often not a source of innovation but an effect of innovation decisions[52]. Firms very often seek to innovate by exploiting their existing knowledge assets. Unforeseen problems often emerge, however, and these may require R&D for their solution. From this perspective R&D should be seen not only as a process of discovery but also as a problem-solving activity within already-existing innovation processes
- A key to successful innovation is that firms possess 'absorptive capacity': namely, "the ability of a firm to recognise the value of new, external information, assimilate it, and apply it to commercial ends." [53]. The corollary is that R&D has 'two faces' – one looking outwards for learning and one looking inwards to work with what it learns or knows already[54]

Mowery and Rosenberg's innovation model (see Figure 23) nicely captures some of the newer thinking. It shows a complex process that essentially rests on the **stock** of existing knowledge or the 'state of the art in technology and production' and in which new knowledge adds increments rather than (as the old linear model would imply) being the whole innovation story.

Figure 23 Modern 'Coupling' Model of Innovation



Source: Mowery, D.C. and Rosenberg, N., 'The Influence of Market Demand upon Innovation: A Critical Review of Some Recent Empirical Studies', *Research Policy*, April 1978

There is a substantial research literature on how the innovation process works, which we do not need to explore here. However, it does offer many important clues about innovation processes that can be used to help understand the effects of research and innovation policies. Evaluations need to recognise the complexity of the innovation

[52] Smith, K. and West, J (2005), *Australia's Innovation Challenges: The Key Policy Issues*, submission to the House of representatives Standing Committee on Science and Innovation, Inquiry into Pathways to Technological Innovation, Hobart: University of Tasmania, April 28, 2005

[53] Wesley M Cohen and Daniel A Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation,' *Administrative Science Quarterly*, Vol 35 (1), March, pp128-152

[54] W Cohen and D Levinthal (1989), 'Innovation and learning; the two faces of R&D,' *Economic Journal*, Vol 99, , pp 569 - 596

processes as well as the 'multiple channel' nature of links between public science and industry.

Ben Martin has periodically been reviewing the literature on the links between public research and socio-economic benefits for over ten years. In the latest version of this review, he and Puay Tang list 7 'communication channels' between research and society.

- Channel 1: increase in the stock of useful knowledge
- Channel 2: supply of skilled graduates and researchers
- Channel 3: creation of new scientific instrumentation and methodologies
- Channel 4: development of networks and stimulation of social interaction
- Channel 5: enhancement of problem-solving capacity
- Channel 6: creation of new firms
- Channel 7: provision of social knowledge[55]

They observe that "Case studies and surveys suggest only some of the benefits flow through 'Channel 1' – ie in the form of new useful knowledge that is directly incorporated into a new product or process … although this varies with the scientific field, technology and industrial sector. Hence, attempts to assess the socio-economic benefits of basic research that focus solely on 'Channel 1' will inevitably underestimate the benefits." They go on to point out that in surveys industry tends to focus on Channels 2, 3, 4 and 5 as the most important ones – tending to confirm out point that what you see depends upon where you look. They conclude that

> *… in recent years, much science policy has focused on the 'science push' aspect of innovation rather than 'demand pull'. Yet many of the economic benefits from basic research depend as much, if not more, on the approach that companies take to innovation as on the strength of the science base. In other words, these benefits depend on whether firms adopt a positive and far-sighted approach to drawing on the results of research through all the channels we have identified. Government policy needs to reflect this fundamental point and find effective ways of influencing the thinking of companies accordingly.*

Bearing the complex nature of innovation and the multiple-channel character of science-society interaction in mind, we can now analyse the intervention logics and therefore the appropriate evaluation methods for the classes of instrument listed in Figure 21.

5.2.3 How to Evaluate Intervention Types at Programme Level

5.2.3.1 Creating New, Innovative Firms

These schemes aim to increase the overall rate of innovation by launching new, innovative firms. A lot of literature and attention focuses on spin-offs from the higher education and research sector, but the greater number of spin-offs come from companies rather than research institutions. Interventions focus either on imperfections in capital markets or on providing a sheltered and supportive environment in which start-ups can benefit from advice and services to help them become established.

Overcoming imperfections in capital markets that make investors reluctant to invest in companies at the pre-start-up or early start-up stages, by providing state seed or venture capital funds that take on some of the investment risk. Some funds take all the risk. Increasingly, state funds try to team with private investors so that the private

[55] Ben R Martin and Puay Tang (2007), *The Benefits from Publicly Funded Research*, SPRU Electronic Working Paper Series No 161, Sussex University: SPRU, June

firm plays a significant role in the investment decision and the state funding reduces the risk to the private partner. Complementary schemes focus on entrepreneurs' ability to communicate their business idea to potential investors, supporting them through consultancy or training so that the investors have better, more understandable information on which to base their decision.

Business incubators provide sheltered environments in which to start new firms. Some are attached to universities, on the assumption that tenants will be exploiting intellectual property developed in or with the help of the university. Others will accept any kind of business. Almost all aim to prove advice on running a business, access to advice on intellectual property protection, partnerships and capital.

Both types of intervention are difficult to evaluate. Inventors and their associations lobby – often powerfully – for improved access to capital. Seed and venture capital funds tend generally to complain of a lack of ideas of sufficient quality in which to invest. One approach to testing the relevance of an intervention is to study the existing private capital markets. Do they in practice offer the type of capital sought? Are there ways in which suppliers of capital from distant markets in practice manage to have distribution channels into the country or region suffering a supposed shortage of risk-willing capital?

Business incubators and science parks more generally (of which they are a key sub-set) are rarely evaluated. National and global science parks associations hold conspicuous conferences and do a great trade in comparing 'best' practices but rarely grapple with proper evaluation evidence. Some important issues we have identified in evaluating science parks and incubators associated with universities include

- Minimal job creation, with academics retaining their university roles rather than committing to their companies
- Focus on services, because these are 'clean' and demand limited capital – but this often also means the opportunities for growth and capital accumulation are also limited
- Displacement, where existing companies are attracted from elsewhere to the incubator or science park. The park management claims credit for 'creating' firms that already existed elsewhere
- Limited networking with the associated university

Figure 24 Creating New, Innovative Firms: Evaluation Methods

Issue	Methods
Relevance	Capital market surveys Surveys of actual and potential beneficiaries
Efficiency	Comparisons of practice Benchmarking costs of management and administration
Effectiveness	Beneficiary surveys Case studies Matched pairs comparisons on firms inside and outside schemes*
Utility	Capital market surveys, comparing with baselines Regional studies and surveys of company activity Microeconomic modelling of private benefits
Sustainability	The creation of well-functioning private capital markets is required for sustainability of the first category of intervention. This can be evaluated by survey Sustainability is not an issue for incubators and associated services. Since these interventions deal with events that happen only once in a company's life, there is no scope for company learning, and there is a continual supply of new start-ups in need of support

* Small numbers tend to rule out conventional control group approaches. Even a matched pairs approach is likely to involve high difficulty in constructing meaningful samples

5.2.3.2 Funding and Credits to Reduce Innovation Risk

Funding and credits to reduce innovation risk are expected to increase beneficiaries' rate of innovation because the state effectively shares the risk of undertaking an innovation project with the company. The state may take on the risk through a simple subsidy, by providing a loan (often a 'soft' loan, that has to be repaid only if the project succeeds) of by providing a guarantee to a bank that makes it possible for the firm to take out a commercial loan. Companies and agencies alike often prefer loans and guarantees because they give the company access to money before making any expenditures and the money repaid from the successful projects can be recycled into support for additional firms. While for projects with moderate risk, a loan now may be better in cash flow terms than a grant later, subsidies can be more important where the innovation risks are high, as when developing new technology.

A recurring problem is that agencies providing innovation loans and grants have to be prudent. They develop more or less elaborate ways to reduce risk. This brings the paradox that their schemes start to behave somewhat like commercial banks, thereby avoiding precisely the risks that they are intended to tackle. At the same time no-one wants state agencies to be wasteful in the way the spend money. It is a difficult dilemma.

Success in innovation projects is notoriously 'skewed'. In a portfolio of a reasonable size, many projects will just about break even; many more will fail and lose money; a very small minority is likely to have huge success, hopefully large enough to compensate for the losses and still to generate a substantial surplus. This skew has to be considered when designing samples. A random sample of beneficiaries, for example, can easily miss the one or two projects that make the big money, leading the evaluator radically to misjudge the performance of the scheme. The skew means that some schemes – especially small ones – may produce poor returns through bad luck rather than bad judgement. The normal portfolio principle applies: namely, that larger portfolios are more likely to obtain high returns and are less vulnerable to individual failures than small ones.

Figure 25 Funding and Credits to Reduce Innovation Risks: Evaluation Methods

Issue	Methods
Relevance	Benchmarks or indicators of innovation Innovation surveys Case studies and surveys of actual and potential innovators
Efficiency	Comparisons of practice Benchmarking costs of management and administration Customer surveys Comparisons of investment performance with commercial lenders
Effectiveness	Beneficiary surveys Case studies Control group and matched pairs comparisons on firms inside and outside schemes
Utility	Microeconomic modelling of private benefits Changes in benchmarks or indicators of innovation compared to a baseline
Sustainability	Surveys to determine whether beneficiaries continue to innovate while reducing their use of schemes over long periods To some degree sustainability is not achievable, since these schemes address market failures

5.2.3.3 Increasing Absorptive Capacity

Wesley Cohen and Daniel Levinthal introduced the idea of 'absorptive capacity' to the R&D and innovation literature in a landmark 1990 article[56]. They define it as "the ability of a firm to recognise the value of new, external information, assimilate it, and apply it to commercial ends." According to Cohen and Levinthal, it is impossible to predict what the 'right' level of investment in absorptive capacity is for an individual firm. This depends on individual circumstances. Equally, the concept of absorptive capacity has not been developed in such a way that it is readily amenable to international benchmarking. The best we can do on this front is probably to compare proxies like the proportions of qualified scientists and engineers in business employment and other 'input' measures such as R&D-intensity.

Interventions to tackle absorptive capacity tend to be of two kinds. The first aims to increase companies' awareness of opportunities and of gaps between their own practice and external practice, in the expectation that this will stimulate them to action. In Chile, these include overseas technical missions. Elsewhere technology audits and benchmarking have been used to raise awareness of opportunities. The second category directly addresses the human resource aspect of absorptive capacity by providing, subsidising, training or lending more technically capable staff to firms. Internationally, both categories of intervention tend to receive very positive evaluations.

Some such schemes limit the number of times an individual company may benefit. Since one of the key objectives is behavioural change brought about by the company's greater understanding of improved absorptive capacity, those who benefit from the interventions are highly motivated to come back for further support to do things that they would now do without support.

Figure 26 Increasing Absorptive Capacity: Evaluation Methods

Issue	Methods
Relevance	Comparison and benchmarking of levels of innovation expenditure and employment of innovation-relevant staff Innovation surveys to understand the rate of innovation Case studies in the form of technology audits, to explore needs at the micro level
Efficiency	Comparisons of practice Benchmarking costs of management and administration Customer surveys
Effectiveness	Before/after company performance comparisons Control group and matched pairs comparisons – qualitative as well as statistical Case studies Beneficiary surveys
Utility	Comparison of indicators of employment of innovation-relevant staff with benchmarks Surveys of companies' attitudes to innovation
Sustainability	These interventions aim at making a one-time change in the behaviour of individual firms. Where they are successful, their effects are likely to be sustainable at the level of the individual firm. However, new firm creation provides a constant flow of potential beneficiaries. Achieving the cultural change that leads new firms to start with high levels of absorptive capacity is difficult and will involve many other measures, at least within the education system.

5.2.3.4 Internationalisation

Geography and the desire for economic development through sales to higher-income countries mean that internationalisation has long been a key issue for Chile – and one which the country's success in export markets shows to have been well tackled in the

[56] Wesley M Cohen and Daniel A Levinthal (1990)

past. The recent pattern of globalisation only increases the need to be internationally present.

Current internationalisation measures by CORFO are business rather than innovation-focused. It might be worth exploring whether some form of technology transfer and partnership scheme would be a useful supplement, providing inward technology transfer as an alternative to R&D as a basis for innovation and strengthening access to international markets.

Figure 27 Internationalisation: Evaluation Methods

Issue	Methods
Relevance	Trade and production statistics Sector- and cluster-based assessments of international market opportunities
Efficiency	Comparisons of management practice Participating companies' proportion of total sales in the sector or cluster
Effectiveness	Beneficiary surveys Cost-Benefit analysis examining private benefits Case studies
Utility	Sector and cluster studies comparing economic and innovation performance with baseline
Sustainability	Studies of the extent to which common internationalisation activities survive the withdrawal of state subsidies and support

5.2.3.5 Technology, R&D Infrastructure

The strategy for innovation for competitiveness identified needs to strengthen the research and higher education infrastructure at all levels, including in the regions. There are options to do so partly using existing instruments, such as centres of excellence, though there appear also to be needs for change in areas like governance as well as increases in the absolute amount of funding provided through the university block grants.

Supporting the universities involves treading a delicate line that provides incentives for them to reform and focus their efforts in areas of national needs while at the same time not jeopardising their academic freedom and autonomy. It is easier to be selective in the support to technology institutes, prioritising those whose focus is consistent with the new selective policies. The regional dimension raises important problems of critical mass. Efforts to regionalise research capabilities into the regional colleges and universities in small countries in the Nordic area have been hampered by very significant problems of critical mass. It is not very difficult to start research activities in themes of regional relevance, but it has proved extremely difficult to raise the capabilities and quality of the resulting research groups up to a level where they can compete nationally and during the development process there is a constant attraction of researchers towards the centre, undermining the regional build-up.

Figure 28 Technology, R&D Infrastructure: Evaluation Methods

Issue	Methods
Relevance	Comparisons of international statistics on size and outputs of the research and higher education sector as well as of industrial use of graduate and research manpower
Efficiency	Studies of management and governance Cost-effectiveness studies Peer reviews of quality and outputs Bibliometric and patents studies Economic performance of technology institutes
Effectiveness	Surveys of researchers and industry
Utility	Institutional reviews of universities and institutes Comparison of statistics on size and outputs of the research and higher education sector as well as of industrial use of graduate and research manpower with baseline
Sustainability	Continuing ability of the strengthened institutions to obtain needed funding

5.2.3.6 Technology Support Largely Based on Existing Knowledge

One class of technology support for industry focuses on companies with technological potential. It helps them to innovate by providing additional knowledge or capability that is 'new to the firm' but that is actually based on existing knowledge or the current state of the art. Technology centres or consultants can supply such support. A key characteristic is that delivering the support does not require any new research. This means that expensive research infrastructure is not needed and that a response to company needs can be made quickly and in a way that can be planned. In Chile, the technology nodes for SMEs are intended to perform such a function.

This type of support is often important in established industries, both to improve performance generally and to meet the specific threat posed by increasing capability and rising technological standards of competing firms abroad.

Figure 29 Technology Support Largely Based on Existing Knowledge: Evaluation Methods

Issue	Methods
Relevance	Innovation and other surveys Sector- and cluster-based assessments of technical level and needs
Efficiency	Cost-effectiveness studies Comparisons with good practice Customer surveys Expert reviews of quality and process efficiency
Effectiveness	Beneficiary surveys Case studies
Utility	Beneficiary surveys Matched pairs and control group studies
Sustainability	As with other interventions aimed at supporting firms, behavioural change is an important success criterion. In this case, increased absorptive capacity in the form of technical personnel could be explored via survey and through longitudinal studies that examine how individual firms interact with such supports over time.

5.2.3.7 Technology Support Largely Base on New Knowledge

This category of support is inherently more ambitious than that based on existing knowledge – though in truth the exact boundary between the categories is often fuzzy. Timescales are longer and the outcome of any individual project is less certain. In many countries, 'research associations' have addressed this need, providing three levels of support

- R&D on common needs of all members, for example addressing shared production process problems, norms and standards, quality
- R&D on problems shared by a sub-set of members, in which case the group themselves may make a technical or financial contribution
- R&D done for or with individual members, usually with a significant financial contribution from the beneficiary company

Research associations have increasingly been complemented by other kinds of research institutes and have become more polytechnic in their scope, loosening the links to individual sectors or clusters. Through technology consortia, tax credits for R&D projects done with registered technical centres and FONDEF's university-industry R&D projects, both CORFO and CONICYT provide variants of this intervention.

Internationally, such schemes are widespread. In practice, the research performers rather than industry often initiate projects. This can be positive in that they tend to have the greater knowledge of technical opportunities, but there is also a significant risk that projects are poorly matched to company needs. Norway redesigned its interventions twenty years ago in order to reduce this risk. As a result, Norwegian 'user-directed R&D' projects (where individual companies or groups of companies obtain subsidy for innovation projects the buy from research institutes) were set up. Formerly, the subsidy was paid directly to the performing institute. In user-directed projects, subsidy goes to the beneficiary companies, which in turn pay the institute, creating clearer customer relationships.

Danish policy links this kind of project to the development of research institute capabilities via the Innovation Consortia (formerly, Centre Contracts) programme. Centre contracts were designed as a mechanism for bringing together R&D needs of a group of companies with the research capabilities of a university in order to generate both usable R&D results and re-usable intellectual capital for the GTS applied technology institutes. These institutes would then exploit this intellectual capital in order to provide technological services to other, often non-R&D-performing companies, thereby generating social returns (externalities). Innovation consortia represented a minor adaptation of the centre contract formula to tackle larger networks of companies, institutes and universities. In these projects, company needs tend to provide a 'focusing device'[57] that draws the attention of the knowledge infrastructure to societal needs. The combination of core funding and user-steered money on the one hand gives the institute directors the strategic freedom they need to develop their 'businesses' and the other hand tends to keep the institutes focused on new as well as existing societal needs.

Figure 30 Technology Support Largely based on New Knowledge: Evaluation Methods

Issue	Methods
Relevance	Innovation and other surveys Industry surveys Technology audits and case studies
Efficiency	Reviews of administration and management Expert review of project quality and relevance
Effectiveness	Beneficiary survey Private cost-benefit analysis Case studies In principle matched pairs and control group studies could be one, but the self-selecting nature of the beneficiaries means they tend to have technological and business capabilities that are distinct from non-users
Utility	Sector and cluster studies comparing economic and innovation performance with baselines
Sustainability	These interventions address market failures and are therefore likely to be permanent

At present, Chile uses a system of tax incentives for R&D projects done with registered technical centres. Individual projects have to be approved, so the mechanism is almost identical to the use of direct subsidy and the scheme does not have the advantage normally claimed for a tax incentive: namely, that the beneficiaries themselves control the way the incentive is used. Norway ran a similar scheme – SKATTEFUNN, using the Norwegian technology institute system – for a period at the start of the 2000s before replacing it with a full tax incentive scheme.

Evaluating tax incentive schemes immediately confronts the problem of the counterfactual: it is difficult to deny access to the incentive to a control group. This difficulty can to some degree be overcome by quasi-experiments and comparisons

- Before/after studies of the introduction of schemes
- Using discontinuities, such as changes in scheme conditions or the type of scheme in use
- Inter-scheme (and, therefore, generally international) comparison

Enabling such studies – and, indeed, evaluation studies more generally – to be done requires the systematic creation and management of large data sets. These need to

[57] Nathan Rosenberg (1976), *Perspectives on Technology*, Cambridge University Press

have long time horizons – much longer than the period of the incentive(s) studied – partly so that it is possible to use synthetic data for a control group. Panel data are needed on companies with detailed information over time about R&D activity, inputs, outputs, employment data (including about education levels) and company financials. Particular efforts may be needed to collect data about firms that fall below the size threshold to be included in the normal R&D and Innovation surveys.

5.2.3.8 R&D Support from the Public Research System

The two previous categories of support rely heavily on the presence of companies and their needs to help focus effort on problems whose solution is likely to yield social returns. However, like most systems, that of Chile also has interventions that aim to support innovation by valorising knowledge autonomously generated in the research and higher education system. There are four schemes – two from CORFO and two form CONICYT – attempting to do this. All place the major initiative on the research side and therefore suffer the traditional risk that 'technology push' overcomes signals and needs from the demand side.

A growing number of countries use 'competence centres' (which are essentially variants of the US National Science Foundation's Engineering Research Centres, ERCs) to create academic-industry consortia that together perform R&D over long periods of time (typically 7-14 years). These have the advantage that they use industrial participation to focus research on 'relevant' problems while operating in a timescale that encourages more fundamental research than is normally possible in short-term industry-academic cooperations. Chile lacks such a mechanism.

Figure 31 R&D Support from the Public Research System: Evaluation Methods

Issue	Methods
Relevance	Innovation and other surveys Industry surveys Technology audits and case studies Surveys of valorisation opportunities in higher education and institutes
Efficiency	Studies of administration and management Comparisons and benchmarking with similar schemes Beneficiary surveys Expert review of project quality and relevance
Effectiveness	Beneficiary surveys Case studies Cost-benefit analysis of private benefits
Utility	Beneficiary surveys Case studies Surveys of universities' and institutes' valorisation policies
Sustainability	Surveys of universities' and institutes' valorisation policies, indicating changed behaviour

5.2.3.9 Researcher-Driven Research

CONICYT runs a number of programmes in its primary role as the national research council, whose aims are the traditional ones of ensuring the quality of Chilean university research and continuity within the research community. From the innovation perspective, these are primarily interesting because they produce a supply of human capital that is important to both private firms and public production. Martin and Tang [58] have documented the wider set of links between research and the

[58] Op. Cit.

economy and an earlier version of their findings has been used[59] in the past as an intervention logic for evaluating the basic research grants scheme in Ireland from an innovation perspective, but such evaluations are extraordinary.

Figure 32 Researcher-Driven Research: Evaluation Methods

Issue	Methods
Relevance	General indicators of research intensity and performance Specific analyses underpinning thematic foci in the national strategy for innovation for competitiveness
Efficiency	Management process analysis and benchmarking Peer review of quality
Effectiveness	Bibliometrics Surveys of researchers and universities Case studies
Utility	Review general indicators of research intensity and performance against baselines
Sustainability	Institutional evaluations of universities

5.2.3.10 Science Communications

Most OECD systems have activities that aim to promote the status and attractiveness of science and innovation. Aims tend to be diffuse, including the promotion of scientific understanding as a rational basis for democracy, making science and engineering careers more attractive so as to counteract the 'flight from science' subjects among students in many OECD countries, and increasing the public's willingness to fund science and innovation by demonstrating their benefits and importance. CONICYT's Explora programme has all these aims, as well as a special concern to provide access to science for young people with limited education.

The main difficulties with evaluating science communications programmes are the nation- and culture-wide nature of their goals and the fact that some of these goals are rather 'soft'. In so far as science communications aim to reach the whole of society, the existence of any 'non-treatment groups' is itself an indication of programme failure.

Figure 33 Science Communications: Evaluation Methods

Issue	Methods
Relevance	International surveys of scientific literacy and educational attainment Surveys and statistics about subject choice in schools and universities Surveys about manpower and skill needs
Efficiency	Process review of management and administration Comparison of methods with those used internationally We are not aware of the availability of any meaningful benchmark data
Effectiveness	Public opinion surveys Surveys of public scientific literacy Case studies of focused projects and of initiatives and policies sparked by the intervention
Utility	Comparisons of manpower and educational statistics with baselines Comparison of scientific literacy surveys with baselines
Sustainability	Sustained performance in international attainment surveys and in meeting national needs for educated manpower

59 Erik Arnold and Ben Thuriaux (2001), 'The contribution of basic research to the Irish national innovation system,' *Science and Public Policy*, Volume 28 Number 2 April, pp 86 - 98

5.2.3.11 'Horizontal' Initiatives

The main 'horizontal' policy prioritised in Chile's is the CNIC's pursuit of selectivity through cluster development. This is at an early stage. So far, there has been limited selectivity at the level of research and innovation programmes – for example, the kind of Technology Programmes pursued with great success by TEKES in Finland have been absent from the agenda. Over time, greater involvement of stakeholders in programming in the Nordic style – supported by appropriate checks and balances – is likely to increase the ability of the funders to focus resources on points of industrial need and potential growth.

As they stand, the horizontal nature of the clusters means that each may involve a wide range of interventions and instruments. Their 'micro' nature, however limits the opportunities for using control group and modelling approaches to evaluation.

Figure 34 'Horizontal' Initiatives: Evaluation Methods

Issue	Methods
Relevance	Industrial and sectoral studies of economic performance, combined with prioritisation processes (as already done by BCG)
Efficiency	Cluster governance and management reviews and comparisons with practices elsewhere Reviews to demonstrate cluster-specific use of wider research and innovation funding instruments Development of new, cluster-policy-specific and cluster-specific instruments
Effectiveness	Surveys of in-cluster beneficiaries Case studies
Utility	Industrial and sectoral studies of economic performance, compared to baselines
Sustainability	Industrial and sectoral studies of economic performance, compared to baselines

5.3 Use of Indicators in Evaluation at Programme Level

In OECD countries, with the generalisation of the systemic approach to innovation over the last decades, the emphasis of innovation policy has evolved from the support of large-scale pre-competitive R&D and technological programmes to a broader range of programmes and instruments focusing on the relationships between the determinants of innovation and the mechanisms of knowledge diffusion and transfer. In the portfolio of incentives to firms relative importance of direct support to projects has diminished while that given to measures designed to promote collaboration and broaden the scope of innovation opportunities has increased.[60]

This evolution which marks a shifting paradigm from a "market failure" approach to innovation policy to the broader one of "systemic failure" has had important consequences on the design and practice of evaluation and, consequently, on the range of indicators on which evaluations are based. As highlighted by Papaconstantinou and Polt[61] "The proliferation and widening coverage of policy initiatives have led evaluation to increasingly adopt a portfolio approach, rather than focusing on individual projects; to a greater use of performance indicators; and to a greater convergence between the activities of ex-post evaluation and monitoring."

[60] Georghiou, L. (1997)

[61] Papaconstantinou. G. and W. Polt, in OECD (1997)

While straightforward evaluations such as cost-benefit analyses remain useful on their own right, notably for budgetary processes purposes,[62] more complex or diversified approaches using quantitative and/or qualitative information have thus been developed, mainly since the '90s.[63] Over the years a number of national or international initiatives[64] aiming at improving the information base on the performance of the agents have facilitated this development and lead to the production of new families of statistics and indicators that pertain both to the public and the private sector R&D and innovative activities and their use in policy design and evaluation

- The development of innovation surveys and their systematisation in European Union countries and beyond on the basis of successive editions and improvements of the OECD/Eurostat Oslo Manuals that broadened the scope of innovation to encompass non-technological innovation and put a greater emphasis on the collaborative aspects of the firm innovative activities
- The generalisation of the collection of enterprises micro-data databases providing information related to innovation performance[65] and that could be matched among themselves as well as with the information stemming from innovation surveys and/or used in econometric analyses of the determinants of innovation[66]
- The development of longitudinal panel surveys that allow to monitor the evolving behaviour and performance of firms over time, distinguishing between those that are beneficiaries of support programmes from those that aren't
- Increased attention to the performance and governance of public research institutions as regards their technology transfer activities have led many governments to systematically develop performance indicators that were used to shape policy actions and resources allocation procedures according to criteria that encompass the various missions of these institutions

In response to the frequent multidimensionality nature of the notion of performance of support programmes, qualitative indicators have often been associated to quantitative ones in programme evaluations. This practice has developed both at national and international levels, notably in evaluations of technology transfer programmes, the promotion of innovative clusters and public/private partnerships for research and innovation. In these instances there is an important distinction to be made between the evaluation criteria pertaining to the programme itself and those pertaining to the variety of its applications to specific groups of beneficiaries in terms of institutions, firms, clusters or sectors.

Several recent OECD initiatives focus on evaluation issues and methods at programme level that could be interesting to emulate in Chile.

---

[62] One of the main limitations of cost-benefit analyses is that they cannot -or are nor designed to- capture externalities associated to a support programme or instrument.

[63] OECD (1997) provides a well balanced state of the art of evolving evaluation practices in the S&T and innovation area in OECD countries.

[64] The OECD has played a pioneering role in this matter, notably through the periodic organisation of events on "Blue Sky S&T Indicators" aiming at exploring how the development of new indicators could better respond to evolving policy needs. For the proceedings of such events, see OECD (2001), *STI Review*, No. 27, Special Issue on New Science and Technology Indicators, Paris and OECD (2007), *Science, Technology and Innovation Indicators in a Changing World – Responding to Policy Needs*, Paris

[65] Analyses based on micro-data sets are being increasingly used in the OECD and some of its member countries; it is to be noted however that their reach is often bounded by confidentiality restrictions imposed by statistical offices.

[66] Mairesse, J. and P. Mohnen, (2001), "To Be or not to Be Innovative: An exercise in Measurement", in OECD (2001)

A first one launched in 2003 consists of country case studies that review the funding, governance, management and performance of public/private partnerships for research and innovation (PPPs) programmes.[67] Its methodology is based on the mapping of the various components that are deemed to determine the conditions of success of the PPP programme in terms of

- Creation of knowledge (e.g. co-publications and co-patenting);
- Leverage of public R&D resources on private ones;
- Joint use of S&T infrastructure;
- Involvement of SMEs;
- Mobility of HRST across institutions;
- Commercialisation of research results;
- Sustainability over time, including expansion of research portfolio.

Beyond the global assessment of the PPP programmes which resulted in eventual changes in modalities of financing and governance arrangements, most countries submit individual PPPs to periodic evaluations on the basis of criteria derived from the above ones. These evaluations can lead to an extension of funding over a new fixed term or termination of the programme.[68] Methods developed by the OECD in concertation with reviewed countries could be emulated in the case of the Consorcios tecnológicos supported by CONICYT, CORFO and FIA since 2004.

A second initiative launched by the OECD in 2004 does not explicitly focus on international comparison and evaluation of programmes with similar characteristics as regards their primary objective or modality of support, but rather on the behavioural additionality of various types of R&D support programmes implemented in member countries.[69]

Input additionality captures the increase of a firm's - or a set of firms'- R&D expenditure deemed to be induced by a unit value of public support. Positive additionality points to an effectiveness of the support instrument in terms of R&D but does not say anything about efficiency in terms of innovation performance.[70] The magnitude of input additionality is usually measured through econometric studies and the use of control groups.

Output additionality is usually defined as the proportion of specific outputs (e.g. patents, share of new products or services in sales, skill ratio of personnel) deemed to be induced by the support instrument. Beyond measurement issues, one of the main shortcomings of output additionality is that it does not include all the indirect effects and positive externalities that can be associated with the public incentive or support instrument, such as the development of knowledge assets, the positive effects of learning and collaboration on innovative behaviour etc.

---

[67] See OECD (2004), P*ublic/Private Partnerships for Research and Innovation – Country case Studies*, Paris. The reviewed countries were Australia, Austria, France, Netherlands and Spain.

[68] This is also the case of Israel's MAGNET program in which the government finances up to 66% of approved R&D costs of pre-competitive research consortia which include a number of industrial companies together with at least one academic or research institution. It is to be noted that in Israel resources devoted to such consortia administered my the Ministry of Industry and Trade amount to more than 20% of total public resources devoted to industrial R&D. In this country, periodical assessments have actually led to the termination of consortia.

[69] Ten countries, the EU and the Flanders region of Belgium participated in this initiative. See OECD (2006)

[70] In fact, at the firm level, positive input additionality i.e. increased R&D expenditures induced by public support may lead to undertake over-risky innovation-related investment that may prove unproductive.

The notion of behavioural additionality is rooted in the evolutionary theory of the firm and the building of competitive advantages through knowledge acquisition and collaboration, makes up for the limitations of the input and output additionalities. It is a dynamic concept that reflects an improvement of knowledge management and innovation-related investment decision making through a learning process associated with innovation-related investment that would not have occurred in the absence of public support or incentive.

The most common methodology for measuring or, rather, assessing behavioural additionality effects is that of the survey of beneficiary firms or consortia[71] with questions related to the following issues:[72]

- Improved knowledge and innovation management capabilities;
- Evolving patterns of knowledge acquisition and evolving modes of innovation;[73]
- Lasting effects on collaboration;
- Upgrading and lasting effects on human resources profile;
- Changing patenting strategy;
- In addition questions can also cover issues related to the management and delivery of support programmes by implementing agencies.

To a large extent, behavioural additionality associated to support programmes reflects these programmes' efficiency in strengthening the resilience of an innovation system. Not only through their dynamic impact on beneficiary firms but also, and possibly more importantly, through the build-up of knowledge-enhancement collaboration networks. While the notion of behavioural additionality does not easily lend itself to the direct quantitative measurement of socio-economic impacts of S&T and innovation support programmes it introduces dynamic and systemic qualitative elements in the comparative evaluation of programmes that may have comparable input or output additionality effects. It therefore provides a useful tool for assessing the adequacy of the policy mix of support instruments to address structural weaknesses of the innovation system.

[71] Innovation surveys, customised surveys or questionnaire or interviews.

[72] The specific nature of questions would of course vary according to the nature of the examined programme.

[73] For specific proposals of indicators in this area, see Arundel, A. (2007), "Innovation Survey Indicators: What Impact on Innovation Policy?" in OECD (2007), , Paris

6. Evaluation at the Organisational Level

This Chapter discusses how to use the I-O-O-I idea at the organisational level and what evaluation criteria it suggests for understanding the way organisations (agencies) do their work. It then goes on to discuss what methods to use at this level. Some indicators are relevant for evaluating organisations, but they overleap heavily with those used at the level of the overall innovation system so we postpone discussion of them to the next Chapter.

While there is a huge amount of innovation and research policy evaluation experience at the programme level, there is much less history on which to draw at the level of organisations. Differences in governance also mean that the extent to which individual criteria are relevant to all agencies varies. In particular, some agencies have responsibility for designing programmes and instruments, so their performance of this task must be part of the evaluation. Others, however, have designs effectively handed down to them by their principals. In this case, the programme design function cannot form part of the organisational evaluation.

6.1 Making the I-O-O-I Logic Specific to Organisations

As we indicated in Figure 15, organisations work with inputs in the form of strategic intelligence, management and money. Their activities are mostly programmes, whose outputs are expected to be improved performance by the beneficiary groups to which they are targeted. The outcome of this improved performance should be better performance of the part of the innovation system with which the agency works, which in turn should impact upon the performance of the research and innovation system as a whole.

6.1.1 Relevance

In principle, the relevance criterion involves asking: Based on society's needs, should this organisation exist? In practice, it is rarely felt necessary to ask such a question about research councils or innovation agencies. Where such questions have been raised – in disagreements about whether university research should be funded wholly via the university block grants or via a mixture of block grants and external competitive funding – they have been answered on theoretical grounds[74]. In practice, evaluation of the relevance of an agency tends to involve asking whether it is the right agency for a given task.

Their principals supply agencies with inputs in the form of missions, money and in many cases strategic intelligence. Increasingly, this relationship is codified via a performance contract but that is generally supplemented with close dialogue between the agency and the ministry to which it answers. The principal also defines the **boundary** within which the agency operates: a definition of what is and what is not in scope. This can lead to gaps or duplication in missions or in the instrument portfolio. While strictly this may be regarded as the responsibility of the strategic or ministry level, in practice it appears reasonable also to consider that agencies should identify and address these problems – and take on the coordination tasks with other agencies that therefore become necessary.

74 Erik Arnold, Stefan Kuhlmann and Barend van der Meulen (2001), *A Singular Council: Evaluation of the Research Council of Norway*, Oslo: Royal Norwegian Ministry of Church, Education and Research Affairs,; Erik Arnold (2000), *Research 2000 or Research 1950" Forskning 2000 and the Future of State Research Funding in Sweden*, Brighton: Technopolis

6.1.2 Efficiency

Efficiency is a crucial evaluation criterion for the organisational level, since the function of agents is ultimately to deliver the policies prescribed by their principals.

Evaluation of efficiency needs to ask whether the governance of the agency is well done. Does it receive the right steering signals from above? Does the dialogue with the principal help the agency operate effectively? Are the instructions too open or too prescribed, so that it is hard for the organisation to do it job?

The second major dimension is process efficiency, which we can divide into **core** processes and **non-core** processes. In research and innovation funding, the core processes – those that define the organisation – are

- Programme and instrument or intervention design (This does no apply in all agencies, though arguably agencies that do not design their own instruments should have the ability to criticise instrument and programme design and to propose improvements.)
- Programme implementation, in the form of setting rules of the game for individual programmes, acquiring, monitoring and managing projects and reporting on their individual and collective performance
- In many organisations, especially those that have responsibility for programme or instrument design, there is a need for a 'strategic intelligence' function in the form of a research or analysis department. There should in any case be functions responsible for evaluation and for organisational learning
- Quality assurance of the programmes the organisations run and of the agency itself

Other processes and functions, such as finance, human resources, IT and so on also need to function well according to normal criteria of efficiency and good practice.

6.1.3 Effectiveness

The effectiveness criterion considers the extent to which the activities done by the organisation make a difference to the problems they are intended to address

- To what extent do programmes reach their goals?
- Does the organisation use appropriate instruments?
- Does the organisation add value to its task, over and above delivering programmes?

This involves looking at the outcomes of the agency's actions, that is largely focusing on its effects on beneficiaries rather than the wider society. Of course, the evaluator must take into account the degree to which the organisation designs its own programmes when reaching a judgement.

6.1.4 Utility

The utility criterion addresses the extent to which the activities of the agency have wider impacts on the sector of society it addresses: industry in the case of CORFO, for example. Attributing change to the organisation or to other causes will be a significant problem, since systems performance is affected by many different factors and activities.

6.1.5 Sustainability

Performance improvements at the sub-system level are sustainable only if they have a structural character, leading to persistent behaviour changes. Thus, sustaining an increased output of PhDs requires investments in university infrastructure and increased budgets. Sustaining increased industrial rates of innovation requires greater employment of people with R&D and innovation skills, and so on.

Only some impacts of an organisation can be sustainable. For example, the 'market failure' that leads companies to under-invest in research is inherent to markets. An agency can run programmes that reduce its effects but it cannot abolish market failure. On the other hands, systems failures – such as poor support by universities for protecting intellectual property created on campus – can often be addressed in a more permanent way, in this case by establishing intellectual property support functions in the universities.

6.2 Use of Methods at Institutional Level

Some of the methods relevant at organisational level are different from those that apply at the level of programmes.

6.2.1 Relevance

The need for an implementing agency is a conclusion that can be drawn from the strategy for innovation and competitiveness and the role of the ministries in making rather than implementing policy. The relevance of using a particular agency emerges from the wider organisational evaluation – in particular the parts that deal with the match between the agency's skills and those needed to deliver its programmes. This needs to be complemented at the overall strategic level by considering alternatives in the choice of agency. In practice, such choices are largely theoretical. Mismatches between the skills needed to deliver individual programmes and those of an agency can result in a reallocation of work among agencies. Sometimes agencies need to be reformed or merged. But it is difficult and rare to get rid of the agency function as a whole.

6.2.2 Efficiency

A simple statistical analysis of programmes, projects and beneficiaries is a needed basis for understanding the degree to which the agency addresses its intended beneficiaries.

Understanding the efficiency of a research or innovation agency requires considerable understanding of the specific processes used by such organisations and their relationships with the research and innovation communities. The normal techniques of organisational review are relevant but must be complemented with knowledge specific to research and innovation agencies.

- Governance can be evaluated by comparing practice with the generally accepted principles of the New Public Management about defining clear goals, budgets and reporting systems; use of performance contracts; and the generation and use of adequate strategic intelligence at both the policy level and the agency level. Who governs can be as important as how they govern. The quality of the informal communication and the satisfaction of those involved with the governance also provide important clues about its efficiency. Exploring these issues requires a mix of document analysis and interviews with those involved

- Overall administrative efficiency can be compared in with equivalent organisations internationally in terms of the proportion of the total budget that is spent on administration. There are different norms for research councils and innovation agencies. The fact that different agencies operate different mixes of programmes means there are quite wide ranges of acceptable administrative costs, however

- Core processes need to be studied and compared with good international practice. In some areas, it is possible to acquire benchmark data. For example, many organisations are interested in aspects of performance that are visible to the beneficiaries, such as the time taken between proposal submission and contracting.

- Programme and instrument design processes in research and innovation agencies as well as the quality-assurance of designs are not publicly documented, so they can best be assessed by expert practitioners or evaluators who have studied large numbers of programme designs
- Acquisition of projects through calls for proposals and various types of expert assessment, whether by scientific peers, officials or both and prioritisation. The processes used in research councils tend to be controlled by the scientific community while those in innovation agencies themselves exercise greater control
- Project monitoring and reporting systems can be studied and compared. These are important because they (a) provide a way for agencies to manage project risk and (b) should be sources of performance data that can also be useful in evaluation
- Programme management involves more than simply buying projects and reporting on their results. Depending upon the context, it can involve bringing communities of beneficiaries together, publicising the effects of their work, making links to third parties, and so on. These components of programme added value can be studied through interviews with programme managers and via beneficiary surveys – preferably both
- Depending upon the precise role of the agency in the division of labour about policy and intervention design, different amounts of 'strategic intelligence' are needed and must be provided through an internal analysis unit. The extent to which this unit is in touch with current research on research and innovation, the way in which it acquires and uses evidence and connects this to policy and programme design are important performance parameters likely to influence the efficiency and effectiveness of the agency as a whole. This can best be judged by experts
- Evaluation should provide a key feedback loop from experience to policy and programme design and to process improvement. Sometimes this is incorporated into the 'strategic intelligence' function. Evaluation expertise is required in order to assess the extent to which evaluation is embedded in organisational learning and the degree to which the organisation has developed an 'evaluation culture'. This can be complemented by a meta-evaluation (in the sense of 'an evaluation of evaluation processes')

- Non-core processes such as finance, human resource development, etc can be evaluated against recognised good practice through simple comparison

6.2.3 Effectiveness

At the simplest level, the effects of the programmes it runs measure the effectiveness of what an agency does. If there is a regular cycle of programme evaluation, these evaluations can form the basis of a **meta-evaluation** or systematic review that reviews the findings of the programme evaluations and draws overall conclusions across multiple programmes.

However, if the agency is only evaluated through the 'windows' provided by the individual programmes, much of its importance can be lost. It is often useful to do a beneficiary survey that tackles beneficiaries' views of the aggregate effects of the agency, the adequacy of the programme portfolio for meeting needs (for example, whether there are gaps or overlaps with the work of other agencies), beneficiaries' perceptions of the agency as 'customers' so that it is possible to see the overall agency role. Do companies use a series of CORFO programmes to help them develop and build their capabilities? How do researchers use CONICYT and others' programmes to develop their careers? What problems do they experience in doing this? How do they see the aggregate effects of the agency on themselves over time? And so on.

Opinions differ about whether it is positive or negative for different agencies to have overlapping functions in research and innovation. Functional duplication can increase administrative costs and confuse potential beneficiaries. However, having multiple sources of funding opportunities can also increase the overall take-up of the intervention measure through better overall information and by allowing clients of particular agencies to continue using their established agency relationships. It is possible to reach judgements about individual situations through beneficiary surveys, comparison of client groups of different agencies and assessment of the respective take-up of overlapping schemes, but there is not generally valid principle that shows overlap is always bad.

Sometimes it is possible to do control group analysis. For example, in the evaluation of the Research Council of Norway[75], we were able to distinguish RCN-funded from non-RCN-funded scientific publications and to use bibliometric methods to establish that those scientists funded by the Council tended to publish more and to be cited more than other Norwegian scientists in equivalent fields. This confirmed that RCN's selection procedures tended to choose the better-performing scientists from among the applicants. Depending on the extent of CORFO's 'market penetration' a similar exercise looking at the economic performance of assisted compared with non-assisted companies could be considered.

Almost no evaluations include longitudinal analysis – looking at interactions with individual beneficiaries over time or considering the long-term impacts of interventions. As a result, the effects of interventions over time are likely to be under-estimated. VINNOVA in Sweden[76] has in the last few years done a series of long-term 'effect studies' in areas where its predecessors appear to have intervened some decades earlier. These carefully selected examples show very large, positive impacts. More systematic use of long-term and longitudinal studies would give a more reliable understanding of effects and how they are achieved (or sometimes not achieved) than solely focusing on trying to measure the impacts of interventions that are still in progress or that are too recent to have reached their full effect.

Comparison of the programme portfolio with that of equivalent agencies in other countries is a useful way to identify gaps and opportunities. This can usefully be incorporated into the evaluation.

6.2.4 Utility and Sustainability

Evaluating the overall utility of an organisation's work means going back up to the level of the sector – industry, higher education, etc – whose performance the agency aims to improve and therefore reaching for the baselines set out for that sector in the strategy for innovation for competitiveness and in available indicators. Attributing changes to the work of the agency is difficult – but if there is a body of evaluation that shows programmes to have been effective, it is plausible to attribute at least some of the change to the agency. Evaluations and other studies will also help show the extent to which structural and behavioural changes have been triggered and therefore the extent to which the organisation's impacts are sustainable or need continuous intervention.

[75] Arnold, Kuhlmann and van der Meulen, (2001)

[76] see www.vinnova.se

7. Evaluation at the Strategic Level

This Chapter discusses how to apply the I-O-O-I idea to the strategic level, what evaluation criteria it implies, what methods to use and relevant indicators.

The strategy for innovation and competitiveness is not only a document but also a major act of coordination of policy and implementation. Its evaluation will be a very public process that should mesh with the further development of the strategy. The evaluation needs therefore to enlist key opinion formers and to feed into a policy development cycle. It also has the character of a public performance: it should be overlaid with more or less open meetings where the strategy can be debated and integrated into wider political debate.

7.1 Making the I-O-O-I Logic Specific to the Strategic Level

As we indicated in Figure 15, the strategy for innovation for competitiveness has no other resources than information and influence at its disposal. It undertakes studies, monitoring and evaluation. Its output is a series of policy documents, including the strategy and various kinds of publicity activities intended to influence the development and implementation of policy. Through its influence on the agencies, the outcome of the CNIC's strategic work should be improved performance within each agency's sphere of influence culminating in impacts on the performance of the innovation system as a whole, increased wealth and welfare.

7.1.1 Relevance

At the level of the strategy, the relevance criterion is quite simply whether the CNIC strategy is consistent with real national needs. Do the objectives of the strategy meet national needs? Would it be better to tackle the need for improved innovation and research policy in some other way?

7.1.2 Efficiency

Efficiency relates to the CNIC's process for producing and communicating its strategy and the degree to which it can use it to influence policy. This can include issues such as the timescale and the quality: Did it take too long? Could the quality of the analysis be improved? Did the CNIC involve the right stakeholders? How well does the CNIC do other processes, such as monitoring?

The strategy for innovation for competitiveness covers a very long period. To be efficient, it needs to have a process of periodic review, and the CNIC needs constantly to be doing 'bottleneck analysis': promoting discussions and funding ad hoc studies in areas of apparent problem, in order to see whether the strategy itself should be modified.

7.1.3 Effectiveness

Effectiveness relates to the way the strategy is implemented and the attainment of its immediate goals. To what extent does it in practice influence the actions of policymakers and the agencies? Does it generate an efficient division of labour in the support system that leads to a rational use of resources? Do those best able to perform activities actually do them? Is the implementation well administered? Are the large numbers of sub- and intermediate goals set out in the strategy being reached?

7.1.4 Utility and Sustainability

Does Chile reach or exceed the growth and other goals set in the strategy within the fifteen-year period foreseen? For much of that period, of course, the question will be: Does Chile appear to be on course to double GDP per head within 15 years? And to what extent are the changes achieved structural or behavioural, so that we may reasonably expect them to remain in place?

7.2 Use of Methods at Strategic Level

Much more than at the organisational and programme levels, evaluation at the strategic level relies on the use of indicators and surveys that routinely take place as part of the process of collecting national statistics. These should as far as possible adhere to international standards in order to allow international comparisons and benchmarking.

7.2.1 Relevance

The correctness of the strategy can be tested by reviewing and criticising it, based on available statistical indicators and current theory/research, notably in the research on research and innovation' tradition as well as in economics. The choice of a strategy as the way to focus attention and policy on the problems of innovation and competitiveness can also be discussed in relation to recent research on the governance of innovation systems, which emphasises the need for coordination across different groups of actors, different sectoral ministries and different levels of policymaking and execution.

However, testing the relevance of the strategy needs also to go beyond desk review to an interview-based survey of key decision makers and influencers, who can provide information about the relationship between the work of the CNIC and the way policy is created. The closest we can get to a counter-factual situation is to look at other countries' experience and to consider whether in the past innovation policymaking appeared more effective in Chile without the presence of the CNIC. Differences in governance mean that international comparisons have to be treated with care, but it can also be relevant to compare the Chilean approach with that in some other countries in order to explore the usefulness of alternative approaches.

7.2.2 Efficiency

The same interviews as those needed for the relevance question can be used to explore the process through which the strategy is developed and communicated. Perceptions of efficiency can be at least as important as reality in gaining acceptance of the strategy and the wider work of the CNIC, so it is important to understand the views of those able to make or break the strategy. The evaluator needs also to reach a judgement, based on stakeholder views and her awn analysis, of the degree to which the CNIC and the strategy have been able to tackle and incorporate the implications of unexpected changes and new information, such as that may appear as a a result of the CNIC's 'bottleneck analysis' studies.

Of course, there are few objective efficiency standards that can be applied to such strategic work. A useful technique may be to document the main activities performed, their resource and time consumption and for the CNIC and its secretariat to make a self-assessment of improvement opportunities, which can then be fed forward into the continuing programme of work.

7.2.3 Effectiveness

Effectiveness at this level is about strategy deployment. Evaluation of the implementing agencies should provide key information about their role, which should be supplemented again by interviews that explore key stakeholders' views. Together, these sources should allow the evaluator to identify mismatches between needs and implementation.

We discussed in an earlier chapter the need for the CNIC to monitor progress against the large number of activities and sub-goals set out in the strategy and constantly to update these. This monitoring will be a main source of evidence here about the deployment and implementation of the strategy. Analysis of the monitoring information will both provide a picture of progress in strategy deployment and identify whether there are areas lagging behind plan, which are proving to be bottlenecks to the realisation of the strategy as a whole.

7.2.4 Utility and Sustainability

The main source of objective evidence about progress towards the strategy goals is in the national and international indicators and statistics, which provide baselines as well as measures of progress towards overall goals. In addition, the evaluator should use information from lower-level evaluations and from the interviews associated with the strategy evaluation to identify areas of under-achievement or areas where it appears further strategic reflection is required. Problems identified need to be fed into an ongoing process of strategy development and revision.

7.3 Indicators at the Strategic and Systemic Levels

Over the last four decades since the work that led to the first edition of the OECD Frascati manual in 1962 and the pioneering work of academics on the contribution of R&D investment on economic growth and productivity[77], policy makers, statisticians and academic researchers have grappled with the issues of mapping the performance of S&T and innovation systems through the development of appropriate indicators and evaluation methodologies.

Although much progress has been made in improving the understanding of the determinants of innovation performance and the influence of policies over that performance, the development of appropriate indicators remains a moving target. As highlighted in endogenous and evolutionary growth theories R&D activities and technological progress are at the same time an outcome of socio-economic development and a driver of change of economic structures and potential growth. Moreover, the ways this dialectical relationship operate in any given country depend to a large extent on initial conditions such as physical and human resources endowments, industrial specialisation, business environment conditions, international openness and S&T as well as on its size and institutional infrastructure.

This state of affairs has impinged on the development and management of information systems allowing to monitor the determinants of innovation performance, and assess the outcome and impact of strategic orientations and policy instruments on that performance. While at a given time there is a general agreement on the core set of statistics and indicators such as those compiled by the OECD and Eurostat, information systems must evolve to better reflect the dynamics of innovation systems, the changing conditions of innovation performance, the needs of policy makers and the increased demands of society for accountability.

The issue of the development of new indicators and analytical methods aimed either at better mapping innovation performance and tracking its determinants, or at improving policy evaluation emerges periodically at international level. New coordinated initiatives are often based on new insights provided by academic research, exchange of experiences or a recognised need to fill information gaps. The OECD has played a central role in promoting such initiatives that could inspire the development of indicators for monitoring and evaluation purposes in Chile.

7.3.1.1 The International Experience

In line with the role it has played in the dissemination of the innovation systems conceptual framework from academia to the policy making community of its

[77] See for instance Griliches, Z (1979), "Issues in Assessing the Contribution of Research and Development to Productivity Growth", *Bell Journal of Economics*

membership[78] the OECD work on S&T and innovation statistics and indicators has increasingly been distributed across three interdependent areas.

7.3.1.2 Methodological Work

Methodological work underlying the definition and compilation of internationally comparable indicators on identified inputs, outputs and patterns of innovation and research activities.[79] In view of its expected access to OECD membership Chile will have not only to comply with the guidelines set in this methodological work, but also to comply with the regular provision of the data. This work is certainly relevant for the monitoring of the strategy and the evidence it may provide for the achievement of some quantitative objectives, such as those related to the R&D intensity of the whole economy or that of the business sector, or to the training of human resources in S&T.

As highlighted above in section 3.1 the CNIC should not have a direct responsibility in that type of statistical activities, leaving it to the innovation division for oversight and other agencies for collection and compilation according to international standards.

7.3.1.3 International Benchmarking and Comparisons

Exercises in international benchmarking of innovation performance based on official statistics[80] are of essentially three types:

- International comparisons of countries across a wide array of indicators without any explicit attempt to articulate the relationships among the phenomena that these indicators represent. This is typically the approach taken by the biennial OECD publication Scoreboard of STI Indicators.[81] This publication regroups 72 indicators in 7 broad categories (R&D and investment in knowledge, Human resources in S&T, Innovation policy, Innovation performance, ICT, Particular technologies and Internationalisation of S&T)[82] and for each of them provides a brief review of global trends and highlights some countries' specificities. Although the Scoreboard is mainly a descriptive document that provides international comparisons[83] it provides useful synthetic information on countries positioning with respect to global trends. The CNIC should ensure that Chile collects the necessary information to allow inclusion of Chilean data for most of the indicators to be covered in the next issues of the OECD scoreboard.
- Indicator-based profiles regroup indicators across countries into categories that reflect performance in predetermined strategic areas[84] and map them in a range that allows to position each individual country vis-a-vis the best and worse performers for each indicator. This technique allows to highlight the weak points in and across categories within the framework of a simplified innovation conceptual framework that identifies the contribution of agents to the

---

78 The creation of the OECD Technology and Innovation Policy Working Party in 1995 marks the institutionalisation of the success of this dissemination role.

79 The outstanding examples are the Frascati, Oslo and Canberra manuals, the ongoing pilot work on HRST mobility, and the development work on globalisation of R&D activities.

80 As compared to those based or relying to a large extent on surveys, such as the WEF innovation index.

81 See OECD (2007), *OECD Scoreboard of STI Indicators – Innovation and Performance in the Global Economy*, Paris for the latest version

82 It is to be noted that the categories have evolved over time, with the regular addition of new indicators deemed to reflect important emerging trends related to S&T policy, innovation strategies and technological developments.

83 Along with a definition of the indicators, as well as in many cases the caveats associated their status of proxies of the real phenomena they are supposed to measure.

84 For instance, generation of new knowledge, industry-science linkages and industrial innovation. See Freudenberg, M. (2003), "Composite Indicators of Country Performance: A Critical Assessment", *STI Working Paper 2003/16*, OECD, Paris

performance of the system. The European Innovation Scoreboard (European Commission, 2003) follows a similar approach that complements its countries' trend charts on innovation policy and performance. The EU has also developed a so called "composite index" of innovation that attempts to synthesise in one figure the information provided by the array of individual indicators and rank the countries according to that index. The OECD has rejected such an approach on the grounds of (i) methodological weaknesses and loss of information associated with the construction of composite indices in the S&T and innovation areas; and (ii) dangers of distorted interpretations when countries specificities are left aside.[85] It can be argued that at best indicator-based-profiles can be used to present in a simplified way the gaps suffered by some countries vis-à-vis better performers in some performance areas and at worst they can be misleading in highlighting gaps that are not determining. In no way can such benchmarking methods substitute for policy analyses based on more complex qualitative and quantitative information.

- Cross-country comparisons based on case studies relying of common analytical frameworks and types of quantitative and qualitative information are probably the most useful method to monitor policy performance against the background of strategic orientations and structural weaknesses and other countries better practices. This approach which puts a premium of a systemic view of the determinants of performance and their interrelationships has been taken by the OECD in 2004[86], to a large extent in response to the shortcomings presented above in the case of indicator-based profiles and composite indices. It requires that the participating countries share the premises of the systemic view, focus the analysis on a similar set of issues and provide qualitative and quantitative information in accordance with an agreed framework[87] (and if required, questionnaire). In the quoted OECD work the main policy issue was that of a transition from a predominant paradigm where innovation mainly occurred in processes based on integrative technologies developed in established large companies or supply chains to one where innovation mainly occurs through more open processes, knowledge transactions and collaboration. This method which can be best implemented in the framework of an international organisation such as the OECD that facilitates the development of common projects among peers and can provide the information infrastructure could be usefully pursued by the CNIC in association with partner countries. One should highlight the benefits to be drawn from international co-operation in terms of learning, information sharing and building and feedback on policy design and implementation.

7.3.1.4 Stocktaking and Debates on Indicator Development in Response to Policy Needs in a Changing Environment

Stocktaking and debates between academic, policy and statistical communities held in the framework of international organisations having a leading role in the development of new S&T and innovation indicators and evaluation methodologies do contribute to

---

[85] See Freudenberg (2003) for a detailed account of the weaknesses of the composite index approach.

[86] Results were published in OECD (2005), *Innovation Policy and Performance – A Cross-Country Comparison*, Paris.

[87] In the case of the mentioned OECD study the participating countries were: Austria, Finland, Japan, Netherlands, Sweden and the United Kingdom. The main policy/performance areas reviewed were: Industry/Science relationships, Public/private partnerships, promotion of high-tech SMEs and barriers to growth, globalisation of R&D and international S&T and innovation networks, rationalisation and governance of innovation policy.

new initiatives. This was notably the case in various events organised in the last 15 years by the OECD.[88]

- The first "Blue Sky" S&T indicators meeting held in 1995 in response to S&T Ministers concerns meeting was instrumental in the improvement of the measurement of mobility of human resources in S&T and of innovation output leading to the development of the OECD Patent Statistics Manual.[89]
- The OECD Conference on "Policy Evaluation in Innovation and Technology: Towards Best Practices" held in 1997 highlighted the importance of the notion of behavioural additionality to assess the dynamic impacts of support R&D and innovation support programmes.[90] It fostered the development indicators that captured this notion and of methods that could highlight its impacts, opening the way for improvements of the Oslo manual. That conference was also instrumental in the development of criteria and methods aimed at assessing the performance public research institutions' knowledge transfer activities.[91]
- The second "Blue Sky" S&T indicators meeting held in 2006 (OECD, 2007) has been a sobering reminder that the ambitious goal of a "science of science and innovation policy"[92] allowing to take policy initiatives and develop support programmes on the basis on robust methodologies allowing to simulate outcomes is still elusive. Progress in evaluation of the of policies that reflect strategic priorities rests on incremental improvements in the measurement of systemic factors that affect outcomes over time, and in particular those that are related to the interaction of the elements of the system, and a greater attention to the measurement of externalities. Hence the premium given to the analysis of policy complementarity in strategic evaluations and the need for better matching of databases.

7.3.2 Possible Priorities for the CNIC

As emphasised above, given its pre-eminent role in steering Chile's innovation system through its recommendations to the interministerial Committee, the CNIC should be involved in the development of the quantitative and qualitative information system that should provide it with the necessary intelligence for that steering activity. In principle, the CNIC should not be involved in the management of that system but, with the Ministry of the Economy and the policy implementing agencies it should ensure that a maximum compatibility among databases is obtained so as to allow interdisciplinary approaches and insights into complementarities between policies.Priority actions to be taken by the CNIC could encompass

---

[88] In general these events responded to demands expressed at high political level. For instance, at its meeting at Ministerial level held in 1995, the OECD Committee for S&T Policy agreed in its conclusions that "there is a need for Member countries to collaborate to develop a new generation of indicators which can measure innovative performance and other related output of a knowledge-based economy". It was also agreed that "trends and challenges to the science system need to be studied further by the OECD" and that "special attention should be given to the data required for assessment, monitoring and policy making purposes".

[89] See Guellec, D. (2001) "New S&T Indicators for the Knowledge-based Economy: Opportunities and Challenges", Introductory Chapter in OECD (2001) and OECD (2009), *OECD Patents Statistics Manual*, Paris.

[90] See OECD (1997)

[91] See Hervik, A., "Evaluation of Research Oriented Research in Norway: The Estimation of Long Run Economic Impacts", Chapter 9 in OECD (1997)

[92] As formulated by John Marburger, Director of the US President Office of S&T Policy at the time.

- Macro and structural indicators and analyses[93]
  - Regular monitoring of the impact of investment in knowledge of Total Factor Productivity
  - Sectoral structure of investment in knowledge and export specialisation
- Cross-cutting policies
  - Evaluations that cut across the responsibilities of implementing agencies using approaches highlighted above that emulate those developed in the OECD framework. This could be the case for programmes that foster industry/science relationships, the inter-institutional mobility of human resources in S&T and the development of technological consortia.
- Behavioural additionality analyses.[94]
  - Comparisons between direct and indirect R&D support programmes;
  - Development of innovation surveys in accordance with the 2005 version of the Oslo manual and their matching with other ad hoc surveys so as to facilitate the comparative behavioural additionality of different support programmes. Results of these evaluations should be used to improve the innovation policy mix and reduce the disequilibria highlighted in the OECD review of Chile's innovation policy. They could also lead to the consolidation of fragmented programmes and improved co-ordination among diverse implementation agencies;
  - Differential effects according to size, technological areas and pattern of knowledge sourcing.
- International sourcing of knowledge and R&D networks
  - Development of appropriate indicators to monitor domestic absorption capacity of, and intensity in, foreign knowledge.
- Institutional performance
  - Beyond review of management processes to be carried out internally by public research institutions, the CNIC could determine performance criteria related to technology transfer, commercialisation and diffusion activities that would affect allocation of institutional fundi

93 Mainly econometric.

94 Including the use of control panels and information about beneficiaries' satisfaction.

8. Managing an Evaluation System

This Chapter considers how the innovation evaluation system should be held together, since it spans many actors and levels. That is not something that can be done by *fiat* but will involve cooperation and coordination across large parts of the state structure concerned with research and innovation. It requires agreement on an overall process and timetable that holds the system together. It needs some guidelines and standards. It requires evaluation capacity. And it needs an infrastructure of indicators.

8.1 Process and Timetable

A number of countries regularly produce Bills or White Papers, where the government discusses research and/or innovation policy. In principle, the fact that these processes are regular means that it is possible to build a larger set of activities around them, both creating national arenas in which to discuss policy and triggering various studies and analyses needed in order to inform the strategic documents themselves. In practice, these opportunities are often wasted, owing to lack of coordination among levels, ministries and stakeholders. The CNIC and its strategy provide Chile with a special opportunity to operate a more integrated and coherent process.

We hope that the discussion of the various levels of evaluation in this document has illustrated the fact that evaluation processes are needed at all three levels: strategy; organisations; and programmes. The fact that there is a national strategy means that the levels of evaluation should be interconnected. Programme evaluations are key inputs into evaluating the agencies that manage the programmes. Organisation-level evaluations are key components of the evaluation of the overall strategy.

Figure 35 shows a cycle for coordinating the role of evaluation in supporting the development of the CNIC and the strategy. There are arguments both for and against coordinating this with the electoral cycle. The CNIC has generally taken the position that its work should not be connected to the political cycle as to do so creates perverse incentives. In many countries, the policy cycle is forced to coincide with the electoral cycle in order to place innovation and research on the agenda of each government.

Figure 35 Evaluation Cycle for Innovation Policy

Level	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Strategy			Strategy evaluation	Strategy revision
Organisations	Organisation 1 evaluation		Organisation 2 evaluation	
Programmes	Mid-term and final programme evaluations	Mid-term and final programme evaluations	Mid-term and final programme evaluations	Mid-term and final programme evaluations

Given the number of programmes in progress, programme evaluation is a rather continuous process of mid-term and final evaluations. These are needed by the agencies involved for operational reasons as well as for accountability. The stream of programme evaluations can be tapped, for example through meta-evaluation, at more or less any time and the resulting information can be fed upwards into organisational evaluations.

Organisational evaluations are quite disruptive. Nor do organisations change much in a couple of years, so evaluations of agencies need to take place at intervals of several years. In most countries, this kind of organisational evaluation has not been undertaken before. Where it has been done, it has largely been done only once.

Norway is an exception. There, the second evaluations of the research council and the innovation agency are following the first ones after ten years. Given the need for such evaluations to feed into the top level of the strategy for innovation for competitiveness, they need to follow the shorter cycle demanded by the innovation strategy. That implies that evaluations of CORFO and CONICYT should be alternately 'heavy' and 'light'. The 'light' one could even be a self-evaluation.

The organisational evaluations should be timed so that they feed into the strategy evaluation. Doing both at the same time would place a big burden on the support system and would stretch the capacity of the specialist evaluation supply side, so organisational evaluations should take place sequentially.

The evaluation of the strategy should precede the point at which it is periodically revised – the Chilean equivalent of presenting a new research and innovation bill every four years. Using such a timetable provides the CNIC with a means to coordinate evaluation across the research and innovation system.

8.2 Guidelines on Evaluation

The main purpose of this document is to provide methodological support to such a common evaluation system. It should be treated in the first instance as the CNIC's evaluation manual and subsequently improved in the light of experience and evolving needs.

A number of additional guidelines would be helpful

- A set of evaluation standards, which set out the principles of ethics, independence, respondent anonymity in surveys, etc that evaluators should follow. A number of potential models exist internationally, such as that of the Austrian technology Platform on Evaluation, whose standards have been worked out in combination by evaluators and evaluation customers[95]
- Agreement on how to present and discuss intervention logic. The prevalence of Logical Framework Analysis in Chile, owing to the influence of foreign funders, means that LFA is the most obvious choice
- Standards or models for 'programming documents', ie the documents used to explain and analyse the need for interventions to take place and to describe their intervention logic. Again, there are international models that can be used, ranging from the loose Rationale, Objectives, Assessment, Monitoring, Evaluation (ROAME) statements formerly required in the UK civil service to the highly prescribed formats require by the European Commission for plans for 'Structural Funds' development funding. Crucially, these documents should contain a logic model in an agreed format and should propose baselines
- Agreement that the I-O-O-I model (or something else like it) be used in defining evaluations, so that they ask a common core of evaluation questions. This is needed in order to make meta-evaluation more feasible
- Agreement that evaluations should be made public, except at the project level (where evaluations often amount to critiques of the work of individual people)

The CNIC and its partner agencies should maintain evaluation plans at the annual (detailed) and 4-year rolling level. These should be shared and discussed among the ministries, agencies and CNIC every year, in order to facilitate coordination.

In general the 'waterfall principle' should apply in commissioning evaluations. That is that the government evaluates the CNIC; ministries/CNIC evaluate agencies and agencies evaluate programmes.

95 www.ft-eval.at

8.3 Evaluation Capacity

Our review of Chilean innovation and research evaluation practice indicated that there are few national evaluators operating in Chile. It is normally beneficial to have foreign inputs to national evaluations because these bring fresh and unexpected points of view as well as providing access to comparative experience and benchmark data. In order to maintain contact with the leading edge of evaluation and policy discussion in research and innovation, Chile should continue to involve foreign evaluators – especially in large programmes, organisations and the innovation strategy. At the same time, there is a need for increased domestic capacity – both to do evaluations and to serve as a collective institutional memory about research and innovation policy in Chile. The CNIC, perhaps together with DIPRES, could usefully initiate a national evaluation society.

8.4 Quis Custodet?

Evaluation is intended to be a useful contributor to policy development, implementation, accountability and learning. Having established a coordinated evaluation process based on this handbook, the CNIC should itself consider its usefulness at the end of one cycle.

Appendix A - Evaluation Toolbox

A.1 Introduction 77

A.2 Macroeconomic and Econometric Modelling 80

A.3 Micro-level Economic Analysis 86

A.4 Cost-Benefit Analysis 93

A.5 Useful Statistical Applications 99

A.6 Analysis of 'Spillovers' 103

A.7 Surveys of Beneficiaries 108

A.8 Peer and Panel Reviews 114

A.9 Case Studies 119

A.10 Historical Tracing 124

A.11 Network Analysis 127

A.12 Prospective Studies 132

A.13 Bibliometrics and Patent Analysis 136

A.14 Webometrics 143

A.15 Meta-evaluation 146

A.16 Organisational Review 150

A.17 Benchmarking/Comparative Study 152

A.18 History 154

A.19 Summary of Evaluation Methods 155

A.20 Authors' Views and Assessment of Methods 158

A.1. Introduction

There is now broad agreement in the general R&D evaluation community that individual techniques have important limitations, so it is very important for evaluations to use multiple techniques in parallel (so-called triangulation of methods), in order to increase their reliability[96]. Limitations of technique mean that we do not quite know what we are measuring or that an instrument only captures one perspective of the phenomenon under study. For example, we can ask industrial beneficiaries of R&D programmes for estimates of the cash flow benefits they obtained, but it is rarely clear (except in cases of failure) how good their estimates are or how much of the total benefits they represent. Many of the available techniques allow us to demonstrate with a fair degree of confidence that there **are** effects, and

[96] Erik Arnold and Ken Guy (1997), 'Technology diffusion programmes and the challenge for evaluation,' in OECD, *Policy Evaluation in Innovation and Technology: Towards Best Practices*, Paris: OECD; Luke Georghiou (1998), Issues in the Evaluation of Innovation and Technology Policy, *Evaluation*, Vol 4 (1), pp. 37-51; Luke Georghiou, John Rigby and Hugh Cameron (eds) (2002), *Assessing the Socio-Economic Impacts of the Framework Programme*, June; Gustavo Fahrenkrog, Wolfgang Polt, Jamie Rojo, Alexander Tübke and Klaus Zinöcker (eds) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impacts of RTD-Policies*, August, etc

sometimes to say that these effects may be quite large in comparison with the state's investments in the intervention being evaluated.

However, naturally enough, a key concern of many policymakers is with the **relative** effects of different types of intervention, since they want to optimise the allocation of scarce resources. Our study for the Dutch Ministry of Economic Affairs on methods and practices for relative performance measurement of innovation programmes[97] identified a number of problems with trying to do this. Many of these problems arise from the fact that, if you do not know what proportion of the total effects you are measuring when looking at two different interventions, there is no logical basis for saying that one is bigger than the other.

There are a number of further problems associated with R&D evaluations which make R&D evaluations particularly difficult. These are

- The 'project fallacy' – the fact that what is administratively defined and funded as a 'project' may bear little relation to the beneficiary's 'real' project or agenda, complicating any attempt to understand and disentangle the effects of the administratively defined project
- 'Skew' – the fact that commercially successful R&D projects normally only make up a very small proportion of any portfolio and the fact that the majority of subsidy may therefore be 'wasted'. (As with buying the winning lottery ticket, however, techniques for predicting which projects will succeed are not well developed.)

An important sub-category of R&D evaluations are impact assessments – evaluations that are concerned with the effectiveness of a policy measure. Typical problems associated with (economic) impact assessment are

- The time lag between intervention and effects differs from programme to programme. There is no single moment in time when the ultimate effects can be compared between more than one instrument
- Given the attribution problem, there is a risk of double counting the effects of various policy measures used simultaneously by companies. The more instruments are included in the comparison the more difficult it will be to define control groups (and in some countries, such as Norway, it can be coherently argued that there are so few companies not involved in the state's R&D subsidy schemes that the idea of a 'control group' is incoherent)
- The variation in risk levels of different policy instruments where some low risk programmes have many incremental and short term effects whereas high risk programmes have fewer but potentially more radical effects in the longer term. Comparing the two in the medium term would always favour the low risk programme and therefore lead to a certain risk averseness of public action, whereas the 'market failure' justification assumes that government acts when risks are too high for the private sector
- The failure to quantify the 'softer' effects that governments want to achieve in changing the behaviour of the target groups in the current cost benefit analyses
- Possible changes in the context of the 'problem' that a government action wants to address. Even though an instrument performed perfectly well in terms of cost-benefit analysis, it could be that the context of the firms and the innovation system has changed drastically in the meantime, making the same instrument ineffective

[97] Patries Boekholt, Maureen Lankhuizen, Erik Arnold, John Clark, Jari Kuusisto, Bas de Laat, Paul Simmonds, Susan Cozzens, Gordon Kingsley and Ron Johnston (2001), *An international review of methods to measure relative effectiveness of technology policy instruments*, report to EZ, Technopolis: Amsterdam

in the future. This would argue against using ex-post cost benefit analysis as the sole input for policy decisions

- Effectiveness measurements typically measure only some of the effects of programmes. In particular, cost-benefit analyses tend to focus on the private returns to intervention rather than the social returns or 'externalities,' which normally justified the intervention in the first place, but which are harder to measure. Because the ratio between the measured and unmeasured effects of individual programmes is not known, different cost-benefit measures are typically incommensurable and cannot validly be compared
- Policy instruments should be seen in the context of their role in the innovation system and the specific objectives and target groups they address. 'Relative effectiveness' in economic terms assumes that interventions are mutually substitutable and can lead to allocation of resources away from vital innovation system functions, such as postgraduate education. The idea that a modern economy can run without postgraduate technologists is about as plausible as the notion that you can make your car go faster by buying a bigger engine and throwing away the wheels

In the next section, we provide overviews of a range of methods for R&D evaluation. These are based on a combination of our own experience, a number of methodological surveys and several key research papers[98]. These sometimes make reference to the use of background statistics from national statistical agency sources, such as production statistics, which we do not discuss further here.

[98] Patries Boekholt, Maureen Lankhuizen, Erik Arnold, John Clark, Jari Kuusisto, Bas de Laat, Paul Simmonds, Susan Cozzens, Gordon Kingsley and Ron Johnston (2001); Luke Georghiou, John Rigby and Hugh Cameron (eds) (2002); Gustavo Fahrenkrog, Wolfgang Polt, Jamie Rojo, Alexander Tübke and Klaus Zinöcker (eds) (2002); Rosalie Ruegg and Irwin Feller (2003), *A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade*, July; Louis Lengrand & Associés (2006), PREST, ANRT, Reidev Ltd, *SMART Innovation: A Practical Guide to Evaluating Innovation Programmes*, A study for D-G Enterprise and Industry, Brussels: European Commission; etc

A.2. Macroeconomic and Econometric Modelling

A.2.1. Description

Early pioneering work on quantitative assessment of the impact of technical change at the macroeconomic level was carried out by Solow[99] who estimated the extent to which economic growth could not be explained by increases in the stocks of productive assets, particularly capital and labour.[100] This and subsequent analyses have overwhelmingly indicated the importance of technical change on economic growth and increases in productivity.

More recent approaches to analysis and evaluation of developments at the macroeconomic level (i.e. relating to nation states) or the meso level (covering broad economic sectors or industries) are mainly based on econometric models, taking advantage of advances in techniques, data availability and computing power. Such models consist of a set of interlinked equations intended to explain variations in one or more dependent variables in terms of variations in independent (explanatory) variables. The calculation procedure involves estimation of the 'best' multiplicative coefficient assigned to each independent variable, a measure of its influence on the dependent variable. The extent to which each equation successfully attributes variations in its dependent variable to the given set of independent variables, and to each independent variable separately, can be measured, providing indications of the overall explanatory power of the model and allowing systematic adjustments to be made to its structure to improve it.

Most industrial countries employ a variety of econometric models for planning purposes, with finance ministries in particular making use of models with many hundreds of equations to forecast future trends in such aggregates as GDP, employment, public finances and trade performance, and to analyse the potential effects of policy initiatives, such as changes in taxation. Other government ministries, many large businesses and academic researchers also employ large econometric models.

Perhaps surprisingly, R&D does not normally appear explicitly as a factor in national econometric models. Macroeconomic effects of R&D are usually analysed using smaller specialised models, comprising a few equations and with a relatively simple structure. In the latter models, dependent variables typically include the growth of (national or sectoral) GDP or productivity, and perhaps employment and trade-related variables. The primary independent variable is normally R&D expenditure, often disaggregated into areas such as public, private and overseas. R&D-expenditure variables may be either exogenous (i.e. specified externally to the model in the form of data inputs to it) or endogenous (calculated within the model from separate equations). In the latter case, business R&D, for example, might be specified as dependent on public R&D, and/or on the extent of fiscal or other government incentives to stimulate private-sector R&D.

Non-econometric macroeconomic approaches to assessing the impact of R&D and technical change include general equilibrium models and input/output models. The former are based on microeconomic theory and comprise a set of simultaneous equations which are solved under the assumption of perfect competition – markets clear and the prices and quantities of all sets of goods are determined, including, at least in principle, provision of research activity. Input/output models derived from national input/output tables, represent inter-industry flows of goods and services

---

[99] R.M. Solow (1957), Technical Change and the Aggregate Production Function', Review of Economics and Statistics, 39, pp.312-320,

[100] More specifically, by use of a production function, typically of the Cobb-Douglas type $Y(t) = A(t)K(t)^{\alpha}L(t)^{\beta}$, where K and L are stacks of capital and labour, Y output and A the 'residual' contribution to output growth over time, accounted for by technical change.

within the economy, and have been used to provide proxy measures for technological flows between sectors.

A.2.2. Uses of Macroeconometric Modelling

To be seen to be effective, public R&D support has to satisfy the objective of ensuring that levels of R&D are raised (assuming that, due to 'market failures' they were suboptimal in the first place – see section A.6), coupled with persuasive evidence that the resultant increased R&D gives the taxpayer value for money in terms of its overall effect on the economy.

Econometric models have been applied to address these issues. In particular, they have been used to examine:

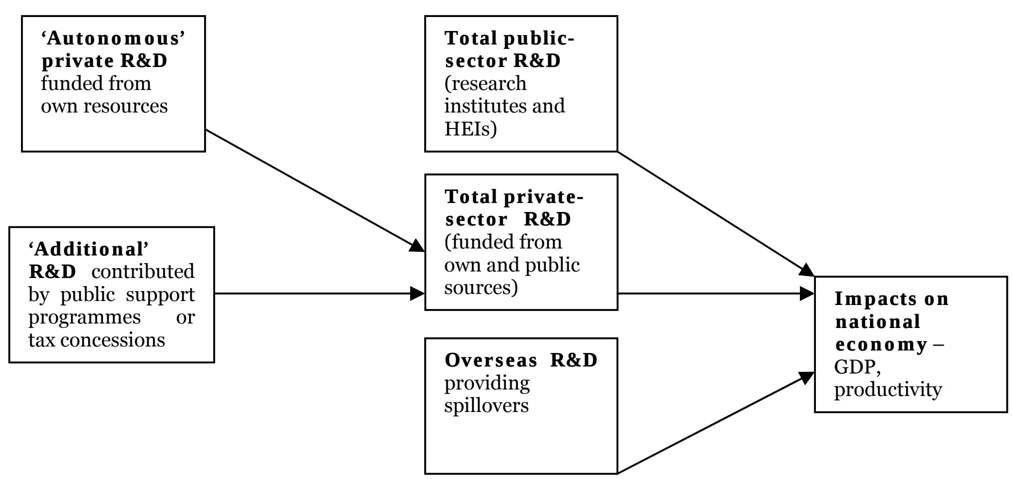
- the effects of public R&D expenditures (including spending on R&D actually undertaken within the public sector, and programmes involving state grants to firms to carry out R&D) on the combined total of business and public R&D – in particular, whether public expenditures stimulate or substitute for business R&D
- the effects of R&D overall on business and the economy as a whole, in particular on national GDP and productivity.
- the effects of R&D tax concessions on levels of business R&D

The first of these has been widely addressed at the firm level, and is discussed in our section on microeconomic approaches (section A.3). Examples of approaches to the latter two issues are presented in this section..

A.2.3. Econometric Models - R&D and the Wider Economy

Figure 36 shows a model capturing impacts of R&D support programmes. The mechanisms of R&D impacts shown (via private, public and foreign R&D) is along the lines suggested by Guellec and van Pottelsberghe[101], work discussed in more detail below (Box 1).

Figure 36 Impacts of R&D support programmes



[101] Dominic Guellec and Bruno van Pottelsberge (2001), R&D and Productivity Growth: Panel Data Analysis of 16 OECD countries,' *OECD Economic Studies*

Econometrically estimated macromodels are used mainly to provide very broad indications of the overall effectiveness of R&D in impacting on the economy as a whole. Total national R&D is input, and its influence on aggregate economic growth or productivity is estimated. Macro-econometric models can thus indicate whether R&D in general is effective in promoting positive economic developments, and the 'average' return in economic growth from each dollar of research expenditure. It is not realistic to expect that it will be possible to isolate the impact on national GDP from a particular R&D project or programme from this type of analysis – there are just too many factors simultaneously influencing the evolution of the aggregates. However, model results can be used to translate the R&D undertaken in a project or programme into macroeconomic benefits *under the assumption that the R&D undertaken is 'typical'* of R&D in general – an assumption which can be useful, but which needs justification.

Box 1 **Macroeconometric models of the impact of R&D**

Guellec and van Pottelsberghe have attempted to assess the impact of R&D from different sources on the growth of total factor productivity of industry across OECD countries. Three sources of productivity growth are considered: the stock of business R&D, the stock of public (government and higher education institutions) R&D, and the stock of 'foreign business' R&D. For each of these, 'elasticities' (the percentage effect on productivity of 1% increases) are estimated. Overall effects on the growth of total factor productivity, and hence on industrial output and GDP, can then be estimated, providing the link to the macro-economy.

Annual data from 16 OECD countries covering the period 1980-1998 is used. The estimated equation is based on the following Cobb-Douglas formulation:

$$P_{it} = A_{(it)} \,.\, B^{\alpha}{}_{(it-1)} \,.\, F^{\beta}{}_{(it-1)} \,.\, P^{\gamma}{}_{(it-2)} \,.\, U^{\delta}{}_{(it)} \,.\, G^{\varepsilon}$$

where, for country i at time t, P is an index of total factor productivity of industry, and B, F and P are stocks of R&D capital relating to business, foreign and public R&D, respectively. B and P are computed using the 'perpetual inventory method'; as with physical capital, it is assumed that these R&D stocks accumulate over time, being enhanced by the latest year's R&D and diminished by an assumed depreciation in the value of earlier R&D (the depreciation rate is taken as 15%). F is a weighted sum of the business R&D capital stocks of the 15 other countries in the dataset.

U is set equal to 1 minus the unemployment rate, with the intention of capturing effects of the business cycle, and G is a dummy set equal to 1 for Germany in 1991 and zero otherwise, to taken account of the exogenous shock of German unification. α, β and γ, the respective elasticities of total factor productivity with respect to the three components of R&D, are the key parameters in the estimation. Principal findings are:

- α, the elasticity of industrial total factor productivity to business R&D, is 0.13, suggesting that an increase of 1% in business R&D generates 0.13% in productivity growth, the effect being larger in countries with research-intensive business sectors, and tending to increase over time
- β, the foreign R&D elasticity, is in the range 0.45 to 0.5, this high figure implying that spillovers from other countries R&D can have more effect on productivity than domestic R&D, providing 'absorptive capacity' is present
- γ, the public R&D elasticity, comes out at 0.17, and is higher in countries where business R&D intensity is higher, suggesting important links between the two

Since R&D represents just a few per cent of overall national economic activity, these elasticities showing the effect of 1% increases in R&D on industrial productivity reinforce the importance of R&D for productivity and economic growth suggested by the results of numerous other studies.

Dominic Guellec and Bruno van Pottelsberge, 'R&D and Productivity Growth: Panel Data Analysis of 16 OECD countries,' OECD Economic Studies, 2001

A.2.3.1 Econometric Models – Impact of Fiscal Measures to Stimulate Business R&D

Most developed countries have a system of tax incentives designed to encourage R&D activities within firms. Schemes differ in many respects; the basic distinction is between 'volume' schemes, where the benefit applies to all R&D (possibly up to a maximum limit), and 'incremental' schemes, where the concession is based on the *increase* in company R&D compared with some base period. Schemes also vary in terms of categories of R&D included, rates of tax concession (sometimes variable according to firm size), types of expenditures included, and taxes to which the concessions are applied. Econometric attempts to evaluate concessions have been subject to extensive review, as exemplified below.

Box 2 **Effectiveness of tax incentives for R&D**

The literature on the effectiveness of these fiscal incentives in stimulating private R&D has been systematically reviewed by Hall and van Reenen. They describe various tax treatments across countries, and discuss approaches used to evaluate their effectiveness, which attempt to establish the extent of the additional R&D induced by tax concessions. These evaluations are primarily based on econometric estimation, and fall into two main categories:

- In the first, an equation is estimated with level of R&D (r(i,t)) expressed as a function of variables like lagged R&D, output, expected demand, and, characteristically, a dummy set to one or zero depending on whether the tax credit scheme is in operation. The coefficient of the dummy gives the R&D induced by the credit. An advantage of this approach is simplicity – there is no need to calculate the actual subsidy to each firm. The downside is that there is the implicit assumption that all firms are subject to similar R&D costs, irrespective of individual circumstances, which is not empirically valid
- The second type of econometric estimation involves inclusion of a price variable (the user cost of R&D) in the equation, capturing the marginal cost of R&D. The response to this price variable gives the price elasticity of R&D, a measure of the response induced by a given tax reduction. Multiplication by the reduction in user price of R&D caused by the tax initiative gives the extra R&D induced.

Hall and van Reenen discuss in detail the alternative specifications, and the results obtained in a number of studies based on them. While the range of figures produced by the various studies is (unsurprisingly) quite wide, they are able to conclude that 'a tax price elasticity of around unity is a good ballpark figure' – in other words, a tax dollar foregone by the government as an R&D concession stimulates a dollar of additional R&D expenditure. Other findings are that the response to an R&D tax credit tends to be small at first, but builds up over time, and that an impotent criterion for success is *consistency* and *sustainability* of the policy, a conclusion supported by several studies.

Given rates-of-return estimates for R&D, this suggests that tax incentives are very good value for public money. Some caveats are appropriate here – for example:

- It is not certain that the additional R&D is 'typical' in terms of potential returns. It may be less productive than 'core' business R&D, which might explain why it seems not to be a priority for the firm's own resources.
- $1 increase in R&D expenditure may not mean a $1 increase in 'real' R&D – there may be price as well as real effects, an issue which also applies to other forms of public R&D support (see 'weaknesses' below).

Bronwyn Hall and John van Reenen, *How effective are fiscal incentives for R&D? A review of the evidence*, Research Policy, 29, pp.449-469, 2000

General equilibrium methods are of little use in analysing the (essentially disequilibrium) effects of technical change, and are not discussed further here. Input/output analysis is covered in our discussion of analysing spillovers (section A.6).

A.2.4. Strengths

Econometric models are susceptible to rigorous testing and examination by third parties, provided full documentation is provided. The estimated coefficients indicate the influence of each explanatory factor (such as R&D) on the dependent variable (for example GDP growth) based on sound statistical theory, and the analysis produces a robust statement regarding the statistical significance of the estimated relationship.

While there are a wide range of estimates of the effects of R&D on the wide economy for different countries at different times, there has been an overwhelming consensus from macroeconometric modelling exercises that the effects are very significant. Evidence strongly suggests that R&D stimulates economic growth to the extent that the national economy gets back many times its investment in R&D.

It is a strength of the method that it has provided such a clear indication of the value of R&D in stimulating economic growth, with consequent benefits for informing policy formulation.

While by no means sufficient as an evaluation tool, econometric methods may justifiably form part of the core of an evaluation system.

A.2.5. Weaknesses

Macroeconomic models are notoriously 'data hungry' – they generally require considerable volumes of data to produce convincing and significant results, which are of course ultimately dependent on data quality. This can be an important issue in the R&D field, where definitional issues (for example of what constitutes R&D) can be problematic.

The 'quantity' of R&D is defined as the money spent on it, but this is not necessarily proportional to the volume of work carried out. The issue of elasticity of supply of R&D resources is important here. One study[102] suggests that a major effect of a fiscal incentive may be to raise R&D workers' wages, rather than to increase the amount of R&D work actually done. Goolsbee has suggested that a large proportion of government R&D assistance goes merely to increase the wages of scientists and engineers – an inflationary rather than a real effect. He regresses real income against total R&D expenditure as a proportion of GDP, and growth rate of GDP, and several dummies reflecting the attributes of individuals, and concludes that a 10% increase in R&D expenditure increases incomes by about 3%. Overall, he suggests that conventional estimates of the effectiveness of R&D policy may be 30%-50% too high. In his words, 'a major component of government R&D spending is windfall gains to R&D workers'.[103] In addition, it should be borne in mind that all results are both temporal and cross-sectional (firm, sector, country) *averages*.

For these reasons, qualitative support for macroeconometric analyses are frequently needed to avoid misinterpretation.

The rigour and sophistication of econometric models has advantages, but also has the disadvantage that considerable expertise is required for their construction. This means that the required skills may not be readily available, and there are many traps for the unwary or inexperienced in applying the technique. If the input data has certain properties, results can be totally misleading unless appropriate transformations are

[102] Austan Goolsbee (1998), 'Does R&D policy mainly benefit scientists and engineers?' American Economic Review, May, pp. 298-302

[103] On the positive side, higher R&D salaries may, in the longer term, promote an increased flow of workers to R&D activity

made to the data prior to analysis.[104] In addition, methodological sophistication means that policy makers by and large need to take results 'on trust' and may be disinclined to accept them if the basis for them is not fully understood. Sophistication also has the drawback that the necessity for extensive testing can be extensive and time-consuming.

A.2.6. Quality Control

Clarity of exposition, to enable peer-group assessment, is the key feature of quality control.

Econometric estimation comprises so many potential traps for the unwary or inexperienced that a funded peer validation exercise to examine critically the veracity of the analysis.

While macroeconometric modelling may not be an essential feature of an evaluation system, understanding of essential features of the technique is important for assessing the validity and usefulness of the work of others.

A.2.7. References

Rachel Griffith, 'How Important is Business R&D for Economic Growth and should Government Subsidise it?, Institute for Fiscal Studies Briefing Note 12, 2000

Austan Goolsbee, 'Does R&D policy mainly benefit scientists and engineers?' American Economic Review, May, pp. 298-302, 1998

Dominic Guellec and Bruno van Pottelsberge, 'R&D and Productivity Growth: Panel Data Analysis of 16 OECD countries', OECD Economic Studies, 2001

Bronwyn Hall and John van Reenen, 'How effective are fiscal incentives for R&D? A Review of the Evidence', Research Policy, 29, pp.449-469, 2000

Rosalie Ruegg and Irwin Feller, 'A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade', NIST GCR 03-857, 2003, pp. 29-34

Robert .M. Solow, 'Technical Change and the Aggregate Production Function', Review of Economics and Statistics, 39, pp. 312-320, 1957

[104] So-called multicollinearity and heteroscedasticity represent two of the most important potential pitfalls in econometric estimation – these are discussed in most textbooks on the subject (e.g. Ronald J. Wonnacott and Thomas H. Wonnacott (1979), Econometrics', Wiley)

A.3. Micro-level Economic Analysis

A.3.1. Description

Micro-level economic analyses take the firm, or a particular innovation, as the unit of analysis, and attempt to assess the payoff to the firm's R&D activities or the innovation.

In this section we cover two main methodologies

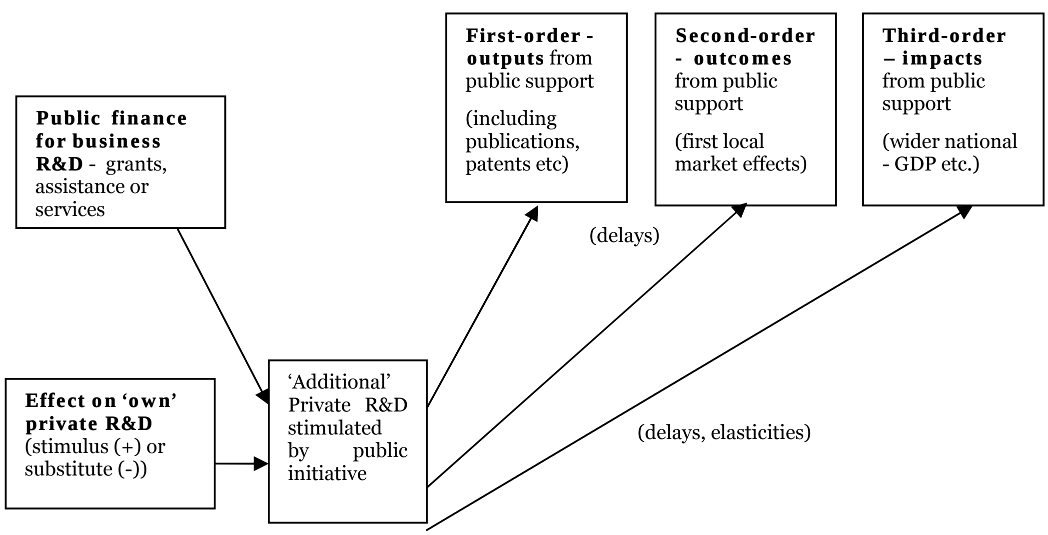
- microeconometric models, analogous to the macroeconometric approaches considered in the previous section, but with firm-level dependent and independent variables.
- microeconomic analysis of the impacts of innovations, particularly in terms of rates of return from the innovation to the R&D-performing firm and to wider society.

As with macro-level analyses, most studies are based on econometric estimation, but there have also been important contributions from applications of microeconomic theory not involving estimation.

Econometric studies are typically concerned with firm-level responses to policy initiatives aimed at stimulating R&D and/or innovation. Unlike their macroeconomic counterparts, where the whole economy or a broad industrial sector is under consideration, only a small subset of firms – those subject to, or participating in, an initiative - is considered. This leads to the use of particular techniques for isolating the effects of the initiative, including control groups and control variables in the estimation.

Figure 37 shows factors of interest to be analysed in such models. The public support, in the form of a grant or tax concession affects the total R&D performed, partly directly and partly indirectly by further enhancing or substituting for autonomous private-sector R&D ('crowding in' or 'crowding-out'). The R&D subsequently contributes to 'outputs' (direct effects such as publications and patents, not themselves income-generating); 'outcomes', benefits to the R&D-performer in terms of increased productivity or sales; and 'impacts', including spillovers to other producers and to the wider economy, which require separate analysis.

Figure 37 Factors of interest in micro-economic models



Following Arvanitis[105], we can characterise a 'typical' microeconometric study in the evaluation of R&D policies as involving estimation of an equation of the form:

$$Y_{i,t} = \alpha_i + \beta_i X_{i,t} + \gamma_i Z_{i,t} + \delta_i P_{i,t} + e_{i,t}$$

where Y, the dependent variable for firm i at time t, is a target variable for the policy in question; α_i is a firm-specific constant; $X_{i,t}$ is a set of explanatory variables; $Z_{i,t}$ are a set of control variables; P_i takes the value 0 or 1, depending on whether or not firm i is participating in the support programme in year t; and $e_{i,t}$ is the error term.

For a policy designed to assist in the commercialisation of a firm's R&D, for example, Y may represent R&D outputs or outcomes, such as patent applications or growth in firm size or productivity, with X including measures of R&D activity. Alternatively, for policies designed to stimulate R&D, the dependent variable itself may measure R&D activity, with X replaced by lagged or historical R&D levels $Y_{i,t-1}$.

Control variables Z may include firm size and industrial sector. δ_i then provides a measure of the impact of the policy; a positive and significant value indicates a degree of 'success' in that participant firms have achieved higher values of the target variable than non-participant firms, other things (as captured by the control variables) being equal.

As in all areas of econometric work, there are large numbers of alternative specifications, structures and approaches, often with greater sophistication with a view to providing greater explanatory power or to avoiding estimation problems. Many of which are covered in the references to this section.

Non-econometric micro studies frequently have much in common with ex-post cost-benefit analyses. Costs include expenditures on the R&D, and costs associated with marketing and manufacturing the innovation. Benefits accrue both to the innovating firm and to wider society, for example via spillovers to other firms and to consumers, through improved products or cheaper substitute products.

In both econometric and non-econometric studies, a key measure of overall benefit is the 'rate of return' to R&D expenditure. [106] It is important to recognise that *definitions* of rates of return can vary considerably, both conceptually and (in the case of social rates) in the *range* of externalities covered.

Despite this, there seems to be a remarkable degree of consensus in the ranges of values obtained.[107] Conventional estimates of the magnitude of the *private* rate of return (from firm-level econometric studies) suggest something in the region of 15-25%. Econometric studies suggest that *industry* rates of return (where spillover benefits to other firms in the same industry are also taken into account) are about double this. Total *social* rates of return, including spillovers *outside* the industry, are

---

105 Spyros Arvanitis (2002), *Microeconomic Approaches to the Evaluation of RTD Policies*, Swiss Federal Institute of Technology Working Paper 55

106 There is sometimes rather loose usage of the terms 'rate of return' and 'elasticity'. In the R&D literature, the following relationships are frequently used:
Elasticity of output with respect to R&D stock
=(DY/Y)/(DZ/Z), (where Y is output and Z the stock of R&D, and D represents an incremental change)
= (rate of return to R&D) x (R&D stock/output)
= (DY/DZ).(Z/Y).
So the elasticity is the proportionate (percentage) increase in output generated by a given percentage increase in R&D, and the rate of return is the *absolute* increase in output for a given *absolute* increase in the R&D stock. However, this appears to be an approximation to the usual definition of rate of return as equivalent to a discount or interest rate.

107 Commenting on the 'surprisingly uniform' estimates of social rates of return, Griliches (1992) observes that 'While one must worry whether this is not just the result of self-imposed publication filters, my own involvement in this work and my acquaintance with many of the other researchers in this area leads me to believe in the overall reality of such findings.'

normally estimated to be much greater still.[108] There has been a consistent story for decades, notably through the pioneering work of Griliches[109] and followers, that spillovers are very significant.

Estimates of rates of return over the last three decades or so have been primarily and increasingly based on econometrics. Griliches[110] suggests the following reasons for this:

- Individual innovations/case studies are not 'representative', and tend to focus on successful innovations
- Such studies are difficult, require primary data collection, and expose the researcher to criticism 'by those who actually know something about the subject'
- Econometric studies are more inclusive and general
- The growing availability of computer resources.

A.3.2. Uses

A.3.2.1. Microeconometric Models – Effects of R&D Support Policies

A wide range of R&D-related policy initiatives can be evaluated by means of micro-econometric models, including the following:

- Publicly-supported research programmes, where proposals are invited from companies for research in specified thematic areas, successful proposals being supported (typically at around the 50% level) from public funds, with the objective of producing additional research with both private and (greater) social returns, the latter justifying the use of public funds. The EU Framework Programme and the US Advanced Technology Programme provide particular large-scale examples.

  The major alternative to R&D programme subsidies is fiscal incentives, a more general measure normally covering all areas of R&D activity and discussed under discussed in the 'macroeconomic modelling' section of this report.

  The effect of R&D subsidies on privately funded R&D – in particular, whether subsidies stimulate additional private R&D or act as a substitute for it, the 'additionality versus substitution' question – has been the subject of a number of studies (several of which are summarised in David et al.[111]). The level of private R&D expenditures are typically taken for the dependent variable, a key explanatory variable being the level of public subsidy. Conclusions generally support the case for 'additionality' rather than 'substitution', but results are somewhat mixed.

- The promotion of particular technologies, in particular computer-aided manufacturing processes, by the provision of services of various kinds, such as information and training or subsidies for consultancy services. A factor indicating the extent of adoption of the technology is a key dependent variable; this variable may in turn form an explanatory variable in a second equation, designed to investigate the impact of adoption of the technology on firm performance.

---

[108] For example, work by Jones and Williams (1998), 'Measuring the Social Rate of Return to R&D', *Quarterly Journal of Economics*, 113, pp. 119-135,), using modern growth theory, suggests that, from a societal viewpoint, optimal R&D investment may be four times larger than actual investment

[109] Zvi Griliches (1958), 'Research Costs and Social Returns: Hybrid Corn and Related Innovations', *Journal of Political Economy*, 66, pp. 419-431

[110] Zvi Griliches (1992), 'The Search for R&D Spillovers' *NBER Working paper 3768,* specifies a simple model which suggests that R&D accounts for up to 3/4 of growth in total factor productivity, most of which comes from the spillover component.

[111] Paul David et al. (2000), 'Is Public R&D a Complement or Substitute for Private R&D?', *Research Policy*, 29, pp. 497-529

- Assistance to small- and medium-sized enterprises (SMEs) is a common policy among OECD countries, with extension services providing assistance in various areas including technology, finance and taxation. Econometric specifications typically involve control variables for sector and firm size, with participation or non-participation in the support programme being tested as an explanation of variations in firms' performance.
- Assistance to firms in forming collaborative R&D consortia. A particular example of this application is detailed below (Box 3).

Box 3 **Payoffs from collaborative R&D**

This example relates to initiatives by the Japanese government involving sponsorship of R&D consortia. The best-known example of this is the very large scale integration (VLSI) semiconductor project running between 1975 and 1985, involving all major Japanese semiconductor producers and 22% government finance, following which Japan became world leader in the field.

The analysis covers 237 government-sponsored R&D consortia which operated between 1959 and 1992, and involved a questionnaire survey in addition to econometric analysis.

A first estimated equation has (log of) R&D spending as dependent variable, with the number of consortia in which firms are involved as an explanatory variable, with controls for firm size (proxied by its capital stock) and industrial sector. The purpose here is to test whether intensity of participation is positively related to increases in R&D. At the margin, participation in additional consortia has a positive and statistically significant impact on R&D expenditure.

A second equation attempts to assess whether the R&D of firms with different levels of involvement in consortia (independent variable) is more or less 'productive', productivity being measured by patents generated per year (dependent variable), while controlling for R&D spending, industrial sector and company attributes. There is a positive and significant relationship between participation in the consortia programme and patenting.

A third equation attempts to provide indirect estimate of the impact of consortia on knowledge spillovers. This is carried out with an extension of the second equation, whereby frequent participants in consortia are distinguished from non/infrequent participants; separate terms measuring the influence of the 'potential spillover pool' (a weighted sum of other firms' R&D, weights reflecting technological 'distances' between the subject firm and others) of these two groups on patenting are included. The influence on patenting appears to be greater for frequent participants.

The conclusions of this econometric work are thus that:

- Participation in R&D consortia tends to be associated with higher levels of R&D spending of participating firms
- Participation in R&D consortia also seems to raise the research productivity (as measured by patenting activity) of participating firms
- At least one channel through which consortia have positive effects may be through effectively augmenting knowledge spillovers.

Correspondingly positive results are found from the questionnaire analysis, which finds that consortia increase the volume of R&D and quicken its pace; government sponsorship of consortia is found to complement, rather than substitute for, private R&D.

Sakakibara, M (1997): 'Evaluating government Sponsored R&D Consortia in Japan', *Research Policy*, 26 (4/5) 447-473

A.3.2.2. Microeconomic Innovation Studies

Non-econometric microeconomic studies (which can be regarded as examples of case studies or cost-benefit analyses) typically represent narrower, but deeper, approaches to evaluation. Much of this work is concerned with individual innovations, as pioneered in particular in the 1950s by Griliches[112] work on hybrid corn. This approach can be useful in analysing economic impacts of particular innovations arising from publicly sponsored research programmes, individual projects within which are generally highly skewed in terms of their outputs and impacts, such that the effectiveness of the overall programme can often be investigated solely on the contributions of one, or a few, constituent projects.

Box 4 **Returns to innovation – case studies**

Mansfield et al. have carried out case studies of 17 specific industrial innovations, estimating the social and private rates of return from investments in each case. Three kinds of innovations are distinguished – product innovations used by firms, product innovations used by households, and process innovations – however the methodology is broadly similar in all cases.

In outline, the methodology is as follows. Costs are incurred by the innovating firm, and possibly elsewhere also - Mansfield et al. note that R&D expenditure by other firms in related innovative activity (whether successful or unsuccessful) should be taken into account, but they find that this is generally small – to a good approximation, private costs equal social costs. On the benefits side, profits to the innovator (private returns) are estimated by deducting costs from revenues accruing from the innovation. For estimating the components of social returns, user firms are assumed to operate in a competitive environment (with prices eroding away profits) and gain no benefits; beneficiaries are the innovator and the final (household) consumers, via a drop in price of the user-industry product (giving a 'consumer surplus').

Data for these calculations, which were made for each of several years following the innovation, are obtained from interviews with executives of the innovating firm and of a sample of 'user' firms, and a review of internal company reports and financial records.

From the above, the net private (social) benefits from the innovation for each year (up to m years subsequent to the innovation) is estimated, and the private (social) rate of return, the interest rate making the NPV of net private (social) benefits equal to zero.

The median estimated social rate of return found for all these innovations is 56%, regarded for various reasons by the authors as a conservative lower bound. The median private rate of return comes out at 25%. Variations between projects are high, in terms of private rates, social rates, and (interestingly) in terms of the differences between them – in one case (that with environmental disbenefits) the social rate is (marginally) below the private rate, while in another it exceeds it 29-fold. The authors observe that 'in about 30% of the cases, the private rate of return is so low that no firm, with the advantage of hindsight, would have invested in the innovation. But the social rate of return was so high that, from society's point of view, the investment was well worthwhile'.

The sample size is too small for conclusions on differences between industries to be drawn, and although there is some apparent difference between product and process

[112] Zvi Griliches (1958), 'Research Costs and Social Returns: Hybrid Corn and Related Innovations', *Journal of Political Economy*, 66, pp. 419-431

innovations (the former have a larger average social-private return 'gap'), it is not statistically significant.

More positive conclusions are that the 'gap' is significantly influenced by the size of the net social benefit (positively), and by the estimated cost of imitation by competitors (negatively). Hence the gap appears greater for 'more important' innovations and those that can be imitated relatively cheaply.

Studies prior to Mansfield et al. were concerned primarily with returns to agricultural R&D, apparently because the data were of better quality than those for other sectors. Estimates of social rates of return by Griliches– 37% for hybrid corn, Peterson 18% for poultry and Schmitz and Seckler for the tomato harvester, were of the same order as those of Mansfield at el., - but slightly lower, perhaps because the analyses were less comprehensive in terms of the social benefits incorporated.

Mansfield, E., Rapoport, J., Schnee, J., Wagner, S., and Hamburger, M. (1977), 'Social and private rates of return from industrial innovations', *Quarterly Journal of Economics*, 91, pp.221-240.

A.3.3. Strengths

Microeconomic studies are more specifically targeted than macro studies, and are better suited to the evaluation of R&D projects and programmes, both ex-ante and ex-post. Being based on established theory and techniques, they are in principle amenable to peer scrutiny and (subject to data availability) amenable also to reproducibility by other researchers.

A.3.4. Weaknesses

The cost of the more specific focus offered by micro studies is that attribution problems are particularly severe. For example, because the coverage of R&D inputs needs to identify those specifically stimulated by the initiative, correction for other influences on R&D is needed, via, for example, control groups or control variables. And comparisons with a control group can be problematic, particularly since participation or non-participation is subject to decision by one or more parties, so that neither the group of participants nor the non-participant controls can be considered as random selections. Because of this, the performance of non-participants may systematically differ from that which participant firms would have experienced in the absence of support, even if there is no interaction between the two groups. But there is a high likelihood of interaction – either in the form of spillovers (tending to benefit non-participants and increasing the impact of the programme) or in the form of competitive pressures (tending to damage non-participants and reducing programme effectiveness). Klette et al.[113] discuss these problems and suggest some ways of ameliorating them using recent econometric advances.

Similarly, on the output side, focus on the subset of participant firms risks the exclusion of externalities, which may need separate treatment.

If the analysis is based on a subset of firms participating in a particular exercise, attempts to generalise the analysis by 'grossing up' (see section A.5.1) will introduce additional uncertainties.

Like their macro counterparts, microeconometric analyses are expensive, are demanding in terms of quality and quantity of data, and are time-consuming and skill-intensive.

[113] Tor Jakob Klette et al. (2000), 'Do Subsidies to Commercial R&D Reduce Market Failures? Microeconometric Evaluation Studies', Research Policy, 29, pp. 471-495

A.3.5. Quality Control

Confidence in the results of a microeconomic analysis is enhanced if a full, and ideally reproducible, description of the methodology and data are provided. Awareness and recognition of the manifold technical problems potentially associated with such exercises, which otherwise can lead to highly perverse results, is essential.

Confidence that conclusions are robust is enhanced if they can be shown to hold under a variety of alternative model specifications. In the same vein, the parallel use of other evaluation techniques such as questionnaires (as in the Sakakibara example above) can add substantial weight to econometric findings. Use of multiple approaches both within the microeconomic analysis and outside it, can help ameliorate the kind of dilemma that can be involved in interpreting evaluation results highlighted by Klette et al (2000) :

> *"…we face the paradoxical situation that if an evaluation study finds little difference between the supported firms and the non-supported firms, it could either be because the R&D program was unsuccessful and generated little innovation, or because the R&D program was highly successful in generating new innovations which created large spillovers to non-supported firms".*

A.3.6. References

Spyros Arvanitis (2000), *Microeconomic Approaches to the Evaluation of RTD Policies*, Swiss Federal Institute of Technology Working Paper 55

Paul David et al. (2000), 'Is Public R&D a Complement or Substitute for Private R&D?', *Research Policy*,29, pp. 497-529

Zvi Griliches (1958), *'Research Costs and Social Returns: Hybrid Corn and Related Innovations', Journal of Political Economy,* 66, pp. 419-431

Zvi Griliches (1992), *The Search for R&D Spillovers*, NBER Working Paper 3768

Tor Jakob Klette et al. (2000), 'Do Subsidies to Commercial R&D Reduce Market Failures? Microeconometric Evaluation Studies', *Research Policy*, 29, pp. 471-495,

Charles Jones and John Williams (1998), 'Measuring the Social Rate of Return to R&D', *Quarterly Journal of Economics*, 113, pp. 119-135

Edwin Mansfield et al. (1977), 'Social and private rates of return from industrial innovations', *Quarterly Journal of Economics,* 91, pp.221-240

Wim Meeusen and Wim Janssens (2001), *Substitution versus Additionality*, University of Antwerp, mimeo.,

Mariko Sakakibara (1997): 'Evaluating government Sponsored R&D Consortia in Japan', *Research Policy*, 26 (4/5) 447-473

A.4. Cost-Benefit Analysis

A.4.1. Description

The UK Treasury 'Green Book'[114] defines cost-benefit analysis (CBA) as:

> *Analysis which quantifies in monetary terms as many of the costs and benefits of a proposal as feasible, including items for which the market does not provide a satisfactory measure of economic value.*

In principle, CBA encompasses the objectives of all evaluation and appraisal techniques, in determining whether, and to what extent, a planned (ex-ante) or completed (ex-post) project or programme has been worthwhile. If all costs and benefits (economic, social, political, environmental), including direct and indirect effects, opportunity costs, and associated timescales, can be accurately identified and quantified, desirability of the undertaking is specified, and no further analysis is needed.

In practice, of course, this represents an unattainable situation. CBA is associated with a particular set of techniques through which it is hoped that this ideal will be approached to a degree sufficient for evidence-based ex-ante decisions to be taken, or instructive information gained from past actions. Details vary, but the set of activities most usually associated with a CBA are as follows:

- Identification of groups of stakeholders positively or adversely affected
- Quantification, where possible, of economic and non-economic costs and benefits to each group, with timescales
- Use of discounting to obtain the present values of costs and benefit
- Consideration of unquantifiable costs and benefits, perhaps with the aid of scoring and weighting techniques
- Comparison of costs and benefits.

These steps are briefly discussed in turn, with particular reference to R&D evaluation.

A.4.2. Identification of Stakeholders

In the case of evaluation of publicly-supported industrial R&D (for example by targeted support programmes or tax concessions), groups affected include:

- the taxpayer (who effectively 'foots the bill' for the initiative, and for whom benefits might be expected to exceed costs overall if the support is to be considered justified)
- companies undertaking the supported R&D
- companies supplying participating firms
- companies in the same industrial sector not participating or directly benefitting from support, who may nonetheless benefit from the results of participants' R&D, or suffer from increased competitive pressures
- Other companies and potential users, who may benefit from 'spillovers', discussed in detail in section A.6.2, which constitute a major justification for public R&D support.

[114] HM Treasury (2003), *The Green Book on Appraisal and Evaluation in Central Government,* London: HMSO

A.4.2.1. Quantification of Costs and Benefits

The following table shows indicates the types of *economic* costs and benefits accruing to each of these classes of stakeholder. The methods by which these can be quantified include many of the techniques discussed elsewhere in this report.

- Direct support and other R&D costs should be readily identifiable
- Results from econometric techniques may be invoked to estimate macroeconomic benefits
- Future expected sales from innovations may be estimated via interviews or surveys
- Spillover effects may be estimated on a case-study basis, or by the use of input/output approaches

Figure 38 Costs and benefits of publicly-supported R&D, by stakeholder

Stakeholder	Costs	Benefits
Taxpayer	Public contribution, administrative costs	'Spillovers' from improved products consumed by householders. Macroeconomic benefits (effects on TFP) from impact of supported R&D and any extra R&D from 'crowding in' effect. Social/environmental benefits
Participant companies	R&D project costs minus public contribution, proposal costs	Net present value of future sales from innovations resulting from the R&D x profit rate
Non-participant companies in same sectors as participants	Costs of later adaptations of technologies developed by participants. Possible depreciation of own R&D stock, possible 'squeezing out' of competitor companies	Knowledge spillovers, leading to market benefits as for participants
Users (other than project participants)	'learning', 'imitation', 'adaptation' costs	Spillovers from failure of producers to appropriate all benefits (through prices) from improved quality or reduced production costs
Suppliers (other than project participants)	Some (as part of old declining markets) may be 'squeezed out' or substituted	New markets
Companies in other industrial sectors	Costs of later adaptations of technologies developed by participants	Knowledge spillovers

Valuation of non-economic costs and benefits is not normally an issue in the evaluation f R&D projects and programmes, but can be extremely important in some evaluation or project-appraisal contexts relating to technology. One example is nuclear power, where a small risk of a major disaster with high social and environmental costs requires consideration of non-economic factors. There is a substantial literature on methods of valuing non-market costs and benefits, some of which are listed in the bibliography. The following are illustrations of approaches that have been used.

Valuations of human life may be pertinent to R&D initiatives relating to health or road safety research, where benefits may include reductions in fatalities (and/or injuries), or where otherwise beneficial technologies may pose potential risks. A number of approaches have been used, based, for example, on insurance premiums, typical court awards for various kinds of injury, or estimates of economic impact as measured by loss of income resulting from injury.

'*Willingness to pay*' or '*revealed preference*' techniques make use of observed behaviour to impute values to benefits such as a peaceful environment (for example, from the extent to which house prices are higher in 'noisy' locations than in 'quiet' ones).

A.4.2.2. Discounting

Discounting is a means of converting a stream of costs and benefits which have accrued, or are expected to accrue, in different time periods, to their values at a particular point in time – their 'present value' - so that they can be added and compared. Discounting is based on the concept that individuals prefer to consume now rather than later, so that a benefit available now is worth more than the same benefit becoming available next year. The 'discount rate' measures the rate at which a given benefit (or cost) successively declines in value, year on year, over a future time horizon.

As a simple example, suppose that a project with a one-off initial cost of $X which is expected to generate a benefit of $1m each year for the next 10 years. To be worthwhile in cost-benefit terms, the total *discounted* income stream - its net present value (NPV) - should be larger than X, the NPV of the income stream being lower for higher discount rates d. For d=3.5%, NPV is $8.32m, while with d=10% it becomes $6.14m, illustrating the significance of discounting and the rather high sensitivity to the (not necessarily rigorously defined) choice of discount rate.[115]

It is important to recognise that the discount rate is *not* equivalent to an assumed interest rate, which measures the rate at which money loses its value and which needs to be accounted for separately. Discounting relates to individual's 'time preference' and measures the cost incurred for deferred consumption. The discount rate can be estimated from research on the extent to which individuals discount future consumption over future consumption; on the basis of such evidence, the UK Treasury, for example, recommend use of a discount rate of 3.5% in cost-benefit analyses.

A related concept used in CBA is the 'internal rate of return', which is defined as that discount rate at which the net present value is zero. This allows comparisons with returns on other investments, and hence whether the proposed undertaking represents a relatively advantageous use of available funds.[116]

A.4.2.3. Unquantifiable Costs and Benefits

There will frequently be important costs and benefits which cannot realistically be assigned either direct or proxy monetary values. These have to be considered separately and judgement has to be used. Their importance can be subjectively weighted and scored between themselves, and in relation to the monetised costs and benefits. Such judgements can be made by use of expert panels, which are considered elsewhere in this report (section A.8).

A.4.2.4. Comparisons of Costs and Benefits

From estimated net present values, together with evidence on non-monetised costs and benefits, an overall conclusion can be reached on the expected viability of a proposed initiative, or on the degree of success of a completed one.

A.4.3. Uses

CBA has been widely used to assess a proposed project or programme, and also for options appraisal – the process of comparing, in cost-benefit terms, alternative means of producing a desired outcome. While primarily an ex-ante procedure for planning purposes, it can also be used ex-post to evaluate a completed undertaking.

[115] The mathematical expression for Net Present Value (NPV) is $NPV=\sum(B_t-C_t)/(1+d)^t$, where B_t and C_t are benefits and costs, respectively, incurred in year t following the base (present) year, the sum being over all years in which costs and/or benefits are incurred.

[116] The internal rate of return is thus calculated by setting the above expression for NPV equal to zero and calculating the resultant value of d.

Applied to a proposed R&D project, a CBA will need to cover the R&D phase and the subsequent phases of commercialisation and production.[117] During the R&D phase, costs but no benefits (revenues) will be incurred, and there is the risk that the intended output (normally a new product or process) will not, for technical reasons, be realised. Assuming the R&D process yields a technically satisfactory prototype, further costs (for example in advertising and 'scaling up') are incurred in the commercialisation phase, where the risk becomes commercial rather than technical.

During subsequent production and sales, both costs and revenues accrue. Benefits include both private returns (to the producer) and social returns (to other groups, from spillovers) – see Figure 38. A pattern of evolution for product sales (such as the traditional 'S' curve of technological diffusion) might be assumed to represent the time profile of revenues.

NPV analysis can then be applied to the expected stream of costs and benefits. If the technical and market risks can be quantified, the benefit stream can be estimated at each stage as an *expectation value*: 'expected revenue if successful' multiplied by 'probability of success'.

A.4.4. Strengths

The major strength of CBA is that it offers a *systematic* and *comprehensive* procedure of appraisal or evaluation. In principle, all the important positive and negative consequences of an initiative are accounted for and given due weight in the analysis. In addition, it is normally relatively easy for assumptions to be understood and assessed by others.

A.4.5. Weaknesses

Weaknesses of CBA relate mainly to difficulties in quantification, which may arise either from uncertainties about the future or from arbitrariness in assigning monetary values to non-monetary outcomes.

Uncertainty is of course always present, and in the case of application of CBA includes risk of technical or market failure, unexpectedly high R&D, commercialisation or set-up costs, and overoptimistic (or pessimistic) expectations regarding revenues. On large-scale investment projects, in particular, cost overruns seem to be almost a fact of life; the UK Treasury has advocated recognition of an 'optimism bias' - 'the demonstrated systematic tendency for appraisers to be over-optimistic about key project parameters' – which, it suggests, should be explicitly accounted for in project appraisals.

Box 5 **US Advanced Technology Program: Medical Technologies**

The objective of this study was to evaluate medical research projects carried out under the US Advanced Technology Program (ATP), and in particular to assess the social return on the public investment. A CBA framework is used to construct and compare two scenarios, respectively with and without ATP funding. Differences between the with-ATP (which of course incorporates the cost of ATP funding) and without-ATP scenarios include (1) the duration of the R&D phase, (2) private-sector R&D investment and its consequences for the likelihood of technical success, and (3) breadth of the technology's applications. Various empirically-based assumptions are used to estimate these effects.

Costs and benefits considered include

- medical benefits to patients. The impact of ATP supported technologies on health outcomes are modelled, a 'quality adjusted life years' (QALYs) parameter being

[117] E.g. Rosalie Ruegg and Irwin Feller (2003), *A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade*, NIST GCR 03-857, p. 144

used to compare changes in well-being made possible by the ATP-developed technology. A QALY is a measure of the utility associated with health outcomes that combines morbidity and mortality into a single measure of annual well-being. QALYs assign each health state a value between zero and one, where zero corresponds to death and one to a year in perfect health. The scale is based on the idea that the value of a year of life varies depending on a persons state of health. In the study, empirical values from the literature for various health-states are used. Monetary values are then assigned to each QALY based on published estimates of people's 'willingness to pay' for avoiding illnesses and accidents. A diffusion model is used to estimate numbers of beneficiaries.

- changes in the cost of health care, categorized as direct medical costs, indirect costs, and intangible costs. Direct medical costs equal the total cost of medical treatment. Indirect costs are the societal costs associated with loss in productivity due to illness and unpaid caregiver time. Intangible costs measure the costs due to the pain and suffering of the patient.
- revenues to companies. Expected private returns are modeled as depending on the following factors: probability of technical success; expected investments and costs for R&D, commercialization and production; and expected revenues.
- private investment and costs, and public investment in ATP funding.

The with-ATP scenario and the without-ATP scenario can differ with respect to three mechanisms of ATP impact:

- project acceleration
- probability of technical success
- project scope

Time profiles for the two scenarios are constructed, costs and benefits are discounted at 3%, and discounted costs and benefits compared. Sensitivity analysis has been carried out using discount rates of 1% and 5%, and with a range of values for various other model parameters. Regarding sensitivity to the discount rate, it is found that the composite social returns on all projects are about 40% lower than baseline at a 5% discount rate and about 75% higher than baseline at a 1 percent discount rate. Sensitivity to estimates of health benefits and to cost parameters also seems high.

Estimated rates of return are large. The expected social return on ATP public investment in these technologies, or the increment to social returns attributable to ATP funding, is estimated at $34 billion in net present value. The expected social rate of return on ATP public investment in these technologies is estimated at an annual rate of 116%.

This study is an impressive application in CBA in terms of its comprehensiveness, attention to detail and imaginative use of existing literature to quantify costs and benefits, but the large number of assumptions required, and the sensitivity of the results to them, raises serious questions rearding the robustness of the results.

Martin, S.A., Winfield, D.L., Kenyon, A.E., Farris, J.R., Bala, M.V., and Bingham, T.H. (1998), *A Framework for Estimating the National Economic Benefits of ATP Funding of Medical Technologies*, Research Triangle Institute, mimeo

Assessment of non-market and non-monetisable costs and benefits will necessarily be potentially subject to serious error. The choice of discount rate is also arbitrary to a degree, and the sensitivity of results to this choice can be high, and need to be explored.

A.4.6. Quality Control

Here one can only advocate that the most important determinants of the conclusions reached in a particular CBA are investigated and recognised, and that the veracity of the assumptions behind them are scrutinised. There is a need also for complete and precise specification of assumptions, for scrutiny and assessment by third parties.

Sensitivity analysis is very important. A CBA typically requires a large number of assumptions to be made which have high inherent uncertainty – the Martin et al. study described above is a case in point – and there is a risk that, *considered in combination*, the uncertainties will render conclusions based on a central estimate all but meaningless. Many large-scale exercises have in the past failed this test.

A.4.7. References

Boardman, A, Greenberg, D, Vining, A and Weimer, D (1996), *Cost-Benefit Analysis: Concepts and Practice*, Upper Saddle River, N.J. Prentice Hall

HM Treasury (2003), The Green Book on Appraisal and Evaluation in Central Government, London: HMSO

Little, I M D and Mirrlees, J A (1994), 'The Costs and Benefits of Analysis: Project Appraisal and Planning Twenty Years On', *in* R Layard and S Glaister eds. (1994), Cost Benefit Analysis, 2nd ed , Cambridge University Press

Martin, S.A., et al. (1998), *A Framework for Estimating the National Economic Benefits of ATP Funding of Medical Technologies*, Research Triangle Institute, mimeo.,

Rosalie Ruegg and Irwin Feller (2003), 'A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade', NIST GCR 03-857

A.5. Useful Statistical Applications

A number of univariate and multivariate statistical techniques, in addition to those incorporated into econometric work discussed separately, find frequent application in econometric work. These include:

- *Significance tests*, such as t-tests and analysis of variance (ANOVA) or chi-square tests for nominal data. These are often applied to test the significance of observed differences in performance between 'active' groups, such as firms participating in a publicly-supported R&D programme, and control groups (non-participants). The techniques, based on comparisons of variations *between* the groups compared with variations *within* them, allow precise statistical statements to be made regarding the degree of confidence (e.g. 95%) that one can have that an observed difference in (for example) average productivity growth between the groups did not occur by chance, and hence the probability that the programme was (at least to some degree) effective.
- More generally, *hypothesis tests* seek to determine whether data indicate that a particular hypothesis is supported at a predetermined level of confidence. A test statistic (such as t) is chosen, its value calculated from the data, and the calculated value compared with the critical value obtained for the chosen confidence level. Hypothesis testing has been used in evaluation of the US ATP programme[118] to investigate, among other things, the extent to which an ATP award affects the level of spillovers among members of a joint venture (as measured by cross-citation of patents between firms) .
- Use of *statistical distributions* can be helpful in deriving 'best estimates' of the overall benefits from a multi-project research programme. This type of enquiry is based on an analysis of the benefits obtained from a restricted sample of the constituent projects which may have been subject to in-depth case studies. Care is needed here, particularly in view of the skewed distribution of benefits typical of R&D programmes (see A.5.1).
- *Multivariate analysis* such as cluster or principal components analysis can be used to examine relationships between variables in a large dataset, such as may be obtained from a questionnaire survey of programme participants. Such an analysis can be used as a way of simplifying and clarifying information from a complex dataset, by extracting its underlying structure (see A.5.2).

A.5.1. Aggregating or 'Grossing Up' from a Sample

In evaluating R&D programmes of perhaps a few dozen – or a few hundred – individual research projects, it is rarely practical to carry out in-depth assessments of each project. It is usually appropriate for a programme evaluation to include detailed studies of a few cases and the question arises of how 'typical' they are, and whether outputs, outcomes and impacts for the whole programme can be estimated from the sample.

The key characteristic of the spectrum of project results is that they are typically highly skewed. One, or a few, projects typically realise the vast majority of the total benefits, with many projects yielding little or nothing. In view of this, it is unwise to carry out in-depth analysis of a random sample of projects and pro-rate the results across the programme, because conclusions are likely to represent severe over- or underestimates, depending on whether or not the most successful project(s) are included in the sample.

[118] David C. Mowery, Joanne E. Oxley, and Brian S. Silverman (1998), *The Role of Knowledge* Spillovers in ATP Consortia, Draft report, ATP

One approach is to case-study those projects which appear to be the most successful (these can often be clearly identified) and obtain a 'lower-bound' estimate for the programme as a whole by assuming that these are the only projects to contribute. If benefits from these projects exceed total programme costs, this may provide a useful means of demonstrating overall success. An 'upper bound' for the programme may be obtained by assuming that these projects are typical, but this is likely to be unrealistically overoptimistic.

An intermediate estimate can be derived by assuming that the range of project benefits follows a particular (skewed) distribution – possibilities include the Pareto and log-normal distributions. The former has been found empirically to provide a reasonably good fit to data in a number of areas, including the distribution of city sizes and distribution of incomes. A popular version of the Pareto distribution is the assertion that roughly 80% of effects or outputs arise from 20% of the inputs or causes. [119][120]

Regarding R&D-related issues, there unfortunately appears to be no theoretical underpinning for any particular distribution of outcomes. Empirically, Scherer[121] reports on the distributions of profits from US patents, returns from marketed pharmaceuticals, and stock-market returns from high-technology venture start-ups, and concludes that some distributions are close to log-normal, while others (notably the survey of patents) conform 'tolerably well' to the Pareto distribution. Scherer and Harhoff[122] consider the size distribution of returns from eight sets of data on inventions and innovations, and find that the top 10% of sample members captured from 48% to 93% of total sample returns, illustrating the high degree of skewness. The skewness might be even more pronounced in the case of a set of research projects, some of which will fail to reach the stage of an invention or innovation.

Any distributional assumption used in 'grossing up' needs to be applied with care and to be accompanied with appropriate caveats, but the empirical evidence does suggest that 'rules of thumb' can prove useful. It seems reasonable to assume, for example, that the top 10% of projects will account for at least 50% of the outputs, and the top 20% at least 80%. Judgement based on evidence from the programme in question might allow more specific estimates in particular evaluation exercises.

A.5.2. Extraction of 'Deep Structure'

The UK Alvey programme was widely regarded as one of the first stereotypical Advanced Technology Programmes, comprising the key attributes of

- cost-sharing between industry and government (with industry typically providing 50% or so of the total finance)
- involving pre-competitive, 'basic', 'enabling' or 'generic' technologies, i.e. long-term developments too far removed from the market, and too general, to be of direct immediate value in enhancing the competitive position of individual firms
- high-risk projects, with potentially high, but uncertain, returns
- Collaboration between firms, and/or between firms and academic institutions, or other research groups, to encourage the sharing of costs, risks and expertise.

[119] A United Nations study found in 1992 that 20% of the world's population controlled 82.7% of the world's income

[120] The general form of the Pareto distribution is $N=kV^{-\alpha}$, where N is the number of cases with a value of V or more, and k and α are positive constants. With α=1 this becomes 'Zipf's rule'; the largest output is twice the second largest, three times the third largest, etc.

[121] Frederic Scherer, 'The size distribution of profits from innovation' in Encaoua, D. et al., 'The Economics and Econometrics of Innovation', Springer, 2000

[122] Frederic Scherer, F.M., and Dietmar Harhoff (2000), 'Technology policy in a world of skew-distributed outcomes' *Research Policy*, 29, 4-5

As part of the programme evaluation process, a questionnaire was used to survey programme participants on various aspects of the research undertaken (a very similar questionnaire was subsequently used for evaluations of Advanced Technology Programmes run by the EU and by various European states, see **¡Error! No se encuentra el origen de la referencia.**). One section of the questionnaire covered perceptions of the ways is which those surveyed felt that participation was valuable to their organisation. Respondents were asked to rate 25 potential motives and goals on a 1-5 scale of importance – the 25 are listed in the table below, in order of average score obtained (scores as shown in the third column).[123]

A principal components analysis was carried out on the raw data, by which the set of 25 (correlated) variables were reduced to six uncorrelated variables (the principal components, or 'factors') which capture sets of the original variables which have common statistical properties. The final six columns of the table show which of the original variables contribute most heavily to each of the factors; the key to the symbols is

*the factor captures more than 40% of the variation in the original variable

**the factor captures more than 60% of the variation in the original variable

*** the factor captures more than 80% of the variation in the original variable

Figure 39 Principal component analysis

Rank	Activity	Importance Rating	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	Develop new tools and techniques	4.18	*					
2	Accelerate R&D	4.14				**		
3	Build on R&D base	3.99				***		
4	Maintain R&D presence	3.91				**		
5	Deepen understanding	3.86				**		
6	Enhance image	3.83	**					
7	Establish new academic-industry links	3.80	*					
8	Upgrade skills	3.77		*	*			
9	Enter new R&D area	3.74	*	*				
10	Use new tools and techniques	3.72	*		*			
11	Develop new prototypes	3.71						**
12	Enter international collaborative R&D programmes	3.53	**					
13	Enter private sector R&D ventures	3.40	*					*
14	Achieve critical mass	3.36		*				
15	Establish new industry-industry (or academic-academic) links	3.30	*				*	
16	Access industry know-how	3.26	*				**	
17	Access academic know-how	3.22			*			
18	Enter other national R&D programmes	3.20	**					

[123] Ken Guy, John Clark and James Stroyan (1998), 'The Nature of Advanced Technology Programmes', Report to OECD

19	Spread costs	3.18			***			
20	Keep track of peripheral R&D	2.82		**				
21	Spread risks	2.79			***			
22	Enter new non-R&D collaborations	2.76						**
23	Develop new products	2.76						***
24	Use new standards	2.45			*		*	
25	Influence new standards	2.27					**	

The following interpretations can be placed on the six factors.

- Factor 1 groups together entry into follow-on R&D programmes, enhanced image and reputation, the establishment of new collaborative links, entry into new R&D areas and the development and use of new tools and techniques. It reflects an expansion of opportunity for an organisation, broadening its horizons through further collaboration and new areas of research. Interestingly, the perception of *overall benefits* is correlated more closely with this factor than with any other
- Factor 2 also captures entry into new R&D areas, but this time links it with tracking developments in peripheral R&D areas, establishing a critical R&D mass and upgrading skills. Whereas Factor 1 describes a positive and active grasping of new opportunities, Factor 2 reflects a more cautious or defensive exploration of alternative technical possibilities.
- Factor 3 is strongly influenced by risk and cost reduction. It is also correlated with accessing know-how and technology from academic organisations, upgrading skills and familiarity with tools and standards. Overall, it seems to represent the quest for security, a strategy to enhance competitiveness based on ensuring state-of-the-art know-how in core areas and cutting costs by minimising 'wasteful' expenditures
- Factor 4 links the expansion and maintenance of existing areas of R&D with acceleration and deepening. It represents a classic strategy of enhancing the knowledge base of an organisation, and there is an association with benefits
- Factor 5 associates the establishment of industry-industry links, and accessing technology from industrial organisations, with gaining familiarity with, and influencing, standards. It represents a strong industrial networking factor
- Factor 6 links the development of new products and prototypes with entry into private sector R&D collaborations and non-R&D-based collaborative ventures. It represents an overt, commercially-oriented industrial exploitation factor.

A.5.3. References

Ken Guy, John Clark and James Stroyan (1998), 'The Nature of Advanced Technology Programmes', Report to OECD

David C. Mowery, Joanne E. Oxley, and Brian S. Silverman (1998), *The Role of Knowledge Spillovers in ATP Consortia,* Draft report, ATP,

Frederic Scherer (2000), 'The size distribution of profits from innovation' in Encaoua, D. et al. (2000), *The Economics and Econometrics of Innovation*, Springer,

Frederic Scherer, F.M., and Dietmar Harhoff (2000), 'Technology policy in a world of skew-distributed outcomes', *Research Policy*, 29, 4-5

A.6. Analysis of 'Spillovers'

A.6.1. Description

Arguments justifying state support for R&D are normally based on the belief or observation that social rates of return to R&D exceed private rates. As a result, society benefits from private R&D expenditure; firms carrying out R&D cannot appropriate all the benefits from their investment in it, and hence tend to invest at a suboptimal level. Some of the benefits 'spill over' to other firms and to society at large.[124] And, as Jaffe[125] points out, to the extent that the policy justification for state-supported R&D programmes lies precisely in creating such spillovers, any attempt to evaluate the success or failure of such programmes needs to attempt to measure the spillovers.

A.6.1.1. Categories of Spillover

Two categorisations of spillovers are commonly employed in the literature. One of these, based on the economic relationship of the beneficiary of the spillover to the innovator, is between *horizontal*, *vertical* and *international* spillovers:

- *Horizontal* spillovers are those enjoyed by other firms in the same industry, whether direct competitors of the innovator or not, who learn about the research or development work and benefit from it without having to incur development costs. The channels through which these horizontal spillovers typically occur are imitation, transfer of knowledgeable individuals, and spin-offs
- *Vertical* spillovers are benefits accruing to firms and other economic agents not in the same sector, principally those in upstream (supplier) or downstream (user) industries. Suppliers may benefit from new markets, and/or induced improvements (and higher prices) for their products. More commonly, users may find that the costs of their purchases has declined, and/or their quality has improved – a familiar example is personal computers, showing quite spectacular cost reduction *and* quality improvement over the last couple of decades
- *International* spillovers have been increasingly recognised in recent years, particularly in the econometric literature, where clear positive relationships have been established between, for example, measures of productivity growth and the extent of R&D in other countries. As with horizontal spillovers, however, the degree to which international spillovers occur will depend on the 'absorptive capacity' of the potential beneficiary – the ability to understand and adapt new knowledge will itself require some prior relevant skills and expertise.

The second classification is based on the type and mode of transmission of the spillover; the key distinction is between knowledge spillovers and market (or rent) spillovers, to which Jaffe adds the category of network spillovers. These categories have the following characteristics:

Knowledge spillovers refer to avenues through which the knowledge generated within research and development processes (broadly rather than narrowly defined) is transferred in the form of information and knowledge consequentially accruing elsewhere in the economy. They may be transmitted by:

- Publications

---

[124] Spillovers would *not* occur if R&D leading to increased product quality, or increased process efficiency, was virtually fully captured in the price faced by the user, who would be almost indifferent between the new product/process and those previously available – the advantage to them of adopting the innovation would be positive but marginal. Such might be the case with a monopolistic producer able to capture all financial benefits, and also able to protect the intellectually property involved, so that others could not access or make use of the technology in any way. But cases where this approximates to the actual situation appear to be very unusual.

[125] Adam Jaffe (1996), *Economic Analysis of Research Spillovers: Implications for the Advanced Technology Program*, NIST GCR 97-708

- Patents, which give protection but also impart knowledge, potentially facilitating new applications
- Movement of researchers with unique or specialist knowledge between organisations and sectors
- Know-how embodied in products, within which new technical ideas are at least partially revealed – the very process of commercial exploitation promotes disclosure

In short, knowledge spillovers capture know-how transmitted to others mainly through four 'Ps': publications, patents, people and products.

Figure 40 Factors influencing spillover potential

Factors increasing market spillover potential	**Factors reducing market spillover potential**
• competitive markets in user industries • short lead times/learning curves • IPR protection difficult or limited • licensing of technology to others important • many potential uses of innovation • co-specialised assets for some applications are not possessed by consortium members	• consortium members have monopolistic or extensive power in relevant markets • lead times convey market power to consortium • other firms will lose markets through their technologies being rendered obsolete
Factors increasing knowledge spillover potential	**Factors reducing knowledge spillover potential**
• generic or path-breaking technologies • 'proof of concept', or development of key component, would facilitate related R&D by others • generation of knowledge distinct from technical objectives of project	• output is process innovation that can remain undisclosed • consortium members have expertise providing important advantages for the development of follow-on technologies
Factors increasing network spillover potential	**Factors reducing network spillover potential**
• technology includes attributes of a 'standard' • output is a product suitable for selling/licensing to other researchers • project success depends on success of a 'critical mass' of sub-projects (with public support providing means of co-ordination)	• output is product innovation that can be securely patented, kept within consortium

Market spillovers are characterised by the realisation of benefits to non-innovating firms *through the operation of market forces*. In the case of the launch of an improved *product*, for example, prices do not fully capture the extent of the improvement, so there is an additional increase in social welfare accruing to consumers. Computers are an obvious example.[126] Similarly, users of a *process* innovation benefit from reductions in their manufacturing costs – both these users and final consumers may benefit, relative benefits between these groups depending on the extent to which the cost reductions are passed on to consumers.[127]

Network spillovers arise where a new technology improves the prospects for other, related technologies, such as when mobile telephone systems 'enable' new connections to be made among previously unconnected producers and users. The principle of network externalities is that the larger the number of participants that join a network, the larger will be the per capita benefits (little is to be gained by having a telephone system with only one telephone). The benefits can also arise in knowledge as well as

[126] E.g. Griliches (1992)

[127] Which, in turn, will depend on the competitive pressures within the industry concerned

market terms; for example, a 'critical mass' of researchers in related fields may be needed in order for the research of any particular firm to be cost-effective.

Figure 40 is a summary of the factors identified by Jaffe as tending to increase or decrease the probability of significant market, knowledge and network spillovers.

Jaffe's enumeration does not include spillovers in the social and environmental arenas. Clearly, environmental/social spillovers can be negative – the rise of environmental movements in the 1960s and 1970s was largely prompted by the recognition of the negative effects of pollution from industrial production. Negative spillovers can also occur within the other categories. For example, while new suppliers may benefit from market spillovers resulting from an innovation, more traditional suppliers may be displaced, with overall net disbenefits in the short term.

Beneficiaries of spillovers include producers in the same industry as the innovator, and producers in other industries (knowledge, market and/or network spillovers); suppliers and users (mainly market spillovers); and the consumer and society at large (market spillovers, any economic growth generated by overall stimulus to aggregate demand, and social and environmental benefits). Figure 38 summarises the costs and benefits of publicly supported R&D to various stakeholders.

Overall, the single most important factor likely to lead to important spillovers is *significant commercialisation* of a product or process arising from an R&D programme.

Box 6 **Spillovers in the US Advanced Technology Programme**

Popkin adopted an input-output approach to assessing spillovers from the US Advanced Technology Program (ATP) – specifically, to identify sectors most likely to benefit from spillover effects. Key features of this approach, which is regarded as preliminary, are

- Identification of the industries within which new products or processes are created under the auspices of ATP, for a particular range of SICs (35-38 – Electronics, Computer Hardware and Communications)
- Identification of those user industries within the same broad range of SICs
- Four measures of inter-industry flows are constructed

5. Proportion of shipments from the producer industry consumed in the *same* industry, a measure of 'own industry' market and knowledge spillovers
6. Proportion of shipments from the producer industry consumed in various particular *user* industries
7. Purchases by the *user* industries from the producer industry as a percentage of those user industries' total intermediate purchases
8. Proportion of shipments from the *user* industry consumed in the *same* industry, a higher proportion being taken to indicate higher levels of diffusion of information – knowledge spillovers – within the user industry

- Weighted sums of the extent of types of spillovers are made for each ATP project areas are ranked according to their potential for producing spillovers.

On this basis, Popkin identifies areas with the greatest potential for spillovers, which in general are those where relatively high proportions of the purchases of the identified *user* industries are from industries within their own sectors and from the innovating industry.

Joel Popkin and company (2003), *Inter-industry diffusion of technology that results from ATP projects*, US National Institute for Standards and Technology, Dept. Of Commerce, NIST GCR 03-848

A.6.2. Uses

A.6.2.1. Input-Output Analysis – Horizontal and Vertical Spillovers

A quantitative procedure for estimating spillover effects – or spillover potential – is to make use of measures of the magnitude of inter-industry flows of goods and services, particularly those provided by input-output tables. The broad reasoning is that large volumes of sales from industry A to industry B increase the potential for firms within B to benefit from the innovative activity of industry A firms. Market spillovers in particular are likely to be involved, with, for example, B firms able to benefit from increased production efficiencies from utilising a new product from an A firm, the corresponding reductions in production costs being less than fully offset by the costs to B firms of the product (for an example see Box 6).

A.6.2.2. Econometric Approaches

Econometric treatment of spillovers generally involves extension of the 'basic' production function – where output is expressed as a function of labour and capital inputs - to include a stock of R&D capital and a spillover variable as additional determinants of output or productivity. The spillover variable for a particular firm is generally expressed as a weighted sum of the R&D stocks of other firms (in the same industry for 'horizontal' spillovers, and in other industries for 'vertical' spillovers); or, in the case of analyses of international spillovers, the spillover variable of a particular country is represented as a weighted sum of other countries' R&D capital stocks.

The weights used in the weighted sum of the 'external' R&D capital stocks are designed to reflect the 'closeness' of the external firm or country to the beneficiary of the spillovers. A particular example is the use of input-output coefficients as weights to measure other- or own-sector spillover potential, as discussed above – the subsequent econometric analysis will hopefully turn the 'potential' into robust measures of actual spillovers.

More generally, there are a range of spillover variables that can be specified, corresponding to the (wide) range of spillover mechanisms that have been identified by various authors. Cincera and Van Pottelsberge[128] have carried out a survey, with particular reference to international spillovers, which is summarised in (Box 7).

A.6.3. Strengths

Spillover studies are important in that the notion of spillover benefits is central to the justification for the use of taxpayer's money to support business R&D. To the extent that the studies have shown an overwhelming consensus in suggesting that such benefits are substantial, they have done considerable service in supporting the application of public policy in R&D.

Conceptually, spillover studies have yielded important insights into modes of transmission of technological advances, and contributed significantly to understanding in this area.

A.6.4. Weaknesses

There is considerable ambiguity in the way that 'spillover variables' might be defined, all the proposed specifications are to some degree proxies for the effects that one is trying to capture. 'Technological proximity', for example, is a rather nebulous concept which defies precise specification. In practice, some studies have been found to produce unstable results.

[128] Michele Cincera and Bruno van Pottelsberge, 'International R&D Spillovers: A Survey', mimeo.

Box 7 **International R&D spillovers**

The authors identify three kinds of market spillovers:

- Input-related, associated with imports of intermediate goods
- Investment-related, corresponding to trade in capital goods
- Patent-related, concerning the use by one country of patents granted in another

In addition, there are potential knowledge spillovers, facilitated by the technological proximity of firms in the two countries. There are various possible mechanisms for knowledge spillovers, which might be associated or combined with market spillovers.

Spillover variables are constructed for each of these four mechanisms. In each case the variable for the (spillover beneficiary) country (j) is a weighted sum of the R&D capital stocks of the R&D-performing country/firm (denoted by i, over which the sum is performed), with weights as follows :

Spillover type and corresponding weight

- **Input-related:** (j's imports from i)/(i's output)
- **Investment-related:** (j's imports of capital goods from i)/(i's output)
- **Patent-related:** The share of i's patents likely to be used by j
- **Knowledge spillovers:** Indicator of i's technological proximity to j

A number of alternative measures of technological proximity have been used – see e.g. Mohnen (1996).

Cincera and Van Pottelsberge consider around 20 studies of international spillovers, which overwhelmingly conclude that these are significant – and the gains often exceed those from 'own' R&D. It should, however, be noted that the two are not independent – the authors also cite evidence that a high level of domestic R&D enables greater exploitation of foreign R&D, suggesting the importance of 'absorptive capacity' and the need for R&D expenditure to fully benefit from international spillovers.

Michele Cincera and Bruno van Pottelsberge, 'International R&D Spillovers: A Survey', mimeo.

Pierre Mohnen (2002), 'R&D Externalities and Productivity Growth', STI Review, 18, OECD, Paris, pp.39-66

A.6.5. References

Michele Cincera and Bruno van Pottelsberge, 'International R&D Spillovers: A Survey', mimeo.

Adam Jaffe (1996), 'Economic Analysis of Research Spillovers: Implications for the Advanced Technology Program', NIST GCR 97-708

Joel Popkin and company (2003), 'Inter-industry diffusion of technology that results from ATP projects', US National Institute for Standards and Technology, Dept. Of Commerce, NIST GCR 03-848

Pierre Mohnen (2002), 'R&D Externalities and Productivity Growth', STI Review, 18, OECD, Paris, pp.39-66

A.7. Surveys of Beneficiaries

A.7.1. Description

In practice, a great deal of R&D evaluation relies on surveys of beneficiaries. Surveys often are key inputs to micro-economic models and control group approaches. Hence, surveys are more of a data collection tool than an evaluation method per se.

However, they also produce interesting information in their own right. They give information about motivations for participating in interventions, beneficiaries' self-evaluations of their own achievements and their accounts of outputs and results (including behavioural changes as a result of programme participation – so-called 'behavioural additionality'). Survey results are typically presented in aggregate, without identifying individual results, using tabular and graphical summaries of data.

Surveys ask multiple parties a uniform set of questions about activities, plans, relationships, accomplishments, value, or other topics, which can be statistically analysed. Surveys generally use a series of precisely worded, close-ended questions but can also include a few open-ended questions. Computation of survey statistics requires consistency across units of analysis in terms of questions asked or ranking systems or scales used. Surveys are typically web-based or by telephone. Mailed questionnaires have become rarer with the advent of the Internet but are still used e.g. in the Community Innovation Survey (see box below).

Statistical inference, the process of using sample data to make inferences about the parameters of a population, reduces the time and cost of collecting data by survey from an entire population. Sample design should be sufficiently described to enable calculation of sampling errors. Establishing a sampling frame—the list from which a sample is drawn—is essential. Samples may be randomized or stratified. Often, however, all participants in a programme are surveyed (census), in order to obtain a high enough response rate. The marginal cost of surveying another participant is low if a web-based survey is used. Sometimes non-participants – often unsuccessful applicants – are surveyed as well as a control group.

In contrast, interviews typically use more open-ended questions and discussion (so-called semi-structured interviews), leading to more varied data to which statistical analysis cannot be applied. Also, in many cases purposeful sampling rather than random sampling is applied. In other words, interviews are a typical data-gathering instrument used in qualitative research. In general, interviews are face-to-face or possibly by telephone (see also section A.9 on case studies).

Box 8 **The role of innovation surveys in R&D evaluation**

Firm-level innovation surveys have been conducted regularly in most OECD countries since the beginning of the 1990s. In the EU, EUROSTAT has taken the lead role and co-ordinated national endeavours to improve innovation measurement by launching the Community Innovation Surveys (CIS). So far four harmonised innovation surveys have been conducted, comprising more or less all member states as well as some non-EU countries (e.g. Norway, Switzerland). After almost 20 years of development, the Community Innovation Surveys are an established tool for addressing information needs of policy makers and policy analysts.

At present most innovation surveys are based on the OECD's so-called OSLO Manual. It has been highly influenced by Schumpeterian ideas and the chain-link model of innovation, which views innovation as the result of market opportunities, appropriability conditions, technological opportunities and the firm's internal capabilities base and external linkages. The 1997 second edition recognised the increasing importance of innovation in services, dealing explicitly with collecting data on innovation in services. The third edition, published in 2005, takes into account the field of non-technological innovation and the linkages between different innovation types. It also includes an annex on the implementation of innovation surveys in developing countries.

Key issues on which data are collected regularly in the CIS comprise

- Firms' innovation activities and expenditure
- Sources of information and co-operation for innovation
- Characteristics of innovating firms
- Constraints on innovation

Innovation surveys can be used in a variety of ways for the evaluation of innovation policies. However, this presupposes that the innovation survey contains questions on whether or not a firm took part in certain government programmes. This was the case, for example, in the third CIS which distinguishes between policy interventions at various levels of government (EU, Member States, regional/local).

Possible uses of innovation surveys in R&D evaluation include:

Analysis of the structure of participants: data from innovation surveys allow to uncover the structure and characteristics of participants in government interventions. For example, they can provide answers to the question as to who profits from government R&D subsidies

International comparisons: Due to the internationalisation of the core questionnaire, CIS data can be used to make international comparisons of the structure of participants in innovation programmes. The data gives insights into the relationship between national support structures and innovation activities of firms, e.g. which government support measures attract SME participation

Econometric analyses: More sophisticated and often more interesting insights into the impact of government support can be gained if econometric methods are applied. Large scale innovation surveys address not only innovative firms but also non-innovative firms, allowing comparisons between the two groups. What is more, innovation surveys cover firms with and without government innovation subsidies, allowing a control group approach and the examination of spillovers. Finally, as innovation surveys are not conducted to evaluate a specific programme, strategic answering by firms tends to be less likely.

To summarise, innovation survey data can be used to evaluate different types of government intervention, such as R&D subsidies and tax credits, specific project-based programmes and programmes to foster science-industry linkages. However, most existing innovation surveys are not specifically designed as an evaluation tool but rather as a tool for providing information about the structure of national systems of innovation.

Georg Licht and Giorgio Sirilli, Innovation Surveys, Gustavo Fahrenkrog et al. (eds.), Gustavo Fahrenkrog et al. (eds.), RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies, Sevilla, 2002, pp. 71-81

OECD and Eurostat, OSLO Manual. Proposed Guidelines for Collecting and Interpreting Technological Innovation Data, 3rd edition, 2005

A.7.2. Uses

Surveys of beneficiaries are a widely used instrument. It is particularly suitable for evaluating government instruments that finance R&D or promote technology transfer and science – industry linkages. Beneficiaries typically are firms and public research institutes but can include any target audience of an intervention. Beneficiaries cannot only be surveyed by independent evaluators but also by research councils and innovation agencies for monitoring purposes (see Box 9).

Box 9 **Using surveys to measure economic benefits – practical examples**

The approach adopted by the Research Council of Norway to calculate the return on its investment in 'user-driven R&D' programmes is among the simplest. It involves longitudinal telephone surveys of company beneficiaries. The so-called Beta Method developed at the University of Strasbourg extends this approach by surveying and attempting to value other non-monetary benefits of participation. This results in the addition of incommensurable values and double counting. These techniques may provide useful – but partial – estimates of benefits (results) to beneficiaries but do not tackle the wider impacts, which justify state funding of the intervention.

Some organisations (TEKES or the now defunct Advanced Technology Programme of NIST) maintain running estimates of the economic benefits of individual projects in cooperation with project managers as part of routine monitoring and for a period after the end of the project. Firms supported by TEKES report estimates about the impact of the project on sales, exports, new and safeguarded jobs, which are then used to calculate estimates of the economic benefits of the project. Quality assuring and updating such claims, as well as attribution, are important problems. However, the data collection exercise is probably itself useful in focusing attention on project progress, and tends to alert the funder to the 'skewed' success cases that can provide strong anecdotal support to funding the programme.

Arild Hervik, 'Evaluation of user-oriented research in Norway: the estimation of long-run economic impacts,' in OECD, 1997

Bach, L., M.-J. Ledoux, M. Matt, Evaluation of Large Scale Programs in Europe; Bach, L. et al. 'Evaluation of the economic effects of Brite-Euram Programmes on the European industry' *Scientometrics*, vol. 34, n. 3, p. 325-349, 1995

Patries Boekholt, Maureen Lankhuizen, Erik Arnold, John Clark, Jari Kuusisto, Bas de Laat, Paul Simmonds, Susan Cozzens, Gordon Kingsley and Ron Johnston (2001), *An international review of methods to measure relative effectiveness of technology policy instruments*, report to EZ, Technopolis: Amsterdam

Box 10 **The 'Alvey questionnaire'**

A key technique used in the evaluations of the Framework Programme since the end of the 1990s and up to 2004/5 is sending questionnaires to project participants, normally using a variant of the so-called Alvey questionnaire. The Alvey questionnaire was developed in the early 1980s by the Science and Policy Research Unit (SPRU) at the University of Sussex and the Centre for Science and Technology Policy and Business Research PREST at the University of Manchester during the evaluation of the Alvey Programme for Advanced Industrial IT Research in the UK. Subsequently, this questionnaire has become pervasive in European R&D programme evaluations at national and international levels.

Typical elements of the questionnaire are:

- Participation in the programme

- Was this the first participation?

- Was the project co-ordinated or linked with other projects within the company or organisation?

- Did the project contain partners with whom the participants had (not) worked before?

- What was the primary focus of the project (basic research, applied research etc.)? Characteristics of the project?

- Motives, goals, outcomes and achievements

- Knowledge-oriented goals and achievements

- Network-oriented goals and achievements

- Exploitation-oriented goals and achievements

- Strategic management goals and achievements

- Outputs and impacts

- Importance of different outputs in assessing the success of project participation?

- Will the outputs and results of the project be demonstrated, developed further, and used regularly by the organisation, by partners in the project or by other organisations?

- Who will be the subsequent users of the results?

- Commercial returns for the organisation? Plans for the future commercial exploitation of project results?

- Expected impacts of project over the next ten years at regional, national and European level?

- What would have been the impact on the project if the organisation had not received EU funding support?

- How do costs and benefits associated with the project participation balance out?

- Procedures

- Ease of application? Clarity of documentation? Speed with which application is processed?

- Input received from European Commission officials? From National Delegates?

- Satisfaction with different administrative arrangements (project payments, project reporting, programme evaluation)?

The questionnaire is generally complemented by the use of other methods in order to permit triangulation. Control groups are rarely used (Austria and Denmark are exceptions) – often because of budget pressure on the evaluations. It is noteworthy that the studies do not make use of baselines – indeed baselines are not drawn up, though this is admittedly difficult to do in the context of R&D programmes. This means there is no 'before' picture to compare with the situation 'after' the intervention. Overall, the picture is of rather homogenous methods – a homogeneity that probably ought to decline as evaluations tackle higher-level, more systemic and more policy-oriented questions in future (see also Box 27).

Ken Guy, Luke Georghiou, Paul Quintas, Hugh Cameron, Michael Hobday and Tim Ray (1991), *Evaluation of the Alvey Programme for Advanced Information Technology*, London: HMSO

A.7.3. Strengths

A strength of using survey-based descriptive statistics in evaluation is that it provides an economical way to gather aggregate level information about a programme and its participants, even in its early stages, and it accommodates the use of control and comparison groups or the collection of counterfactual information (see Box 11). Moreover, surveys provide representative data provided that a random sample has been drawn or a census has been conducted.

Other strengths are that diverse audiences can usually understand the approach and results, and many people find statistical results credible and informative.

Furthermore, once collected, survey data can be analysed and reanalysed in different ways. Surveys can provide information about participants and users not available through other sources. Surveys are also useful because they provide the figures and the high numbers of cases necessary for statistics to be applied.

A.7.4. Weaknesses

A disadvantage of survey statistics is that they do not convey the richness of individual projects that stakeholders tend to find interesting and memorable. A further limitation is that the responses on which descriptive statistics are based are often subjective in nature. Also, respondents may not be truthful, or they may have faulty recall.

Responses have a systematic positive bias – whether respondents are asked about future cash flows,[129] scientific and technological performance[130] or additionality[131] – but the fact that the bias is systematic makes it possible to interpret survey findings.

Box 11 **Counterfactual questions and comparison group**

One survey design approach sometimes used to provide evidence of impact is a non-experimental design with counterfactual questions to learn from participants what they think would have happened had the government programme not existed. Counterfactual questions typically entail asking survey respondents several hypothetical questions: Would they have proceeded with their research project had they not received government funding for it? If they had proceeded without government funding, would the project have been different?

These questions were asked for the first 50 completed projects funded by the Advanced Technology Program (ATP). 59% of ATP grantees responded that they would not have proceeded with the project, while 41% would have proceeded but with a delay of between 6 months and 5 years. The results were used as one line of evidence of programme impact.

Another ATP survey used a quasi-experimental design with a comparison group. This approach was used to test if unsuccessful applicants of ATP grants continued with their proposed research projects, and if so, what was changed from the plan submitted to ATP. In other words, rather than asking the counterfactual questions the survey asked a group of unsuccessful applicants what they actually did.

The survey was conducted a year after applicants failed to receive the grant, while the experience was still relatively fresh and could more likely be recalled by respondent. The survey found that more than 62% did not proceed with the project, while 29% began the project on a much smaller scale (17%) or somewhat smaller scale (12%) than proposed. The rest began the project at about the same or larger scale than proposed. This was another piece of evidence of the additionality of ATP.

NIST (ed) (2001), *Performance of 50 Completed ATP Projects*. Status Report - Number 2, NIST Special Publication 950-2, U.S. Government Printing Office, Washington, D.C., December

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

A.7.5. Quality Control

Limitations – such as failure to adequately reflect the target population or biases in the results – typically result from weaknesses in survey design. To help eliminate them, detailed review of all aspects of a survey's design, together with pilot testing – i.e. administering the proposed survey to a small sample of people who are the same as those who would be included in a full-scale survey – are advised. Low response rates may limit the reliability of results and may require extra steps to increase responses. In short, the usual methodological quality criteria of survey research apply[132].

[129] Lasse Bræin, Arild Hervik and Bjørn G Bergen (2001), *Brukerstyrte Prosjekter i Norges Forskningsråd 1999: Portføljeanalyse (PROVIS), trendanalyse av nye prosjekter 1995-99 og undersøkelse av et utvalg avsluttede prosjekter 1999*, Molde: Møreforsk

[130] Erik Arnold, Bavid Bannister, Claude Lamure, Lee Schipper, Ben Thuriaux (1996), *An Evaluation of KFB's Energy and Environment Programme 1990-1996*, report to the Swedish National Board for Transport and Communications (KFB), Brighton: Technopolis,

[131] Andreas Schibany, Gerhard Streicher, Nikolaus Gretzmacher, Martin Falk, Rahel Falk, Norbert Knoll, Gerhard Schwarz and Martin Wörter (2004), *Evaluation of FFF: Impact Analysis* (Background Report 3.2), Vienna: Joanneum Research

[132] See for example F.J. Fowler (2008), *Survey Research Methods*, 4th edition, Sage

A.7.6. References

Patries Boekholt, Maureen Lankhuizen, Erik Arnold, John Clark, Jari Kuusisto, Bas de Laat, Paul Simmonds, Susan Cozzens, Gordon Kingsley and Ron Johnston (2001), *An international review of methods to measure relative effectiveness of technology policy instruments*, report to EZ, Technopolis: Amsterdam

Bach, L., M.-J. Ledoux, M. Matt (1995) Evaluation of Large Scale Programs in Europe; Bach, L. et al. 'Evaluation of the economic effects of Brite-Euram Programmes on the European industry' *Scientometrics*, vol. 34, n. 3, p. 325-349

J.R. Fowler (2008), Survey research methodology, 4th edition, Sage

Luke Georghiou, John Rigby and Hugh Cameron (eds) (2002), *Assessing the Socio-Economic Impacts of the Framework Programme*, Manchester: PREST, June

Ken Guy, Luke Georghiou, Paul Quintas, Hugh Cameron, Michael Hobday and Tim Ray (1991), *Evaluation of the Alvey Programme for Advanced Information Technology*, London: HMSO

Arild Hervik (1997), 'Evaluation of user-oriented research in Norway: the estimation of long-run economic impacts,' in OECD, in OECD (1997)

Georg Licht and Giorgio Sirilli (2002), Innovation Surveys, in Gustavo Fahrenkrog et al. (eds.) (2002), RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies, Sevilla, , pp. 71-81

NIST (ed) (2001), Performance of 50 Completed ATP Projects. Status Report - Number 2, NIST Special Publication 950-2, U.S. Government Printing Office, Washington, D.C., December

OECD and Eurostat (2005), *Oslo Manual. Proposed Guidelines for Collecting and Interpreting Technological Innovation Data*, 3rd edition, Paris

Rosalie Ruegg and Irwin Feller (2003), *A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade*, NIST GCR 03-857, pp. 29-34

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes*, A Directory of Evaluation Methods Relevant to Technology Development Programs, U.S. Department of Energy, March

A.8. Peer and Panel Reviews

A.8.1. Description

Peer and panel reviews, while often regarded as synonymous are in fact distinct forms of review processes with significant differences between them. Both methods, however, are based on the principle of using a group which combines the most relevant and expert knowledge and experience available – expert review – to make a judgement of worth and value. The experts generally render their verdict after reviewing written or orally presented evidence or making direct observations of activities and results.

Panel and peer review methods are used prospectively, i.e. ex ante, and retrospectively, i.e. ex post, but the typical use of ex ante peer review is for the allocation of funding to research grant applications and by editors to select and screen submitted manuscripts, while the most typical form of the Panel Review process is the ex post evaluation of R&D programmes.

Strictly, review by scientific peers is used as a component in proposal assessment and in evaluating the scientific value of results. However, in recent years peer review has become increasingly 'extended' in the evaluation of R&D with social purposes. This extension has involved both introducing non-scientists (such as professional evaluators and stakeholders) into panels and asking scientists to assess not only scientific quality and track record but 'relevance' criteria as well.

Box 12 **Example of a classical peer review**

In the U.S. Department of Energy's Hydrogen, Fuel Cells, and Infrastructure Technologies Program (HFCIT), research and other activities performed by industry, universities, and national laboratories are evaluated annually at the Hydrogen Program Merit Review and Peer Evaluation meeting. Independent expert panels review the project portfolio in accordance with set criteria, which helps guide the program managers in making funding decisions for the new fiscal year. This review of the HFCIT program is conducted using the process outlined in the DOE's Energy Efficiency and Renewable Energy Peer Review Guide. In addition to annual peer review at the project portfolio level, external reviews are conducted every two or three years by the National Academies (e.g. National Research Council, National Academy of Sciences) or an equivalent independent group. The program prepares a formal response to the review recommendations.

Rosalie Ruegg and Gretchen Jordan (2007), Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs, U.S. Department of Energy, March

The use of panels composed of peers as the key arbiters in ex post evaluations, as used by the European Commission and others, is called panel review and is an extreme extension of the principle of 'peer review'. The traditional, scientific use of peers is to make judgements about quality within the peers' areas of competence, while this type of extension to the peer review principle means that most of the experts' work in the panel is to tackle issues outside their areas of expertise.[133] Hence, panel review is most likely to give reliable results where general broad brush pictures of the quality of research are carried out and where evaluators and policy makers need a general picture of what impacts on socio-economic development have occurred.

(Expert) panel reviews rely on the ability and experience of a group of individuals as a good way to explore complex situations – often in situations where there are not

[133] For a useful discussion of this issue in the context of foresight, see Dennis Loveridge (2002), "Experts and Foresight: Review and experience" PREST Discussion Paper Series paper 02-09, Manchester University: PREST

adequately developed methods for evaluation. Panels may commission studies to generate evidence needed to support their enquiries.

A.8.2. Uses

Peer and panel review can be used in the following ways[134]

- To conduct in-progress reviews of scientific quality and productivity
- To help answer questions about the relevance, timeliness, riskiness and management of existing programme research activities, and resource sufficiency of new programme initiatives
- To help assess appropriateness of programme mechanisms, processes, and activities and how they might be strengthened
- To provide information to help programme managers make decisions to design or revise their programme, re-direct existing R&D funds, or allocate new funds
- To integrate across multiple evaluation results and render judgements about the overall success of a programme or initiative

Panels are dependant on sound and detailed information on which to base their judgements about a programme's progress or impact, and therefore are vulnerable to poor and insufficient information. The type of data needed for retrospective impact assessment cannot be created in a panel format. Indeed, panellists have found it relatively easy to comment on the technical quality of research but have often found themselves ill equipped to deal with its economic and social impacts. For this reason, panel reviews tend not to be appropriate for evaluating impacts of programmes – except if a peer review panel is provided substantial, reliable results from impact studies based on other methods, and serves the function of integrating multiple studies.

Box 13 **Examples of Successful Panel Reviews**

An example of a panel review of this type that is seen as successful can be found in Finland where senior researchers on research and innovation have played important roles. Like in the latest Five Year Assessment of the Framework Programmes, these had a number of professionally conducted background studies at their disposal. They also had a say in what the background studies would be.

Other examples exist where panel review is a freestanding component in a larger evaluation, for example of the Research Council of Norway, the Austrian Science Fund (FWF) and the Austrian Research Promotion Fund (FFG). This means that professional evaluators manage the panel process, maintaining focus on the evaluation questions in hand, while the panel retains the independence to reach its own conclusions. The other parts of the evaluations provide alternative sources of evidence, allowing triangulation of results.

A.8.3. Strengths

Peer review is a relatively low-cost, fast-to-apply, well-known, widely accepted, and versatile evaluation method that can be used to answer a variety of questions throughout the programme performance cycle, as well as in other applications.

Peer review is often the only method available when specialised knowledge is required. Clearly, when the quality of research has to be assessed, it can only be done by scientists active in comparable fields.

[134] Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

Panels, while methodologically problematic, can be useful devices where budget, time or methods constraints prevent more rigorous approaches and where political legitimation is important. As panel review is a dialectical process, in its ideal form, it has the potential to provide new ways of looking at programmes and new insights into programme performance.

Box 14 **OECD peer review of industrial and R&D policies**

The OECD has a long tradition of 'peer review' of industrial and R&D policies. This type of peer review can be described as the systematic examination and assessment of the performance of a State by other States, with the ultimate goal of helping the reviewed State improve its policy making, adopt best practices, and comply with established standards and principles. It tends to "create, through this reciprocal evaluation process, a system of mutual accountability. … The effectiveness of peer review relies on the influence and persuasion exercised by the peers during the process."

The performance of the reviewed State is assessed against principles, criteria and standards which may differ widely in character and scope. They may include:

- Policy recommendations and guidelines: […] E.g., in the peer reviews, or surveys, carried out by the Economic and Development Review Committee, country performance is assessed in relation to broad economic policy principles and best practices that have been development over the years, the policy orientations of the OECD Growth Project, as well as specific guidelines such as those contained in the OECD Jobs Strategy.
- Specific indicators and benchmarks: indicators and benchmarks provide specific and often numerical targets to achieve, and they are more susceptible than policy guidelines to being assessed according to quantitative measures.
- Legally binding principles: peer review can also be a mechanism to monitor compliance with international norms.

The OECD Review team typically comprises the OECD Secretariat, two countries as 'special' reviewers and consultants. The country under review submits a background report as initial input. Extensive desktop research and interviews with the major stakeholders in the National System of Innovation complement the review. A report is issued at the end of the process.

The choice of examiner countries/examiners is based on a system of rotation among member states, although the particular knowledge of a country relevant to the review may be taken into account. Their task includes the examination of documentation, participation in discussions with the reviewed country and the Secretariat, and a lead speaker role in the debate in the collective body. However, the most labour intensive part of the work is carried out by the Secretariat and the consultants, which may also have the most expertise in the substantive area of the review. Hence, one could say that OECD peer reviews are expert reviews rather than peer review, as the Secretariat appears to have the stronger role and position in the process.

This variant of OECD-type policy review has also been used in work associated with the European Commission's Open Method of Coordination.[135] A problem associated with this type of peer-based policy review is that it is not quite clear what standards countries are assessed against and what qualifies experts to be experts.

Isabelle Collins et al. (2007), Peer Review as a Policy Learning Tool? Presentation given at the UK Evaluation Society Conference, Leeds, 23 November

[135] The "Open Method of Coordination" (OMC) was introduced by the European Council of Lisbon in March 2000. It was a method designed to help Member States progress jointly in the reforms they needed to undertake in order to reach the Lisbon goals.

Fabrizio Pagani (2002), Peer Review: A Tool for Cooperation and Change. An Analysis of the OECD Working Method, SG/LEG(2002)1, Paris: OECD,

A.8.4. Weaknesses

The quality and credibility of peer evaluation is highly dependant on the peers selected and the evaluation questions and criteria used by them. Limitations set on the geographical area from which peers might come can give rise to bias in the judgement. It should also be borne in mind that the peers themselves are often in competition with those whose work they judge. Ideally, peers are very knowledgeable about the subject and free of conflict of interests that could bias their judgement. The sometimes-expressed view that peer review is an "old boys' club" must be avoided. Steps may be needed to calibrate reviewer ratings. Defining appropriate criteria may be problematic when the work being reviewed is highly innovative and/or interdisciplinary. Indeed, peer review has been shown to be conservative and leading to "institutionalised orthodoxy".[136]

In panel reviews, elements of political judgement may be important, and the status of the individuals involved tends to legitimate exercises that are methodologically problematic.[137] The panel approach is vulnerable to being exploited to promote disconnected political or policy agendas, as is argued was the case with the Five Year Assessment of the European Framework Programmes prior to 2002. "Many of the recommendations drew not so much on an evaluation of past Framework activities but on the collective opinions and assessments of the Panel members concerning the general structure and organisation of science, technology and innovation in the EU."[138]

While the more extreme conflicts of interest have been addressed in the composition of recent panels, it is in practice difficult to assemble panels that exclude people from major beneficiary organisations.

A.8.5. Quality Control

The selection of peers for peer and panel review is *the* central issue in quality control. The following is a list of principles of good practice based on a literature review[139]

- It is recommended that experts declare their interest – perhaps even publishing a bias statement and potential conflict of interest – to ensure that the panel's reputation for fairness is upheld. For example, the UK Research Assessment Exercise employs strict rules to ensure that potential and apparent conflicts of interest are avoided. Indeed, concerns that statements of financial interest should be made are rising, even for authors. The publication *Nature* took the step in mid-2001 of asking authors of papers to declare their financial interests as there is increasing evidence that failures to do so are bringing the process of peer review into dispute.
- Restricting either the number of evaluations on which panel members serve or the number of years in which they are active is likely to reduce bias and complacency.

[136] John Rigby (2004), Making Decisions about Science and Technology – between the Devil and the Deep Blue Sea? A brief reflection on Expert and Peer Review, *Plattform Forschungs- und Technologieevaluierung* No. 21, June, pp. 2-11

[137] Luke Georghiou (1995), 'Assessing the Framework Programmes – a meta-evaluation,' *Evaluation*, 1(2), pp171-188

[138] Luke Georghiou, John Rigby and Hugh Cameron (eds) (2002), *Assessing the Socio-Economic Impacts of the Framework Programme*, Manchester: PREST, June 2002, p. 180

[139] Gustavo Fahrenkrog et al. (eds.) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, Sevilla, , pp. 165-166

- Broadening the panel as much as possible without introducing those who do not have relevant skills is a means to ensure effective debate and discussion and the generation of new ideas.
- The number of panel members is a problematic area, with a major research effort in social and developmental psychology given to the effective functioning of groups of different sizes. No studies are reported in the literature of the relationship between group size and decisions made specifically within the context of panel review.
- Steps can be taken to publicise the area of expertise of a particular panel member, thereby protecting the credibility of the expert process itself and the assessments made by the panel, although openness about the suitability of particular individuals' expertises could also be used to question or even to undermine the judgements which the panel makes.
- It is suggested that those within any panel who have more specialist knowledge of an area should not be allowed to make decisions as to quality and value of proposals but to submit their views to their peers within the panel itself for a joint decision.
- The appointment of panel chairs should or could be from the membership of previous panels to ensure continuity.
- The panel chair and other experts should all have high reputations in the areas in which they are required to make judgements so as to instil confidence in those affected by the evaluation.
- Occasionally it may be necessary to seek experts or expert witnesses from outside the geographical area where the programme is being carried out.

A.8.6. References

Luke Georghiou (1995), 'Assessing the Framework Programmes – a meta-evaluation,' *Evaluation*, 1(2), pp. 171-188

Dennis Loveridge (2002), *Experts and Foresight: Review and experience* PREST Discussion Paper Series paper 02-09, Manchester University: PREST

Fabrizio Pagani (2002), *Peer Review: A Tool for Cooperation and Change. An Analysis of the OECD Working Method*, SG/LEG(2002)1, Paris: OECD

John Rigby (2002), Expert Panels and Peer Review, in Gustavo Fahrenkrog et al. (eds.) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, Sevilla, pp. 163-172

John Rigby (2004), Making Decisions about Science and Technology – between the Devil and the Deep Blue Sea? A brief reflection on Expert and Peer Review, *Plattform Forschungs- und Technologieevaluierung* No. 21, June, pp. 2-11

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

Michael Stampfer (2004), 'Some Developments in Peer Review', *Plattform Forschungs- und Technologieevaluierung*, No. 21, June 2004, pp. 12-16

A.9. Case studies

A.9.1. Description

Case studies are in-depth investigations into a programme, project, facility or phenomenon, usually to examine what happened, to describe the context in which it happened, to explore how and why, and to consider what would have happened otherwise. The case study approach uses multiple methods of data gathering and analysis, in which a range of respondent types are involved within a number of different applications settings. It entails the collection of both quantitative and qualitative data, including interviews, surveys, content analysis, statistical analysis of secondary data and observation, and the interpretative synthesis of these different data sources to provide an overall interpretation of each case.[140] In R&D evaluation, interviews as a data collection instrument typically have an important role, while observational techniques are used only rarely. The potential scope of case studies is broad – ranging from brief descriptive summaries to long complex treatments.

Case studies usually start with qualitative information from direct observation, programme or project documents, and interviews with key project managers. Programme and project documents are useful for establishing key dates, budgets, initial plans and goals, specific outputs, key staff, and other critical information helpful in framing a study. Interviews promise significant depth and understanding of effects which cannot be known in advance. They are particularly useful in the context of identifying new processes of impact. To extend the available information, the evaluator may bring in results from one or more of other evaluation methods, such as survey results or bibliometric results, to enhance the story.

There is no one-to-one match between case studies and particular policy instruments. They could be applied to financial incentives to firms, encouraging networks among research and development actors, dissemination initiatives etc. The key determinant of appropriateness are the characteristics of the policies and programmes being evaluated and the kinds of evaluation questions being asked.

How to select individual cases for an evaluation is a crucial question. Perhaps nothing better captures the difference between quantitative and qualitative methods than the different logics that underlie sampling approaches. Qualitative inquiry typically focuses in-depth on relatively small samples, even single cases, selected purposefully, while quantitative methods typically depend on larger samples selected randomly. The logic of purposeful sampling lies in selecting information-rich cases for study in-depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research.[141] For example, if the purpose of an evaluation is to increase the effectiveness of a programme in reaching non-innovating SMEs, one may learn a great deal more by focusing in-depth on understanding the needs, interests, and incentives of a small number of carefully selected SMEs than by gathering standardised information from a large, statistically representative sample of the whole programme.

A.9.2. Uses

Case studies have two principal uses in R&D evaluation:

---

140 Robert K. Yin (2009), *Case study research: Design and methods*. Fifth edition. Thousand Oaks, CA: Sage

141 There are several different strategies for purposefully selecting information-rich cases. The logic of each strategy serves a particular evaluation purpose, see Michael Quinn Patton (1990), *Qualitative Evaluation and Research Methods*, Second Edition, Sage: Newbury Park, London, New Delhi, pp 169-181

One is exploratory[142]. Case studies are particularly helpful in understanding general propositions, and in identifying key relationships and variables. At an early stage in an evaluation, case studies are a useful way to 'go fishing': understanding typical projects; looking for categories of unintended effects; testing programme logic against participants' decision-making processes.

Their second use is in reporting, to provide comprehensible illustrations of success or failure. Using a 'story-telling' approach, the evaluator may present the genesis of ideas, give an account of the human side of the project, explain goals, explore project dynamics, and present outcomes. Case studies enable the exploration of interventions from the perspective of the participant or beneficiary, and can also be used to construct theories about programme and project dynamics. Case studies can also be valuable in identifying exemplary or best-practice experiences. Used in formative evaluations, they can guide agency behaviour or serve as a benchmark for other programme recipients.

Multiple case studies may be conducted with uniform compilation of information across cases to provide aggregate statistics for a portfolio of projects. Indeed, it is becoming increasingly common to undertake comparative case studies that seek to locate examples in a particular setting or look for associations between different attributes of the case study concerned. In a similar vein, Kingsley and Klein[143] have shown how meta-evaluation of structured case studies can be used to build a more systematic understanding of the effects of interventions. Using case studies as raw material for what then amounts to a survey enables systematic exploration of qualitative source material.

Case studies are particularly suited for the identification of the following effects:

- Identification of products and services: When RTD expenditure leads to products or ranges of products, the most appropriate methods to deal with the issues of attribution are case studies and interviews. However, when research seeks to identify externalities / spillover from participants in research to those nominally outside, the techniques which should be used are those from cost-benefit analysis (see sectionA.4) and econometrics (see section A.6).
- Identification of capabilities: Case studies are appropriate for the identification and assessment of the capabilities which firms develop as a result of their participation in RTD activities directly or from the externalities which result from participation in them by other firms. These capabilities are often divided into two categories, capacities and absorptive capacities. Knowledge of these capabilities are normally derived from interview. Capabilities and competences are often implicit hidden and particular, and need therefore to be identified by questioning. Evidence of them is not normally available in conventional data sources. They are not always disclosed in company information sources and legal disclosures. As capabilities often vary from period to period, comparisons between different periods of time are difficult. Consequently, time series, and other parametric statistical techniques are inappropriate.
- Identification of network effects: Socio-economic impact within innovation process itself has increasingly been seen to be the major benefit of RTD policy and innovation policy. Innovation networks are complex, existing within sectors, regions, and internationally. Innovation policy increasingly seeks to promote such

[142] Many (social) scientists believe that case studies are appropriate only for the exploratory phase of an investigation and that they are inappropriate for explanatory and causal explanation. This view has, however, been questioned by various authors who point out that some of the best and most famous case studies have been explanatory case studies, see e.g. Steinar Kvale (1996), *Interviews. An Introduction to Qualitative Research Interviewing*, Sage: Thousand Oaks, London, New Delhi,

[143] Gordon Kingsley and Hans Klein (1998), 'Interfirm collaboration as a modernization strategy: a survey of case studies,' *Journal of Technology Transactions*, Vol 23 No 1

entities through networks of excellence and similar instruments. Case studies and interviews are able to assess the role of individual actors within innovation networks and these accounts can then be assembled to generate a coherent description of the effect of a policy.

- However, in developing an assessment of network and innovation system impacts, the basic unit of analysis need not be at the level of the single firm. Larger entities can be assessed, for example regional innovation systems or clusters, which reflect the performance of firms and their inter-relationships. The relevant tools include network analysis and bibliometrics (see sections A.11 and A.13).

Box 15 **The Advanced Technology Programme ATP**

This example illustrated the extensive use of descriptive case study (in combination with the systematic collection of indicator data) to provide progress reports on all completed R&D projects funded by the Advanced Technology Programme. The case studies are written three or more years after an ATP project is completed. They briefly tell the "story" of what happened during the course of the project and after completion. They explain why ATP funding was sought for the project, how the research could benefit the U.S. economy and society, how the project was conducted, and what technological and commercial achievements have occurred since the ATP funded portion of the research ended. Data for key project inputs, outputs and outcomes are systematically collected and aggregated to provide interim performance metrics for use by the programme. These include numbers of publications, patents, prototypes, commercial products in the market and products to be expected soon, employment effects, and awards received from third parties in recognition of scientific accomplishment and business acumen.

A report available online (http://www.atp.nist.gov/eao/sp950-2/contents.htm) assesses the first fifty completed projects – approximately 10% of the projects funded by the ATP from 1990 to 2000. The majority of the first 50 completed projects are single-applicant projects led by small businesses. Although only 16 percent are joint ventures, 84 percent involved collaborative relationships. The estimated benefits far outweigh the entire cost of the ATP to date.

The set of project case studies includes all levels of project performance, ranging from poor to outstanding based on success criteria. A four-star rating system, based on the collected case-study metrics, is used to score the performance of each project and to show the distribution of project performance across the portfolio of completed projects.

Sources: NIST (ed) (2001), *Performance of 50 Completed ATP Projects*. Status Report - Number 2, NIST Special Publication 950-2, U.S. Government Printing Office, Washington, D.C., December

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

A.9.3. Strengths

Case study methods provide both detailed understanding of effects of research, technology and development programmes and can also be used to produce sets of comparable cases that may assist evaluators with the assessment of net effects and efficiency and effectiveness studies. Also, by bringing in substantial programme/project information on a less restrictive basis than most methods, case studies document events and provide a richness of detail that may prove useful in formulating theories and hypotheses that lay the groundwork for further evaluations.

Another advantage of the case study method is that many decision makers read and process anecdotal evidence more easily than they do quantitative studies.

Box 16 **The Small Business Innovation Research (SBIR) programme**

Mandated as a part of SBIR's reauthorization in late 2000, the National Research Council conducted a comprehensive study of how the SBIR programme has stimulated technological innovation and of how firms use SBIR grants to develop their businesses through innovation and commercialisation.

The NCR included a large set of descriptive case studies as part of its assessment of the SBIR programme. The case studies focused on the five U.S. government agencies that have the largest SBIR programmes, namely, in order of programme size, the Department of Defense, National Institutes of Health, Department of Energy, National Aeronautics and Space Administration, and National Science Foundation[144] – the five agencies responsible for 96 percent of the program's operations. A set of case studies was prepared for each agency's SBIR programme to supplement results and other methods used to assess the programme.

The case studies helped to explain how small companies use the SBIR programme to obtain early seed capital to launch technologies – and often their businesses – and expand innovative capacity and intellectual property portfolio. The case studies discussed company financing and commercialisation strategies, the outcomes of SBIR-funded projects, and summarised the views of company officials on the SBIR programme. This helped explain variations among the SBIR programme at these agencies.

Charles W. Wessner (ed.) (2008), *An Assessment of the SBIR Programme*, National Research Council, Washington DC: The National Academies Press

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

A.9.4. Weaknesses

A disadvantage of the case study method is that the anecdotal evidence provided is generally regarded as less persuasive than quantitative evidence. What is more, the results of one or more individual cases cannot be generalised to a population of cases, i.e. are not representative. Also, case studies tend to be ex post only.

In general, case studies are time intensive at the point of data collection but less time intensive in terms of data analysis. Therefore, one of the conditions necessary for the application of the method is the willingness of practitioners, technology providers, industrial partners and the like to provide access to documents and to make time for planning and interviews.

A.9.5. Quality Control

Methodological rigour applies to case study research – or more generally qualitative research, as it does to quantitative research. This tends to be forgotten. We would therefore like to refer researchers with little experience in qualitative research to text books on case study research or qualitative research (for example Yin 2009, Flick 2008, Kvale 1996).

144 Charles W. Wessner (ed.) (2008), *An Assessment of the SBIR Program at the Department of Energy.* National Research Council, Washington, DC: The National Academies Press,
Charles W. Wessner (ed.) (2008), *An Assessment of the SBIR Program at the National Science Foundation.* National Research Council, Washington, DC: The National Academies Press,
Charles W. Wessner (ed.) (2009), *An Assessment of the SBIR Program at the Department of Defense.* National Research Council, Washington, DC: The National Academies Press
Charles W. Wessner (ed.) (2009), *An Assessment of the SBIR Program at the National Aeronautics and Space Administration*, National Research Council, Washington, DC: The National Academies Press,
Charles W. Wessner (ed.) (2009), *An Assessment of the SBIR Program at the National Institutes of Health.* National Research Council, Washington, DC: The National Academies Press,

Moreover, even if data collection is more time intensive than data analysis in case study analysis, the latter should not be neglected. A common occurrence is that data collection takes so long that there is hardly enough time for careful data analysis, which evidently affects the quality of the case study.

For comparative case studies involving a large number of cases we recommend the use of a qualitative data analysis software, to analyse the data.

A.9.6. References

Uwe Flick (2008), *Designing Qualitative Research*, Sage

Gordon Kingsley and Hans Klein (1998), 'Interfirm collaboration as a modernization strategy: a survey of case studies,' *Journal of Technology Transactions*, Vol 23 No 1

Steinar Kvale (1996), *Interviews. An Introduction to Qualitative Research Interviewing*, Sage: Thousand Oaks, London, New Dehli

Charles W. Wessner (ed.) (2008), National Research Council, *An Assessment of the SBIR Programme*, Washington DC: The National Academies Press

Michael Quinn Patton (1990), *Qualitative Evaluation and Research Methods,* Second Edition, Sage: Newbury Park, London, New Delhi

Rosalie Ruegg and Irwin Feller (2003), *A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade*, NIST GCR 03-857, 2003, pp. 34-35

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programmes, A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

Elliot Stern (2002), 'Field Studies and Case Studies', in Gustavo Fahrenkrog et al. (eds.) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, Sevilla, , pp. 173-182

Robert K. Yin (2009), *Case study research: Design and method*s, Fifth edition. Thousand Oaks, CA: Sage

A.10. Historical tracing

A.10.1. Description

The historical tracing method resembles the case study method in terms of providing an in-depth investigation in a storytelling mode. What sets it apart is its emphasis on tracing chronologically a series of interrelated developments leading from research to ultimate outcomes or from outcomes back to the factors that spawned them.

A.10.1.1. Forward Tracing

When the objective is to evaluate a given project, forward tracing, where the analyst starts with the research of interest and traces the evolution of related events from that point forward, is generally more manageable and cost-effective than backward tracing and produces a relatively complete portrayal of a project's impacts. Forward tracing enables the investigation of all known pathways leading forward from the project and contributes to a better understanding of the evolutionary processes of science and technology.

A.10.1.2. Backward Tracing

In contrast, backward tracing, in which the analyst starts with an outcome of interest and traces backward to identify the critical developments that appear instrumental to the outcome, may or may not lead back to the project of interest. And if it does, the study may have a narrow focus that misses other effects associated with the project. For these reasons, the backward tracing approach seems more appropriate: (1) when the outcome is the central focus, (2) when a particular outcome is of known significance and the programmatic linkage is also known to exist, or (3) when the purpose is to show in a general way how significant outcomes are rooted in a certain type of programmatic funding or in work funded or conducted by certain organizations. An appeal of the approach is that the significance of the outcome is already established rather than evolving.

Box 17 **The Department of Defense's Project Hindsight**

The earliest example found of historical backward tracing used by a U.S. government research programme is "Project Hindsight," conducted by the Department of Defense in the early 1960s. The study traced backwards over 20 years the development of each of 20 major weapons systems supported by DOD in order to identify the key research outputs that contributed to their realization. The study used the approach of interviewing experts. It linked the support of research to a variety of desirable technological outcomes. It examined characteristics of what were identified as critical R&D events to ascertain whether any general principles could be extracted. A major conclusion related to the science-technology conversion process was that the results of research were most likely to be used when the researcher was intimately aware of the needs of the applications engineer.

C.W. Sherwin and R.S. Isenson, "Project Hindsight: Defense Department Study of the Utility of Research," *Science*, 156 (1967), pp. 1571-1577.

A.10.2. Uses

Historical tracing is used in the following ways[145]

- To show the path by which a particular research programme or project led to useful downstream products and processes.
- To increase understanding of the evolutionary processes of R&D and innovation.

[145] Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

- To suggest that the benefits of research outweigh its costs by comparing a proven-to-be-valuable innovation against the costs of a research programme shown to underpin the innovation.

The approach to conducting historical tracing studies has evolved over time. Studies have most often been organized as backward tracing studies to examine key mechanisms, institutions, activities, and processes that seemed to play a key role in an observed innovation. Earlier studies relied mainly on expert opinion solicited by interview to identify and understand key events in the development of an innovation. Each interview would often identify earlier events, people, and organizations to investigate on the backward tracing path. As computerized citation analysis developed, it was found that citation analysis studies could be helpful in identifying a path of linkages to follow. Tools of social network analysis may also be useful to identify linkages among people and organisations. The result has been the evolution of a hybrid approach to historical tracing studies that combines "detective work," expert opinion solicited by interview, and publication and patent citation analyses. Results have been presented as roadmaps leading to and from research programs to successful innovations.

Variations of historical tracing are the Finnish VINDI project (Box 18) and the 'tracing back' case studies Technopolis conducted for the UK Arts and Humanities Research Council AHRC (Box 19).

Box 18 **The Finnish VINDI project**

The impact framework and indicators project VINDI recently launched by Tekes, the Finnish innovation agency, and the Academy of Finland, the Finnish research council, has been working to develop a new, 'reversed' approach. This approach starts out from the end-result, from the question of what kinds of impacts science, technology and innovation are expected to generate.

The VINDI project progresses from impacts and outputs to inputs, thus tracing successive chains of impacts. The key question from the point of view of the VINDI projects is what impacts can and should be expected from science, technology and innovation. In other words, the project combines historical tracing with a normative approach.

The focus is on four core areas of society. These include the economy and its renewal, learning and knowledge, Finns' well-being and the environment. In each impact area the first step was to define the indicators that describe the impact of science, technology and innovation. The impact area 'the economy and its renewal' addresses the economic impacts of science, technology and innovation. The impact area 'learning and knowledge' includes impacts of R&D and innovation activities on the accumulation of knowledge, the skilled labour force and networks of experts. The impact area 'Finn's well-being' consists of impacts on objective and subjective factors of well-being, such as health and social relations. The impact area 'environment' addresses impacts in the face of environmental challenges, such as climate change.

This approach offers several benefits. It not only ensures that indicators are socially relevant, it also ensures that attention turns away from inputs and actions to outputs and impacts. Another significant advantage is that impact indicators are easier to integrate as part of the overall evaluation and development of science, technology and innovation.

Tarmo Lemola, Janne Lehenkari, Erkki Kaukonen and Juhani Timonen (2008), *Impact Framework and Indicators for Science, Technology and Innovation (VINDI)*, Academy Report 6/08, Helsinki (in Finnish)

A.10.3. Strengths

The historical tracing method tends to produce interesting and credible studies documenting a chain of interrelated developments. By providing linkage all the way from inputs to outputs it may shed light on process dynamics.

Box 19 **'Tracking-back case studies' in forensic linguistics**

Technopolis was asked by the UK Arts and Humanities Research Council AHRC to conduct three 'tracking-back case studies' in the areas of forensic linguistics, forensic archaeology and evidence-based medicine. The aim was to show the societal impact of publicly funded research in the humanities and give an idea of how research in the humanities is conducted.

The forensic linguistics case study outlines the areas forensic linguists are active in, not only in terms of research but also in terms of application of their expertise in criminal investigations and court cases. The description of a specific criminal court case – the posthumous overturning of Derek Bentley's conviction for murder for which he was hanged in1952 – showed the way forensic linguists work and the methodology they typically use.

A timeline, starting out with the first mention of the term forensic linguistics in 1968 and tracing the evolution of the discipline in UK academia and its increasing branching out into practice in the UK and in Europe, showed the growing importance of forensic linguistics in court cases and in criminal investigations.

The methodologies used included desk-top research and interviews with experts from the field.

A.10.4. Weaknesses

A disadvantage of the approach is that chains of events tend to be highly complex with many organisations and researchers involved, making it sometimes difficult to know the significance of apparent linkages.

Establishing cause and effect is difficult; antecedents to technological innovations are complex. A given innovation is typically the result of a number of direct and indirect effects, efforts by multiple people and organizations, and the synthesis of advances in knowledge on multiple fronts, often occurring over decades prior to the emergence of the innovation of focus. Hence, historical tracing studies typically require the elapse of considerable time in order for a history to be established. Substantial judgment is required to assess the comparative significance of various research events; significant events may be overlooked or dropped from an investigation. These studies tend to be time consuming and costly.

A.10.5. Quality Control

As attribution and establishing cause and effect are particularly difficult in historical tracing, we highly recommend a mix-method approach. The use of various data sources (data triangulation) and methods (method triangulation) will ensure a multi-faceted perspective.

A.10.6. References

Rosalie Ruegg and Irwin Feller (2003), *A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade*, NIST GCR 03-857, 2003, pp. 50-51

C.W. Sherwin and R.S. Isenson (1967), "Project Hindsight: Defense Department Study of the Utility of Research," *Science*, 156, pp. 1571-1577

Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

A.11. Network Analysis

A.11.1. Description

Networks are an important concept in the study of science and technology. They serve both as guiding metaphors for conceptualising the relationships between actors and, more recently, as techniques to measure structural properties of the ensemble. All network studies share the assumption that the ties between actors, which connect them into a system, are as important – if not more important – than their individual attributes. The network studies available in the literature make a persuasive case for the importance of such ties for understanding the development of science and technology.[146]

Social network analysis is a method of visually mapping and measuring relationships and linkages among researchers, groups of researchers, laboratories, or other organisations. Network analysis is relevant to R&D evaluation because it identifies routes of interactions by which ideas, knowledge, and information flow among participants in R&D, thereby possibly influencing the nature, quality, and quantity of research and innovation, as well as the dissemination of created knowledge through the network.

The underlying concept is that the conduct of science is a social activity collectively performed and resulting in "communities of practice"[147]. Network shape, size, and density can serve as indicators of the strength of communities of practice and signal relative roles and relationships.

Researchers, research groups, and other entities in a network are denoted as nodes. The relationships or flows between entities are called links or ties, denoted as lines linking the nodes. Arrows show the direction of the relationship, with incoming arrows showing that the node is a source of information and outgoing arrows showing that the node seeks information from the linked node. Relationships can also be unidirectional, as is e.g. the case in co-authorship. A sequence of links from one node to another is called a path.

Data for diagramming networks are collected in a variety of ways, such as by conducting interviews or surveys, tracking e-mail flows, observing interactions, counting co-authorship, and analysing CVs of researchers.

In social network analysis, a distinction is made between the *characteristics of relationships (links, ties)* such as strength of the relationship and frequency of interaction and *the characteristics of networks* such as size[148] and density[149] of the network or extent to which a network member is directly connected with others (adjacency). Another important measure is centrality, based on the number of direct links one node has to other nodes.[150]

Box 20 **Communication and co-operation patterns in the German Interdisciplinary Centres for Clinical Research**

Since 1996 the Federal Ministry for Education and Research (BMBF) has been

[146] Juan D. Rogers, Barry Bozeman, Ivan Chompalov (2001), Obstacles and Opportunities in the Application of Network Analysis to the Evaluation of R&D, *Research Evaluation*, 10 (3), December, pp. 161-172

[147] Rosalie Ruegg and Gretchen Jordan (2007), *Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs*, U.S. Department of Energy, March

[148] Usually the size of a network is indexed simply by counting the number of nodes

[149] The density of a binary network is the proportion of all possible ties that are actually present

[150] Robert A. Hanneman and Mark Riddle (2005), Introduction to social network methods, Riverside, CA: University of California, Riverside, (published in digital form at http://www.faculty.ucr.edu/~hanneman/nettext/)

supporting eight model centres for interdisciplinary clinical research at university hospitals.

The decision to examine the communication and co-operation structures at the eight centres was based on the objectives of the programme. The programme aims to encourage junior staff, to establish efficient structures for interdisciplinary clinical research and to enhance interdisciplinarity and scientific quality. As the encouragement of junior staff and the extent of interdisciplinary co-operation are reflected in daily work, an analysis of the co-operation and communication forms and mechanisms was considered suitable for answering the question to what extent the programme met these objectives.

The co-operation and communication analysis was based on a written survey of all persons identified as members of the clinical centres. By taking into account the differing framework conditions of each of the eight centres, it was also possible to identify influential factors for successful communication and co-operation.

The study identified two structurally different types of networks – "old" and "new" networks. "Old" networks are characterised by actors who are older and have a higher professional status. This group of persons is characterised by less frequent interactions, spatial proximity is regarded as less important for initiating and cultivating communication and co-operation relationships. In other word, the higher the levels in the hierarchy, the more inclined researchers are to confine their "social circle" to the own group or to perceive them as possible cooperation partners. In contrast, "new" networks are characterised by a multiplicity of younger staff – PhD students, postdocs or recently habilitated researchers. For the first two groups in particular, frequent interactions and spatial proximity play an important role.

The analysis also showed that interdisciplinarity is a result of the quality of communications, as well as the size and duration of exchange relationships. In centres with old networks it was seen that longstanding communicative relations can, under certain circumstances, compensate for unfavourable research framework conditions, i.e. for example the lack of central meeting places in the form of central labs.

Susanne Bührer and Viola Peter (1999), Kommunikation und Kooperation in den interdisziplinären Zentren für klinische Forschung. Fraunhofer IRB-Verlag,

Susanne Bührer, Network Analysis, Gustavo Fahrenkrog et al. (eds.) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, Sevilla, pp. 183-189

A.11.2. Uses

The use of social network analysis as a technique is comparatively novel in studies of research and innovation and is being incorporated into evaluations on more of an experimental basis. In principle, network analysis should increase in importance as it becomes necessary to explore in more detail the claims about the importance of networks in research and innovation processes. Indeed, the promotion of 'innovation networks', 'knowledge transfer networks', and 'competence networks' has increased dramatically in the last decade.

Social network analysis can be utilised[151]

- To analyse the impact of R&D policies on collaborative activity
- To reveal dominant researchers or research organisations
- To improve understanding of what form collaborations take, and their dynamics

[151] Rosalie Ruegg and Gretchen Jordan, Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs, U.S. Department of Energy, March 2007

- To identify and foster emergent knowledge systems; to assess their strengths and weaknesses. Network analysis can point to weak spots in the communication and co-operation of network members
- To highlight the importance to participants of intangible asset development, and to assess more fully knowledge spillovers
- To provide information to help programme managers make decisions to design or revise their programme, re-direct existing R&D funds, or allocate new funds

Social analysis is not only a measure to find out certain parameters of the network like density, centrality etc., but it is also a tool to promote communication within the networks. This is why network analysis should be used in the context of intermediate evaluations. Experience in different contexts shows that using network analysis as a learning medium to improve network communication is highly appreciated by those who are being evaluated[152]. It is also possible to organise workshops with network members in order to validate the results and discuss ways to optimise the network structure.

So far, network analysis has been used at a small scale in evaluations of phenomena such as networking in health research (see Box 20), and, experimentally, at a much larger scale in understanding some of the co-operative networks associated with the Framework programmes (see Box 22).

In many respects, the greatest success in network-based R&D evaluation has been with bibliometric applications such as co-authorship and co-citation analysis, as data is readily available and relationships are clear-cut. Citations, co-authorships, co-citations, and co-occurrences of keys words (co-word analysis, see 0) are taken as indicators of links between individuals, teams, groups of scientists, or journals. However, the overwhelming majority of the findings from constructing maps of co-authorship and co-citation patterns over time just confirm through sophisticated network means characterisations of changes in modern science like greater networking and collaboration, increased globalisation and growing interdisciplinarity[153].

A.11.3. Strengths

Network analysis provides a systemic perspective on knowledge systems different from and complementary to surveys of beneficiaries or impact assessment evaluations. The analysis of networks can have a multiplicity of surprising results, even for the members of the network.

Because of the high standardisation of techniques of social network analysis, results of separate networks can be compared relatively easily.

A.11.4. Weaknesses

More often than not network analysis has been used as a set of formal techniques that could be applied to the study of a particular problem. As a consequence, studies are largely descriptive. Those that have an evaluative or normative element rarely tie the evaluative elements to the network itself. Hence, it appears that R&D evaluation requires something more than identification and description of networks; one must have some criterion by which networks can be said to have improved or declined or to have ‘good’ or ‘bad’ structures.

---

[152] Susanne Bührer (2002), Network Analysis, Gustavo Fahrenkrog et al. (eds.) (2002), *RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies*, Sevilla, , pp. 183-189

[153] Juan D. Rogers, Barry Bozeman, Ivan Chompalov (2001), Obstacles and Opportunities in the Application of Network Analysis to the Evaluation of R&D, *Research Evaluation*, 10 (3), December, pp. 161-172

Box 21 **Nordic Research Excellence**

In a bibliometric study for Nordforsk, the Scandinavian Research Council, we identified areas of Nordic research strength, first at the level of fields of science and then at a much more specific, sub-field level and suggested funding priorities.

We used social network analysis to map co-authorship at the institutional level, in order to understand whether the linkages in certain sub-fields are fragmented or represent more solid alliances among institutions. We distinguished between Scandinavian and non-Scandinavian institutions.

We found networks of interconnected Scandinavian institutions in four out of the ten sub-fields we examined and recommended Nordforsk to build on them. The sub-fields are: fisheries, ecology, sports science and marine and freshwater biology.

Source: Yann Cadiou and Erik Arnold (2008), *Nordic Excellence, A Bibliometric Exploration of Common Nordic Research Funding Opportunities*, Nordforsk Policy Briefs 2008-5

With regard to the former, a network diagram shows the relationships as of a specific point in time, such that repeating the process after a time interval is necessary to reveal changes in the network over time. It also means that a baseline needs to be established if the effects of a programme that promotes networking is to be investigated.

Box 22 Using social network analysis to study co-operative networks in the FP6

Co-operative networks in the Sixth Framework Programme FP6

The European Commission's Sixth Framework Programme (FP6 2002-2006) had a goal of promoting denser and more diverse research collaborations. The study was conducted to gain insights into the degree to which Information Society (IST) researchers were collaborating and the impact of the new FP6 Instruments on the integration of IST research. It assessed six research areas, known as "Strategic Objectives", using three types if analysis: network analysis to examine the structures of the networks, content analysis of the types of participants in the network and survey research to understand the motivations of participants for joining the networks.

The study found that the Framework Programme has provided a major integrating function and created an intensely linked network of IST research collaborations in Europe. The funding helped to connect universities and businesses and researchers from different disciplines, and to integrate new Member States, key patent-holders, and small businesses into collaborations. The study also found that FP6 resulted in increasing the interconnectedness of network participants more than previous FPs did. However, it also raised concerns that large institutes and companies act as gate-keepers to Framework Programme participation, providing stability over time, while SMEs may bring new ideas and resources but are not as likely as larger organisations to stay connected to the Framework Programme over time.

Caroline S Wagner et al.(2005), ERAnets: *Evaluation of Networks of Collaboration Among Participants in IST Research and their Evolution to Collaborations in the European Research Area (ERA)*, Report to the European Commission Directorate-General Information Society, Amsterdam: RAND Europe

With regard to the latter, there appear to be few fields of social sciences where there is a greater disparity between technique and theory than in network analysis. Network concepts lend themselves to formalisation. The result is the development of high-power technique as reality is increasingly simplified for the simplicity required for formalisation. Given the expansive dominion of network approaches it is surprising how few propositions can be advanced as part of 'network theory'. For example, we do not know what structure a network must have in order to be conducive to innovation. Hence, more theoretical work is required.

Finally, data collection may be time-consuming and costly, as determining networks can require extensive surveying, many interviews or time-consuming analysis of CVs or the like, to trace the network, particularly if the evolution of a network is traced across time and takes account different types of connections.

A.11.5. Quality Control

Social network analysis is a time-consuming exercise. Hence, careful thought should be given as to what the value added of a social network analysis is. Does the mapping of network relationships tell us anything that we do not already know? How useful are the statistics and concepts used in network analysis (density, centrality, cliques and sub-groups) to answer the evaluation questions?

A.11.6. References

Susanne Bührer and Viola Peter (1999), Kommunikation und Kooperation in den interdisziplinären Zentren für klinische Forschung. Fraunhofer IRB-Verlag

Susanne Bührer (2002), Network Analysis, in Gustavo Fahrenkrog et al. (eds.) (2002), RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies, Sevilla, , pp. 183-189

Robert A. Hanneman and Mark Riddle (2005), Introduction to social network methods, Riverside, CA: University of California, Riverside, (published in digital form at http://www.faculty.ucr.edu/~hanneman/nettext/)

Juan D. Rogers, Barry Bozeman, Ivan Chompalov (2001), Obstacles and Opportunities in the Application of Network Analysis to the Evaluation of R&D, *Research Evaluation*, 10 (3), December, pp. 161-172

Rosalie Ruegg and Gretchen Jordan (2007), Overview of Evaluation Methods for R&D Programs. A Directory of Evaluation Methods Relevant to Technology Development Programs, U.S. Department of Energy, March

Yann Cadious and Erik Arnold, Nordic Excellence, A Bibliometric Exploration of Common Nordic Research Funding Opportunities, Nordforsk Policy Briefs, 2008-5

Caroline S. Wagner, Jonathan Cave, Tom Tesch, Verna Allee, Robert Thomson, Loet Leydesdorff, Maarten Botterman (2005), *ERAnets: Evaluation of Networks of Collaboration Among Participants in IST Research and their Evolution to Collaborations in the European Research Area (ERA)*, Report to the European Commission Directorate-General Information Society, Amsterdam: RAND Europe

A.12. Prospective studies

Studies concerned with futures, such as foresight studies and technology assessments, have not traditionally formed part of the evaluation repertoire. To the extent that they inform the design of policies and programmes, however, they have clear importance – especially where evaluations are based on modelling the logic of the interventions.

A.12.1. Foresight Studies

Foresight studies encompass a range of methods, aiming to look into the longer-term future of science and technology in order to identify the areas of strategic research and emerging generic technologies likely to yield the greatest economic and social benefits[154] and to help set priorities and identify policy needs.[155]

Science and technology foresight exercises are becoming increasingly attractive for governments, national research agencies and businesses in their efforts to cope with the increasing complexity of new technologies and decision environments. Since the 1990s, quite a number of major foresight exercises have been launched in many European countries.

Foresight exercises are ways of obtaining opinions, conflicting or otherwise, about future developments. Foresight in this sense is an essential contributor to the creation, either collectively or individually, of models of the future. Such models are important because they are capable of creating synthesis, they are disruptive and interfere with current modes of thought, thus forming and shifting values.

Future developments underlie influences which cannot be assessed exhaustively in advance, thus not predicted. Nonetheless, there is a need to "monitor the future prospectively": the accelerating changes that individuals as well as societies have to adapt to, make it necessary to anticipate these changes before they become reality.

In the context of policymaking, the most important aims of foresight exercises are:

- to find out new demand and new possibilities as well as new ideas
- to identify a choice of opportunities, to set priorities and to assess potential impacts and chances
- to discuss desirable and undesirable futures
- to prospect the potential impacts of current research and technology policy
- to focus selectively on economic, technological, social and ecological areas
- to start monitoring and detailed research in these fields.

The UK's Foresight Programme is a good example of an extensive foresight programme used to inform government policy and strategy, and to improve how science and technology are used within Government and by society.[156]

A methodology frequently used in foresight studies is the Delphi method. The Delphi method relies on a panel of independent experts. The carefully selected experts answer questionnaires in two or more rounds. After each round, a facilitator provides an anonymous summary of the experts' forecasts from the previous round as well as the reasons they provided for their judgments. Thus, participants are encouraged to revise

[154] Ben R. Martin (1995), Foresight in Science and Technology, *Technology Analysis & Strategic Management*, Vol. 7, No. 2, , p. 140

[155] Luke Georghiou (2001), Third generation foresight – integrating the socio-economic dimension, Proceedings of the International Conference on Technology Foresight – the approach to and potential for New Technology Foresight, Tokyo: NISTEP; The Role of Foresight in the Selection of Research Policy Priorities, Conference Proceedings, Seville: IPTS, July,

[156] http://www.foresight.gov.uk/index.asp

their earlier answers in light of the replies of other members of the group. It is believed that during this process the range of answers will decrease and the group will converge towards the "correct" answer. Finally, the process is stopped after a pre-defined stop criterion (e.g. number of rounds, achievement of consensus, stability of results) and the mean or median scores of the final rounds determine the results.

A.12.2. Technology Assessment

In very general terms, technology assessment can be described as the anticipation of impacts and feedback in order to reduce the human and social costs of learning how to handle technology in society by trial and error.

Technology assessment attempts to provide policy-makers with a rational basis for their decisions. It does not provide 'super expert opinions', but points out areas where specialists are in general agreement, where controversy exists, what assumptions or fears lie behind the differences of opinion, and what risks are associated with the various possible options. It highlights facts that policy-makers can safely assume to be accepted and problems that they have to solve through political decisions. By focusing on political discussion of relevant questions, the decision-making process can be made simpler, more objective and efficient.

Technology assessment provides input for the elaboration of technology policies which are not oriented exclusively towards technical efficiency and economic rationality but which also consider the social and ecological consequences of the introduction of specific technologies.

Technology assessment covers

- Critical studies, aiming to understand likely future social and economic effects of the use of certain technologies: this type of technology assessment tries to identify technology-induced risks at an early stage, in order to analyse in detail the range of possible social, economic, legal, political, cultural and ecological effects, present alternative decision-oriented options, and at the same time point out the various social interests and value judgements linked with the development and use of new technologies.
- So-called constructive technology assessment, which aims not only to criticise but to set the agenda for future research and technological development based on social needs: this type of technology assessment is based on the realisation that technical developments are ultimately influenced by society and not determined solely by their own logic. This calls for a wider vision of technology assessment, one which takes account of design-related factors and forces in the R&D process and their application in a specific technology, and which also allows social discussion of the objects and options of technological development. Thus, the core analytical functions of technology assessment are supplemented by constructive design and discussion-provoking functions.

There is no such thing as a 'technology assessment method' as such, because it is not a single method but the variety of methods – or method mix – and specific procedure that characterise a successful investigation.

Foresight and technology assessment can jointly contribute to strategic intelligence about future developments and their value. However, a difference in style and context will remain: while foresight aims at opening up spaces for thinking about new possibilities, technology assessment is oriented towards selecting or at least modifying and modulating developments. The link of technology assessment with judgment, decisions and strategies implies that there will be more and more broadly based controversy than with foresight, which often remains limited to communities of experts.

There has recently been renewed interest in technology assessment, and that this has to do with the increased possibilities of combining private-domain and public-domain technology assessment, and with the role of technology assessment in broader priority setting, technology road-mapping, and articulation of views about new technology.

A.12.3. Uses

Technology foresight results can be useful when evaluating a research institution. The foresight results enable evaluators to get an overview of the fit between perceived future developments in science and technology worldwide and the performance portfolio of a publicly funded research organisation. This can for example be done by constructing an index, which compares the results of a Delphi study with the research activities and/or the staff competences of a given sample of research units.

Since foresight exercises and technology assessments are costly, their conduct just for the purpose of an evaluation would mean a considerable investment. Hence, it is advisable to use the results of existing foresight exercises and technology assessments. This of course presupposes that policymakers initiate and support the conduct of regular foresight exercises and technology assessments.

Box 23 **Evaluation of the German Fraunhofer Society**

The Fraunhofer Society (FhG) is a semi-public research organisation currently consisting of 57 largely autonomous institutes, primarily active in the filed of applied technological research. In 1996, the Fraunhofer Society was evaluated in the framework of a larger evaluation of all German research institutions. Among the most important questions of the FhG evaluation were the following questions: Which technology-related markets promise the largest growth? Is FhG sufficiently represented in these markets? Does the technological portfolio of FhG fit with related technological developments worldwide?

The international panel in charge of the evaluation decided to employ the results of the German Delphi '98 study as a benchmark for FhG's research and technology competencies. The Delphi study offered 'a thousand visions' of problem solutions based on future scientific or technological achievements. For the evaluation, a set of figures of 'important visions' of future developments in science and technology was developed and contrasted with competencies of the Fraunhofer Society. Thus, strengths and weaknesses of FhG's competencies vis-à-vis potential future research markets were revealed.

Source: Stefan Kuhlmann (2002), Foresight and Technology Assessment as Complementing Evaluation Tools, Assessing the Socio-Economic Impact of RTD Policies, Sevilla, , pp. 195-196

A.12.4. Strengths

The enhancement of R&D evaluation with foresight exercises and technology assessment helps to broaden the scope of actual or potential, intended or non-intended impacts and effects of public interventions.

A.12.5. Weakness

An important limitation of foresight is the well known fact that sudden science and technology breakthroughs often have not been foreseen by the majority of main-stream experts but were anticipated by a few unorthodox thinkers only. This is a classical problem of foresight and other methods of "prospection": how to detect feeble signals or the minority views that could be revealed as the very precursors of the future? The paradoxical nature of foresight tools is that they aim at two conflicting goals: building consensus and preserving a variety of visions.

As technology assessments is used in a variety of contexts using a multitude of methods, it is more difficult to identify the weaknesses.

Since foresight and assessment are complex combinations of methodologies in themselves, the options and limitations of the combination with R&D evaluation are inevitably huge. However, there are two basic limitations that should be mentioned:

a) The cost of combining such complex efforts (resources, time) are potentially high.

b) The problem of the causal attribution of potential scientific, technological, social or economic developments to a public policy measure under evaluative scrutiny – a basic problem of any policy evaluation – becomes even more an issue in this case.

A.12.6. Quality Control

Evidently, the normal quality controls apply for foresights and technology assessments conducted on their own.

However, since the enhancement of R&D evaluation with foresight exercises and technology assessment is still in an experimental stage, it is difficult to indicate measures of quality control for this application.

A.12.7. References

Luke Georghiou (2001), Third generation foresight – integrating the socio-economic dimension, *Proceedings of the International Conference on Technology Foresight – the approach to and potential for New Technology Foresight*, Tokyo: NISTEP

Luke Georghiou and Michael Keenan (2006), Evaluation of national foresight activities: Assessing rationale, process and impact, Technological Forecasting and Social Change, Vol. 73 (7), Sept, Pages 761-777

IPTS-JRC (ed.) (2002), *The Role of Foresight in the Selection of Research Policy Priorities*, Conference Proceedings, Seville: IPTS, July

Stefan Kuhlmann (2001), Management of Innovation Systems: The Role of Distributed Intelligence. Maklu: Apeldoorn/Antwerpen

Stefan Kuhlmann (2002), *Foresight and Technology Assessment as Complementing Evaluation Tools*, Assessing the Socio-Economic Impact of RTD Policies, Sevilla, , pp. 192-199

Ben R. Martin (1995), Foresight in Science and Technology, *Technology Analysis & Strategic Management*, Vol. 7, No. 2, , pp. 139-168

Michael Rodemeyer, Daniel Sarewitz, James Wilson (2005), The Future of Technology Assessment, Woodrow Wilson International Center for Scholars, Washington, DC,

http://www.foresight.gov.uk/index.asp (UK Foresight programme)

http://www.itas.fzk.de/home_e.htm (Institute for Technology Assessment and Systems Analysis (ITAS), Germany)

http://www.oeaw.ac.at/ita/welcome.htm (Austrian Institute of Technology Assessment)

http://www.ta-swiss.ch/ (Swiss Centre for Technology Assessment)

http://www.tekno.dk/ (Danish Board of Technology)

A.13. Bibliometrics and Patent Analysis

A.13.1. Description

Publications and patents constitute major outputs of research and innovation programmes, and the large databases created to capture these outputs support the bibliometrics method of evaluation. As the term is used here, bibliometrics encompasses tracking the quantity of publications and patents, analysing citations of publications and patents, and extracting content information from documents. Bibliometrics is used to assess the quantity, quality, significance, dissemination, and intellectual linkages of research, as well as to measure the progress, dynamics, and evolution of scientific disciplines.

Bibliometrics typically compares the relative scientific contributions of two or more individuals or groups. These evaluations are sometimes used to inform research policy and to help direct research funding. Indeed, growing use is being made of bibliometric performance indicators as an adjunct to ex post peer review (for example, in the UK Research Assessment Exercise), such as the impact factors of journals in which people publish and the extent to which their work is cited by others. This is encouraged by the ready availability of such indicators within academia that contrasts strongly with the difficulty and expense of generating or accessing larger scale indicators outside the academic sphere.

The field of scientometrics – the quantitative study and analysis of the outputs of research (normally publications and patents) – is comparatively young. It requires heavy investments in databases and specialised analytical capabilities, so there is a quite limited number of people and organisations able to work with such methods. Capital and skill barriers to entry may partly explain why these techniques have been little used in general R&D evaluation. However, it is also often difficult to detect or attribute impacts of programmes on patterns of development in fields or countries. Even to identify the effects of a national research council's funding on the national science base is hard.[157]

A.13.1.1. Citation Databases

Bibliometric analysis can be conducted on the bases of any sufficiently large publication database in which publications are annotated with the count of citations they received. The bibliographic databases of the Institute for Scientific Information (ISI) have been the most generally accepted data sources for bibliometric analysis. The Thomson Corporation[158] is one of the world's largest information companies providing specialized information content on which financial, legal, research and medical organizations rely to make business-critical decisions and drive innovation. ISI maintains the Web of Knowledge, an online academic database which provides access to many resources, in particular:

- Web of Science (WoS), including the Science Citation Index (SCI), the Social Science Citation Index (SSCI), and the Arts and Humanities Citation Index (AHCI).
- Journal Citation Reports (JCR), containing citation information, and in particular the Impact Factor, defined as the average number of recent citations per paper, of

[157] Erik Arnold, Stefan Kuhlmann and Barend van der Meulen (2001), *A Singular Council: Evaluation of the Research Council of Norway*, Oslo: Royal Norwegian Ministry of Research, Education and Church Affairs

[158] The ISI was founded by Eugene Garfield in 1960. Following ideas inspired by Vannevar Bush's famous article 'As we may think' (Bush, 1945), Garfield undertook the development of a comprehensive citation network, based on post-war sociological theory allowing it to be used to assess the impact of scientific work. The ISI was acquired by Thomson in 1992. In 2007, The Thompson Corporation reached an agreement with Reuters to combine the two companies under the name Thompson-Reuters

the journals tracked by ISI. The JCR is published annually in separate editions for the sciences and the social sciences.

Fierce competitors of Web of Science are Elsevier's Scopus and Google Scholar. Scopus, as Web of Science, is a subscription-based proprietary databases, while Google Scholar is freely accessible.

Box 24 **The SCImago Journal and Country Ranking**

The SCImago Journal & Country Rank a freely available database of journals citation metrics that tracks a greater number of journals than Thompson Scientific. The portal includes the journals and country scientific indicators developed from the information contained in the Scopus database (Elsevier B.V.). These indicators can be used to assess and analyse scientific domains.

This platform takes its name from the SCImago Journal Rank (SJR) indicator, developed by SCImago from the widely known algorithm Google PageRank. This indicator shows the visibility of the journals contained in the Scopus database from 1996.

SCImago is a research group from the Consejo Superior de Investigaciones Científicas (CSIC), University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares, dedicated to information analysis, representation and retrieval by means of visualisation techniques.

As well as SJR Portal, SCImago has developed The Atlas of Science project, which proposes the creation of an information system whose major aim is to achieve a graphic representation of IberoAmerican Science Research. Such representation is conceived as a collection of interactive maps, allowing navigation functions throughout the semantic spaces formed by the maps.

http://www.scimagojr.com/

A recent study[159] has compared the Web of Science (with ISI data), Google Scholar and Scopus with the explicit objective of assessing the extent to which the results of a citation analysis depend upon the data source used, using the task of ranking the faculty in a single library and information science school. The findings showed that Google Scholar was probably too difficult to use for a large-scale citation analysis and that the other two gave similar results overall. Nevertheless, weaknesses in the coverage of certain fields resulted in significant disadvantages to some faculty members, depending upon the database used. Hence, the use of both in conjunction with each other would give the fairest results. In addition, the poor coverage of conferences by both in comparison to Google Scholar illustrates that neither give fair results to academics who publish in fields which emphasise conferences, such as computer science and computational linguistics. Another investigation compared different databases for coverage of social sciences research, finding Scopus to offer particularly good coverage.[160]

Patent data are freely available on the internet. Both the U.S. Patent and Trademark Office[161] and the European Patent Office[162] offer online and offline search capability for patent citations.

[159] Meho, L.I.; Yang, K. (2007), A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar, *Journal of the American Society for Information Science and Technology*, Vol. 58, no. 13, pp. 1-21

[160] Mike Thelwell (2007), Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4), pp. 1-18

[161] http://www.uspto.gov

[162] http://ep.espacenet.com/quickSearch?locale=en_ep

A.13.2. Uses

A.13.2.1. Counting Publications and Patents

Publication and patent counts[163] are among the most widely used bibliometric indicators to evaluate a research institution or a project, as they are an easy output measure to track the quantity of an organisation's or project's publications and patents.

The count may be normalized by research costs or some other measure of input to create an indicator of research productivity. Aggregated across a programme, numbers of publications and patents per pound spent on research may serve as an indicator of programme progress, and trends in outputs may be tracked over time. Adjustment can be made to account for quality differences in publication journals. Care should be exercised in making comparisons among organisations and disciplines on the basis of their counts of publications and patents, as rates of patenting and publishing may vary for reasons other than productivity, and quality differences may not be adequately taken into account.

The impact factor (JIF) for a publication venue (journal or conference) for a specific year is the mean number of citations that occurred in that year to the articles published in that venue during the two previous years. It is an index for publication venues and not for scholars.[164]

A.13.2.2. Citation Analysis

Tracking citations of publications and patents is useful for identifying pathways of knowledge spillovers. Citations may include publications citing other publications, patents citing publications, and patents citing other patents.[165] The frequency with which publications and patents are cited is also used as an indicator of quality and significance. The more other scientists cite a research paper or patent, the greater its assumed relevance, impact, quality, and dissemination, other things being equal.

Normalization approaches can be used to help control for quality differences in the citing journals. An example of a simple normalization approach is to hold the journal constant and compare the number of citations a given paper, or group of papers, receives against the average citation rate of all papers in the journal. A value greater than one indicates the paper, or set of papers, is more heavily cited than the average.

Who is citing publications or patents may also be of interest. Examining who is citing what can reveal where a field of research or a technology is moving, and show knowledge linkages among subject areas. For example, a public programme may wish to know whether nationally owned or foreign-owned firms take up a technology it funded. It may wish to know if its research is supporting other fields of knowledge. Citations of research papers in patents may be of special interest to a research organisation because the citations show how the programme's research findings are being converted into technology and yielding economic benefits. Patents analysis have also been used to evaluate the performance of a country's technology and to identify flows of knowledge transfer between science and technology. For example, one empirical study of patents relative to the Netherlands concluded that the results did

[163] Related indicators are papers per year (the number of papers divided by the author's academic age) and papers per author (computed by dividing each paper unit by the number of authors of that paper and summing the results over all papers)

[164] For more recent bibliometric indicators see http://users.dimi.uniud.it/~massimo.franceschet/bibliometrics/indicators.html

[165] Related indicators are: citations per paper (the number of citations divided by the number of papers), citations per year (the number of citations divided by the author's academic age) and citations per author (computed by dividing each citation count of a paper by the number of authors of that paper and summing the results over all papers)

not fit the existing theoretical models of university-industry relationships, which therefore needed to be reconsidered.[166]

Box 25 **Nordic Research Excellence**

In a recent study we conducted for the Scandinavian Research Council NordForsk, we provided NordForsk with 'strategic intelligence' about areas of research excellence in the Nordic area, where it might be able further to add value through the use of cross-border funding instruments. We identified areas of national research strength, first at the level of fields of science and then at a much more specific, sub-field level, and identified areas of potential Nordic strength – both in the sense of areas where there appear to be good numbers of researchers publishing and by identifying sub-areas where the quality - measured by conventional bibliometric indicators - also appears to be excellent. These provided important clues about areas where NordForsk might focus its funding to promote Nordic strength. Our study was based on bibliometric analysis of publications in the period 2001-4.

The Nordic countries represented 3.45% of World scientific production in 1999-2001, and 3.27% in 2002-2004. Sweden had the greatest share of world scientific publications in 2004 with 1.38%. Denmark and Finland each held 0.71%, Norway had 0.44% and Iceland 0.13%. With the exception of Iceland, the overall world share of each of the Nordic countries has fallen across the period. This result was not surprising, since the world population of publishing scientists is growing.

At the level of the eight large fields of science for which there are good bibliometric data, the Nordic publication pattern is strongly oriented towards life sciences. Medicine is the field with the greatest number of publications. The Nordic Countries collectively are most specialised in: Fundamental Biology; Applied Biology-ecology; Earth and Space Science; and Medicine.

Nordforsk also asked us to identify the 10 strongest sub-fields of Nordic science from a list of more than 170 subject categories. To do this, we ranked the sub-fields in each country, based on World share of publications. We then applied a second criterion: the number of publications should be greater than 50 per year, which is not only a conventional criterion for assuring the quality of bibliometric analysis but also a way to take small fields out of the analysis. Once the fields with few publications were eliminated, we could see which were the 20 top sub-fields for each country.

We checked that our findings were not contradicted by other bibliometric work in the Nordic area and that the bibliometric analysis accorded with the knowledge of most of the Nordic research councils. However, bibliometrics does involve a number of well-known limitations, notably the poverty of bibliometric indicators outside the English language and the 'hard' sciences. Bibliometrics is also backwards looking and poor at picking up the nuances of how scientific fields develop.

Yann Cadiou and Erik Arnold (2008), *Nordic Excellence, A Bibliometric Exploration of Common Nordic Research Funding Opportunities*, Nordforsk Policy Briefs 2008-5

Citation analysis is also a useful adjunct to other evaluation methods. For example, it can facilitate historical tracing studies (see A.10). In addition, citation analysis can be used to support social network analysis by investigating paper-to-paper, patent-to-patent, and patent-to-paper citations to identify potential intellectual linkages and clusters of relationships among researchers and organizations (see A.11).

[166] L. Leydesdorff (2004), The university-industry knowledge relationship: Analyzing patents and the science base of technologies, Journal of the American Society for Information Science and Technology, 54 (11), pp. 991-1001

A.13.2.3. Content Analysis

Extracting content information is another way to use documents in evaluation. Content analysis can help evaluate the historical evolution of research funded or conducted by a particular organisation, or trace the emergence of a field of knowledge from multiple sources. One approach to content analysis is co-word analysis, which uses key words to search text. The frequency of co-occurrence of the key words for a selected database of published articles depicts the evolution of ideas and concepts. This methodology is used for mapping scientific fields and for detecting new emerging fields. It identifies keywords and relates the contents of papers with other scientific publications, grouping papers to show the structure and dynamics followed by science and technology.

A newer approach is database tomography, which avoids the need to pre-specify key words. The texts to be searched are entered into a computer database, and a computer-based algorithm extracts words and phrases that are repeated throughout the database, using the proximity of words and their frequency of co-occurrence to estimate the strength of their relationship.

A.13.3. Strengths

A major advantage of bibliometric methods is that they are widely applicable to evaluations of research organisations as well as programmes that have an emphasis on publishing or patenting. The methods can be used to address a variety of evaluation topics, including productivity trends, collaborative relationships, and patterns and intensity of knowledge dissemination. Existing databases support the methods, and the methods scale easily, making it feasible and economical to apply them to large numbers of documents. The approach is relatively straightforward, and diverse audiences can understand the results. Another important advantage is that the methods do not burden those who are the subject of evaluation because data are obtained from existing databases. Some of the bibliometric methods can be applied to a programme with a relatively short time lag. Finally, the objectivity associated with the methods lends them a high degree of credibility, making it a frequent adjunct to peer reviews.

A.13.4. Weaknesses

A weakness of bibliometric evaluation is that it treats only publications and patents as programme outputs and ignores other outputs and long-term outcomes. Another disadvantage is that time must pass before extensive patent citations can be observed.

Citation analysis rests on the assumption that the more other scientists cite a research paper or patent, the greater its assumed relevance, impact, quality, and dissemination. However, research has shown that this perspective is a simplification of reality: there are many different reasons to cite articles as well as many influences on which articles to select, when multiple options are available.

Potential problems abound in the application of the methods. For example, counts indicate quantity of output, not quality; all publications are not of equal importance; and adjustment approaches may not adequately account for differences in quality and importance. The propensities to publish and to patent differ among organizations, technical fields, and disciplines for a variety of reasons, not just productivity differences, making comparisons difficult. For example, mature technology areas can be expected to exhibit more citations than emerging technology areas. Works of poor quality may be heavily cited. A cumulative advantage process can be at work for highly cited papers, where papers that are initially well cited then tend to continue to be cited partly because they have been cited a great deal rather than for their intrinsic worth ("Matthew effect").[167] Self-citations and friend-citations may artificially inflate citation

[167] Mike Thelwall (2007), Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4), pp. 3-4

rates.

Although databases exist, the databases may be difficult to work with due to inconsistent and incomplete citations. Typical problems are weak coverage of the social sciences and humanities and a bias towards English-speaking publications.

Perhaps the most significant challenge for bibliometrics in the long run is that the new digital libraries are producing large-scale evidence of the usage patterns of academic articles for the first time.[168]

A.13.5. Quality Control

Rates of patenting and publishing vary among organisations and disciplines for reasons others than productivity. Hence, comparing counts of publications and patents is inappropriate.

The same is true for citation counts, because there are widely differing field-based citation norms. For this reason, it is now best practice to field-normalise citation indicators when using them to evaluate an academic department. Even if a set of departments in the same discipline are to be compared, raw citation counts, or average citations per researcher, would not be an accurate reflection of their citation impact because the departments might specialise in fields with different average citation rates. Hence, departments engaging in research in fields with high average citation counts would have an unfair advantage unless the indicators were normalised, for example through dividing each department's citations by the field average. In other words, the evaluative citation analysis goal has shifted from evaluating the impact of research to evaluating its impact relative to a field.

It is also important to keep in mind the well-known difficulty that evaluation of published research faces in the social sciences and humanities. The documentary databases established under the Thomson Scientific are generally considered to be inappropriate for characterising the production of non-Anglophone researchers, in the sense that a number of the journals which are considered important scientific references by non-Anglophones are not included in these databases. The bibliometric evaluation of work published in these disciplines necessitates the construction of a list of journals considered as scientific by the community of researchers concerned, a list from which, then, can be calculated conventional bibliometric indicators.

A.13.6. References

Massimo Franceschet. Bibliometrics in a nutshell, http://users.dimi.uniud.it/~massimo.franceschet/bibliometrics/index.html

Rosalie Ruegg and Irwin Feller (2003), A Toolkit for Evaluating Public R&D Investment. Models, Methods, and Findings from ATP's First Decade, NIST GCR 03-857, 2003, pp. 47-50

Science Metrix (2004). The Use of Bibliometrics in the Social Sciences and Humanities, Report prepared for the Social Sciences and Humanities Research Council of Canada (SSHRCC) – http://www.science-metrix.com/pdf/SM_2004_008_SSHRC_Bibliometrics_Social_Science.pdf

Yann Cadiou and Erik Arnold (2008), *Nordic Excellence, A Bibliometric Exploration of Common Nordic Research Funding Opportunities*, Nordforsk Policy Briefs 2008-5

Mike Thelwell (2007), Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4), pp. 1-18

[168] Mike Thelwall (2007), Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4), pp. 3-4

Van Raan, A.F.J. (2004). Measuring Science. Capita Selecta of Current Main Issues. In: H.F. Moed, W. Glänzel, and U. Schmoch (eds.). Handbook of Quantitative Science and Technology Research. Dordrecht: Kluwer Academic Publishers, p. 19-50

A.14. Webometrics

A.14.1. Description

Webometrics studies the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches.[169] Webometrics was triggered by the realisation that the web is an enormous document repository with many of these documents being academic-related. Moreover, the web has its own citation indexes in the form of commercial search engines, and so it is ready for researchers to exploit. One of the most visible outputs of webometrics is the ranking of world universities based upon their websites and online impact.[170]

A.14.2. Uses

A.14.2.1. Link Analysis

Link analysis is the quantitative study of hyperlinks between web pages. The use of links in bibliometrics was triggered by Ingwersen's Web Impact Factor (WIF), created through analogy to the Journal Impact Factor (JIF), and the potential that hyperlinks might be usable by bibliometricians in ways analogous to citations. The standard WIF measures the average number of links per page to a web space (e.g., a web site or a whole country) from external pages.

The hypothesis underlying early link analysis was that the number of links targeting an academic website might be proportional to the research productivity of the owning organisation, at the level of universities, departments, research groups, or individual scientists. Essentially, the two are related because more productive researchers seem to produce more web content, on average, although this content does not attract more links per page. However, the pattern is likely to be obscured. For example, some researchers produce highly visible web resources as the main output of their research, whilst others with equally high quality offline research attract less online attention.

Subsequent hyperlink research has introduced new metrics and applications as well as improved counting methods. Nevertheless, the wide variety of reasons why links are created and the fact that, unlike citing, linking is not central to any areas of science has led to hyperlinks rarely being used in an evaluative role. However, they can be useful in describing the evolution or connectivity of research groups within a field, especially in comparison with other sources of similar information, such as citations or patents.

A generic problem with link analysis is that the web is continually changing and seems to be constantly expanding, so that webometric findings might become rapidly obsolete. Moreover, the cumulative advantage (Matthew effect) of bibliometrics applies also to web links. On the web, a few pages attract millions of links whereas hundreds of millions of pages attract one or none. This imbalance is due to the fact that when a new link is created it is more likely to target pages that already have many links. The mediating factor is search engines: people are more likely to know about pages that have many links to them because search engines use links to find pages and to rank them. Hence, pages with many links to them are more visible.

A.14.2.2. Web Citation Analysis

A number of webometric investigations have focused not on web sites but on academic publications, using the web to count how often journal articles are cited. The rationale behind this is partly to give a second opinion for the traditional ISI data, and partly to see if the web can produce evidence of wider use of research, including informal scholarly communication and for commercial applications. A number of studies have

---

169 This section draws heavily on Mike Thelwall (2007), Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4), pp. 3-4

170 www.webometrics.info

shown that the results of web-based citation counting correlates significantly with ISI citation count across a range of disciplines, with web citations being typically more numerous. Nevertheless, many of the online citations are relatively trivial, for example appearing in journal contents lists rather than in the reference sections of academic articles. If this could be automated, it would give an interesting alternative to the ISI citation indexes.

In conclusion, comparing the advantages and disadvantages of webometrics, it seems that it is unlikely to replace traditional bibliometrics but can be useful for several other purposes.

- It can be used for fast pilot studies to identify areas for follow-up systematic bibliometric analyses
- It can be used to assess the extent to which researchers are successful in publicising their work online, given that this is an important activity. Third, it can be used for relational analyses of communication in disciplinary or geographic areas of science.
- Finally, its methods can help the analysis of Web 2.0 and online repositories for social sciences and humanities research goals.

A.14.3. Strengths

There are three main appeals of webometrics in contrast to traditional bibliometrics. First, the web can be timelier than the ISI databases. A typical research project might get funded, conduct research, report findings and then submit articles to journals. The time lag between the start of the project and the publication of the results in a journal is likely to be at least two years. Hence ISI-based bibliometrics is inevitably always retrospective, describing the research of years ago. In contrast, a research project might start by publishing a web site and could therefore be analysed with webometrics long before its research is published.

The second advantage of the web is that it contains a wide range of scholarly-related artefacts, including presentations, patents, data, software and general web sites. Hence, webometrics is potentially able to gather a wide range of evidence of research impact or connections.

Finally, the web is free to access for all web users and so it potentially opens bibliometric-style analyses to those who could not access or afford ISI data.

A.14.4. Weaknesses

Research into webometrics has also revealed many shortcomings, some of which are related to its advantages. First, the web is not quality controlled, unlike the ISI publication lists. Hence, web data tends to be of lower quality, which means that webometric results are normally indicative rather than providing robust evidence.

Second, web data is not standardised and so it is difficult to extract all except the simplest data (e.g. link counts). In particular, it is difficult to separate out the different types of publication. For example, there does not seem to be a simple way to separate out web citations in online journal articles from those in online course reading lists. Therefore, webometric results (e.g., link counts, web citation counts) tend to be the total of a mix of sources with variable value.

Third, although web data can be very timely, it can be impossible to find the publication date of a web page and so webometric results typically combine new and old web pages into one data set.

Finally, web data is incomplete in several senses and in arbitrary ways. Although some academic articles are freely available online, the majority are not. Similarly, some researchers and research groups maintain extensive and comprehensive web sites but others do not. Hence, the results reflect the web, which in turn is a very partial reflection of the activities of research.

A.14.5. Quality Control

Quality issues are the main weakness of webometric indicators. Unless more sophisticated methods and indicators are available, webometric studies are by necessity limited and indicative only.

However, webometrics is in a state of flux. New developments happen virtually daily. Hence, new methods and indicators of higher quality could be available soon. For this reason, we suggest the monitoring of developments in webometrics as the most promising measure of quality control.

A.14.6. References

Lennart Björneborn and Peter Ingwersen (2001), Perspectives on webometrics. *Scientometrics*, 50, , pp. 65-82

Lennart Björneborn and Peter Ingwersen (2004), Toward a basic framework for webometrics, *Journal of the American Society for Information Science & Technology*, 55(14), pp. 1216-1227

Mike Thelwall, Liwen Vaughan and Lennart Björneborn (2005). Webometrics, *Annual Review of Information Science & Technology*, 39, pp. 81-135

Mike Thelwell, Bibliometrics to Webometrics, *Journal of Information Science*, 34 (4) 2007, pp. 1-18

A.15. Meta-Evaluation

A.15.1. Description

A meta-evaluation is a systematic procedure for organizing and aggregating findings from a series of evaluations (sometimes also called evaluation synthesis). It brings together existing evaluation studies, performed by different people at different places and at different times, assesses them, and uses them as a data base for answering specific questions[171]. It can answer different kinds of questions – about overall programme effectiveness, about specific versions of the programme that are working particularly well or poorly or about how to organise future evaluation studies to provide more useful information about a programme. Meta-evaluation also involves an evaluation of the quality of the series of evaluations used and its adherence to established good practice in evaluation.

Meta-evaluation can integrate administrative data and findings from studies with either qualitative or quantitative emphasis. The quantitative synthesising of findings of experimental or quasi-experimental studies is normally called meta-analysis.[172] The two basic meta-analytical approaches are a) computing an average effect size[173] or b) conducting a combined significance test[174]. Meta-analysis is particularly important in medicine.

Since R&D evaluations rarely meet the assumptions or contain sufficient information to allow the use of meta-analytical procedures, it is relatively uncommon to use meta-analytical techniques in R&D evaluation, However, this does not mean that meta-evaluation or some systematic review is not possible in R&D evaluation. A careful narrative review, explicit about its analytic procedures, can be extremely valuable. Narrative reviews of collections of evaluation studies may frequently identify methodological weaknesses of certain broad types or groups of studies in a particular topic area. The evaluator will then need to consider these points in deciding whether or not to include these studies in the meta-evaluation and, if they are included, in interpreting findings from these studies.

Techniques of meta-evaluation and systematic review are important in some domains. Strikingly, they are barely used in R&D policy. Our own meta-evaluation of the EU Framework Programmes[175] is a rare exception. As is the study by Kingsley and Klein[176] who show how meta-evaluation of structured case studies can be used to build a more systematic understanding of the effects of interventions. By and large, R&D evaluations also do not use techniques that are cumulative. Rather, they try in every study (or are asked to try) to provide fresh evidence for all the steps in a chain of causation from intervention to long-range social and economic impacts.[177]

---

[171] United States General Audit Office GAO (1992), *The Evaluation Synthesis*, p. 8

[172] Robert Rosenthal (1991), *Meta-analytical Procedures for Social Research*, Sage

[173] When one compares a treatment group to a control group, a common definition of effect size is the difference between the two group averages, expressed in terms of the control group's standard deviation

[174] The relationship between sample size and the power of a statistical test is well known: the larger the sample size, the more likely that a certain effect will be detected as statistically significant. When multiple independent studies all compare two treatments that are similar across studies and the group differences are tested statistically in each instance, one strategy for drawing a single 'grand' conclusion from these results involves combining the separate significance tests into an overall test of a common null hypotheses. A number of procedures using this idea have been suggested by Rosenthal (1991)

[175] Erik Arnold, John Clark and Alessandro Muscio (2005), 'What the evaluation record tells us about Framework Programme performance', *Science and Public Policy*, Vol 32, No 5, pp385-397

[176] Gordon Kingsley and Hans Klein (1998), 'Interfirm collaboration as a modernization strategy: a survey of case studies,' *Journal of Technology Transactions*, Vol 23 No 1

[177] Erik Arnold (2004), 'Evaluating research and innovation policy: a systems world needs systems evaluations,' *Research Evaluation*, April

A.15.2. Uses

The methodology is best applied to areas in which there is a base of evaluation information. Policy concerns for which there is little or no existing study information cannot be satisfactorily investigated. Therefore, the methodology will not be appropriate for new programmes where evaluation studies have not been completed and not existing information base has applicability.

Box 26 **A meta-evaluation: the effects of promoting research partnerships**

The CTI, the Swiss Innovation Agency, has been promoting research partnerships between firms and public research organisations since the 1980s.

In a meta-evaluation published in 2006, fourteen evaluations studies conducted up to 2005 were assessed and synthesised. In a first step, the studies were classified by evaluation questions (effectiveness, relevance, implementation, goal attainment etc.) and by methodological characteristics. Then they were assessed against pertinent evaluation standards set up by the Swiss Evaluation Society SEVAL.

The meat-evaluation found that the evaluations were mainly internal, qualitative, confidential/unpublished, ex post and situated at micro-level. Only research organisations and firms that carried out large numbers of CTI projects were evaluated regularly. However, these evaluations focused more on organisations' research capacity than on the quality and impact of CTI funding. There were differences in quality among the studies, with most evaluation standards fulfilled fairly to very well.

In a second step, the outcomes the studies reported were synthesised. As most of the studies were case studies, the net effect of CTI funding could not be estimated. However, by listing effects at firm-level and at the level of research organisations, strong evidence both of the breadth of effects and of typical outputs and outcomes could be demonstrated.

Barbara Good (2006), *Technologie zwischen Markt und Staat*. Die Kommission für Technologie und Innovation und die Wirksamkeit ihrer Förderung (The CTI and the impact of its funding), Zürich/Chur (with English summary)

A.15.3. Strengths

Meta-evaluations provide relatively inexpensive and comprehensive information, as no primary data collection is required.

Another strength is that the meta-evaluation can increase the power of the individual study finding. Confidence in a number of well-done studies with the same finding is greater than in the finding of a single well-done study.

By drawing together information about a specific question from a disparate number of completed evaluation studies, the meta-evaluation also creates a common knowledge base about a particular topic. It clearly sets out what is known and what is not known about the topic, thus enabling programme managers and evaluation units to determine where they might best commit future evaluation resources. Hence, a particularly valuable feature of the synthesis is the identification of remaining unanswered questions.

Last but not least, meta-evaluation can guide and improve future evaluations. For example, the lessons learnt in the various evaluations of the EU Framework Programmes are certainly useful input for future evaluations of the Framework Programmes (see Box 27).

A.15.4. Weaknesses

The main limitation of the meta-evaluation methodology stems from its reliance on existing data and evaluation studies. Even when a substantial information base is available, the meta-evaluation is limited in that it can answer questions only to the extent that the existing studies have addressed them.

Moreover, the meta-evaluation can only be as good as the evaluations it is based on. Poor reporting also limits the evaluation synthesis. Procedures may have been described in so brief a manner that judgements cannot be made about a study's methodological adequacy.

Finally, the meta-evaluation is only as current as the studies it synthesises and aggregates.

Box 27 **Evaluations of EU Framework Programmes – techniques in use and key criticism**

A 2002 study provides an interesting meta-evaluation of 28 socio-economic impact studies of aspects of the EU Framework Programmes. The studies were classified according to their way of dealing with data collection, data analysis, methodological stance, level of analysis, stakeholders analysed, methodological reflexivity, policy recommendations. Key criticisms were

- More attention is given to short- than long-term effects, because these are easier to identify and measure
- Data are often one-dimensional and not triangulated against other sources
- Data collected are under-analysed (often because of the haste in which evaluations have to be prepared)
- The importance of quantitative information in the evaluations is often over-stated
- Non-participants' views are neglected
- Efforts to trace social – as opposed to economic – effects are underdeveloped
- Lack of coherent and comprehensive impact methodology
- Lack of coherent data collection in the programmes, either for monitoring or for evaluation purposes

The report found that there was room both for improvement within known parameters and for methodological development, to enable better evaluation of programmes.

Luke Georghiou, John Rigby and Hugh Cameron (eds) (2002), Assessing the Socio-Economic Impacts of the Framework Programme, June

A.15.5. Quality Control

It is advisable to include all existing evaluation studies in the meta-evaluation, also those that are difficult to get hold of or are unpublished. Research has shown that published research reports tend to have more significant positive findings than unpublished reports.

There are at least four specific steps we recommend for organising a systematic search for evaluation studies. To begin with, the agency who commissions the study or administers the programme should be asked about relevant studies. Often, they not only know about relevant studies, they also have them in their archives. A second step would be to search for relevant evaluations on the Internet and in appropriate databases. A third would be to examine the lists of references at the end of key evaluation reports. Finally, a fourth step is to ask knowledgeable colleagues around the country.

Major weaknesses of study design, conduct, analysis, or reporting that affect the reliability or validity of each study's findings must be identified and considered in using the study and placing confidence in the study findings. Hence, although the question what constitutes a 'good' study is debatable, it seems reasonable that all evaluation studies to be included in a meta-evaluation should be assigned against basic standards for research design, conduct, analysis, and reporting. It may prove useful to use (some of) the evaluation standards defined and promoted by professional evaluation associations, for example the Program Evaluation Standards defined by the

Joint Committee on Standards for Educational Evaluation or the Canadian Evaluation Society.[178]

A.15.6. References

Erik Arnold (2004), 'Evaluating research and innovation policy: a systems world needs systems evaluations,' *Research Evaluation*, April

Erik Arnold, John Clark and Alessandro Muscio (2005), 'What the evaluation record tells us about Framework Programme performance', *Science and Public Policy*, Vol 32, No 5, pp. 385-397

Luke Georghiou, John Rigby and Hugh Cameron (eds), *Assessing the Socio-Economic Impacts of the Framework Programme*, June

Gordon Kingsley and Hans Klein (2002), 'Interfirm collaboration as a modernization strategy: a survey of case studies,' *Journal of Technology Transactions*, Vol 23 No 1, 1998

United States General Audit Office GAO (1992), *The Evaluation Synthesis*

Ray Pawson (2002), 'Evidence-based policy: In search of a method,' *Evaluation*, Vol 8 No 2, pp157-181

Robert Rosenthal (1991), *Meta-analytical Procedures for Social Research*, Sage

[178] A list of links to professional standards and codes of ethics worldwide can be found on the website of the Swiss Evaluation Society http://www.seval.ch/en/standards/internstandards.cfm

A.16. Organisational Review

A.16.1. Description

The research and innovation evaluation community has no strong traditional of organisational review. Its roots are strongly in the 'research on research and innovation' community's focus on the generation and use of knowledge and very little work has been done on the specifics of organising for research and innovation funding or management. For many evaluation purposes, it is not relevant to review the agencies responsible for delivering services. If they produce results they are effective; if not, then not. However, when the question of choice and efficiency of agencies becomes an issue for evaluation, then organisational review becomes more significant.

There is within the management literature a certain tradition of reviewing the way in which not-for-profit organisations are run. This looks at **how** they do things rather than whether they are effective in what they do. It tends to be highly normative in relation to 'best practice, – but that practice itself is more likely to be defined in quality standards and other less formal norms than having been found to be 'best' on the basis of studies or evidence.

A prominent example is the Minnesota Council of Nonprofits, which has developed a guide for nonprofit board members, managers an staff that sets out principles for

- Role in society
- Governance
- Planning
- Transparency an Accountability
- Fundraising
- Financial Management
- Human resources
- Civic engagement and public policy
- Strategic alliances
- Evaluation

Other approaches, such as that of Peter Drucker, stem more directly from the corporate management tradition. His self-assessment tool focuses on five questions

- What is our mission?
- Who is our customer?
- What does the customer value?
- What are our results?
- What is our plan?

Numbers of other approaches exist, with varying degrees of clarity about the degree to which non-profit organisations' performance needs t be managed differently from that of for-profit organisations. The special characteristics of non-profit organisations tend to involve them in non-market types of planning, more complex issues of governance and specialist functions such as fundraising to which commercial organisations are not exposed.

Research and innovation funding agencies, in particular, perform a number of specialised processes rarely found even in the non-profit sector. We address these in the 'Benchmarking and Comparison' section of this Chapter.

A.16.2. Uses

Because agencies are rarely exposed to market competition, there may be a need to test the degree to which they take up what is considered 'good practice. Identifying failures in this respect can alert the evaluator to potential problems, which can then be more closely examined. For example, a lack of staff training cold be a signal that key processes may be out of date and ineffective.

A.16.3. Strengths

Organisational review provides some evidence about organisational efficiency. For example, reviewing the way human resources are managed allows the evaluator to understand whether staff skills are developed and refreshed in ways likely to underpin administrative efficiency, process improvement and the ability of the organisation to learn. It can provide evidence about incentives (or lack of incentives) for personal performance and therefore for the organisation's ability to reward innovation and improvement. Similarly, it can allow the evaluator to establish whether it is reasonable to suppose that other functions are being well performed.

A.16.4. Weaknesses

The key weakness of organisational review – which is shared with a great deal of the business management literature – is the lack of rigorous evidence for a connection between individual process or functional behaviours and overall organisation effectiveness and efficiency. It is tempting to look for proxies, such as ISO certification of quality and other standards, but there s disagreement within the quality community, for example, whether a standard such as the ISO 9001 quality standard (which has been adopted by some funding organisations) increases or decreases efficiency.

A.16.5. Quality Control

Given the gap lack of evidence about the links between what is conventionally seen as good practice and efficiency or effectiveness, it is difficult to quality control organisation review ay the overall level.

A.16.6. References

Peter Drucker (1998), *The Drucker Foundation Self-Assessment Tool*, Jossey-Bass

Minnesota Council of Nonprofits (2005), *Principles and Practice for Nonprofit Excellence: A Guide for nonprofit board members, managers and staff*

A.17. Benchmarking/Comparative Study

A.17.1. Description

The practice of benchmarking performance is well established in industry. It may involve groups of companies in the same sector in comparing process variables, such as 'How many minutes does it take to set up a certain type of job on a 5-axis CNC milling machine?' Some such clubs operate by holding an anonymous database so that those who contribute data can se how their performance compares to others' without t the same time being able to identify them. Where data are especially sensitive, benchmarking may be done across different sectors, so that common processes can be compared without at the same time revealing information to competitors.

Except to the extent that international statistics (such a the OECD R&D statistics) may be regarded as benchmarks, there s little benchmarking tradition in research and innovation policy. European projects such as ERAWATCH and the Inno Policy Trend Chart contain comparative information about research and innovation funding schemes and organisations, but do not do benchmarking. The only process benchmarking of which we are aware was a project Techopolis ran with the Research Council of Norway, TEKES and NUTEK Teknik (now VINNOVA) in the Nordic region during the 1990s, which benchmarked innovation programme development and implementation. This was subsequently extended in work for the European Parliament to consider the administrative burden imposed on applicants to the EU Framework Programme of Research and Technological Development.

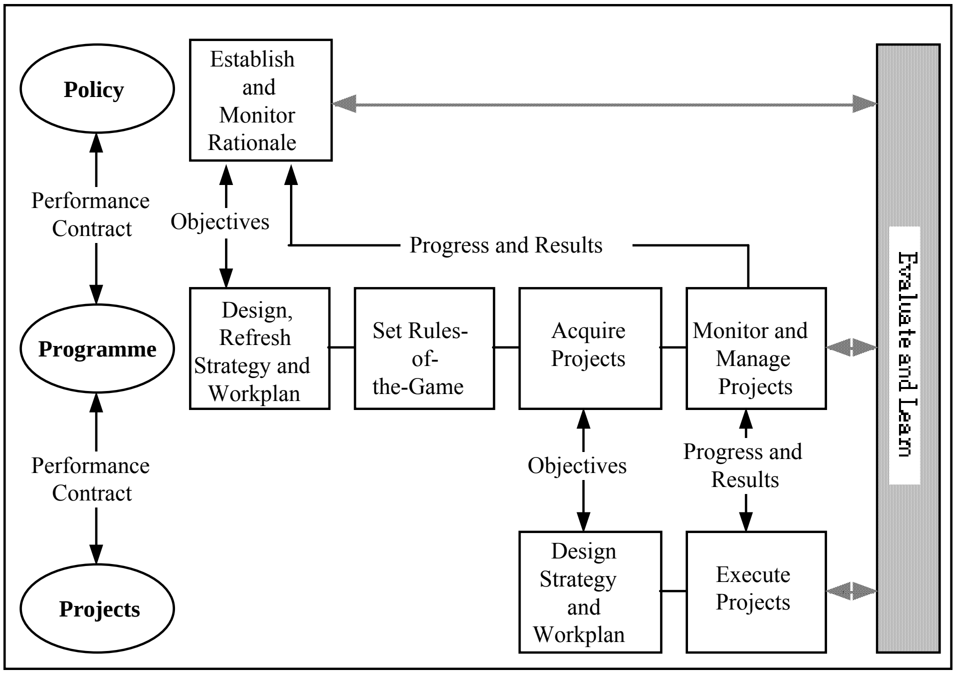
The benchmarking was done on the basis of a simple process model of an innovation agency (Figure 41), which is turn was broken down into a large number of smaller steps, comparable among organisations. Time, cost and process description data were then collected for each.

A.17.2. Uses

The main use for the Nordic benchmarking exercise, which was initiated by the agencies themselves, was to launch a series of learning exercises where benchmark data suggested there were bottlenecks or important differences in practice, which the agencies explored with a view to improving their processes. The data were subsequently re-used and compared with process data from the European Commission, allowing objective testing of complaints about the supposed administrative inefficiency of the Commission. That study showed that some EU-level process were both longer and more burdensome than their counterparts at national level. Some of the inefficiency was caused by the controls imposed by the member states. Others were the result of complex EU financial controls.

The data built up have been incrementally added to in a series of evaluations and have provided some simple rules of thumb for administrative efficiency in different types of funding organisation. Participants generally value the qualitative insights and learning opportunities more than the 'hard' data.

Figure 41 Innovation Agency Process Model



A.17.3. Strengths

The strength of this type of detailed process benchmarking is that it allows both evaluation of efficiency and process improvement. Being activity-driven the data are more comparable across organisations than the global cost-of-administration data sometimes published by agencies.

A.17.4. Weaknesses

Many measurements, however, are difficult to make because agencies tend not to require their employees to fill in time sheets. Thus there are no data about how the biggest resource at their disposal is actually used. The lack of benchmarking clubs in the field also means that such data as can be obtained are spotty and sometimes quite old. Unless process models are very explicit, there is a risk that incompatible data are collected.

A.17.5. Quality Control

Process benchmarking is very labour intensive and is demanding of the time of those being benchmarked – especially in situations such as research and innovation funding where there is no established benchmarking tradition. The main quality control opportunities are attention to detail on the part of those who collect the data and the active participation of those benchmarked in the analysis and exploitation of the data.

A.17.6. References

James Stroyan, Erik Arnold and Jackie Senker (1998), *Comparative Study on Administrative Burdens and Rules of Procedure between the EU Research Programmes and those of the Individual Member States*, (IV/98/06), Report to the STOA office of the European Parliament, Brighton: Technopolis

A.18. History

A.18.1. Description

The explicit use of history in research and innovation evaluation is unusual. It can become especially important in organisational evaluation because it helps explain why things are as they are. Thus, in the evaluation of the Research Council of Norway, one of the background reports gave an historical account of he role of the previous research councils, the growing dissatisfaction with their performance and the political and policy reasons why they were merged to form the single Research Council of Norway. Analysis of documents and interviews with stakeholders from that time clarified the failures involved with the old councils. The evaluation, almost ten years after the merger, found that these failures were still present. This gave a clue that at least some of the key problems were in the governance system and that these systems problems could not be addressed by further changes inside the research council. Rather, they were problems of context that had to be solved – if at all – by the ministries responsible.

An historical analysis in the evaluation of the Austrian research council (FWF) and innovation foundation (FFF) was inspired by the RCN study. By exploring the policy problems in the 1960s that led to the peculiar status of these foundations – funded by the state but governed by their beneficiaries – it became possible to see why that had failed to modernise since then. As a result, several specialised organisations had grown up to take on the new functions that in others systems had been absorbed by the research and innovation funders, hence the Austrian system became very fragmented. As a result of the evaluation, the innovation foundation was merged with some of the newer funders to form an agency properly under the control of a ministry.

These historical accounts were useful because they made sense of the characteristics of the organisations being evaluated and they provided causal hypotheses that could be use both to explain problems in performance and to propose solutions. Unfortunately, the historical materials are respectively in Norwegian and German. References to the relevant synthesis reports are given here.

A.18.2. References

Erik Arnold, Stefan Kuhlman and Barend van der Meulen (2001), *A Singular Council: Evaluation of the Research Council of Norway* (Royal Norwegian Ministry of Research, Education and Church Affairs: Oslo

Erik Arnold (ed) (2005), *Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF)* Synthesis Report, Vienna: BMVIT

A.19. Summary of Evaluation Methods

Figure 13 gives an overview of the strengths and weaknesses of the evaluation methods discussed in the previous sections, while Figure 14 discusses their general applicability.

Figure 42 Overview of evaluation methods

Method	Strengths	Weaknesses
Macro-Economic modelling	• Based on established principles • Susceptible to rigorous testing • Provide consistent accounts of the importance of R&D for the macroeconomy	• Dependent on large quantities of high-quality data • Resource-intensive, skill-intensive and time consuming • Difficult for non-specialists to understand
Micro-economic modelling	• Provide a structured, formal framework for project/programme evaluation • Susceptible to rigorous testing	• Dependent on large quantities of high-quality data • Resource-intensive, skill-intensive and time consuming • Difficult for non-specialists to understand • Risk that important externalities will be excluded
Cost-benefit analysis	• Offers a systematic and comprehensive evaluation procedure • Assumptions can be transparently presented, little specialist knowledge required for understanding	• Quantification of some costs and benefits can be very difficult, and 'guesses' can yield misleading results • CBA models are often very sensitive to realistic changes in the assumptions used
Spillover analysis	• The idea of spillover is central to most justifications for state intervention • Spillover studies have demonstrated the huge social value of R&D • They have also helped identify the mechanisms of spillover, allowing policy to focus in improving these	• Ambiguous definition of 'spillover variable' • Frequent need to resort to proxies • Some studies have been found to produce results that are unstable
Survey of beneficiaries	• Economical • Provides understandable and credible results • Can accommodate comparison groups and counterfactuals • Can be representative • Supplies high numbers necessary for statistical analysis	• Subjective • Positive bias • Does not capture richness of individual projects and experiences
Peer review	• Low-cost, fast-to-apply • Well-known, widely accepted • Versatile	• Peers' independence • Problematic when highly innovative or interdisciplinary work is to be assessed
Panel review	• 'Method of last resort' when budget, time and methods constraint are presents	• Selection of suitable peers/composition of panels
Case studies	• Comprehensive view of phenomenon under study/richness of detail • Can be used for theory-building • Easy for policy makers to read and understand	• Qualitative evidence generally considered less persuasive than quantitative evidence • Not representative, not generalisable to population

Method	Strengths	Weaknesses
Historical tracing	• Produces interesting and credible study • Sheds lights on process dynamics • Suited for the tracing of social impacts	• Attribution and establishment of cause and effect difficult • Time-consuming and costly
Network analysis	• Different and complementary perspective • Can bring surprising results • High comparability	• Theoretical paucity • Largely descriptive • Data collection time-consuming
Prospective studies (foresight and technology assessments)	• Broadens scope of evaluation studies	• Time-consuming and costly • Attribution and establishment of cause and effect difficult
Bibliometrics and patent analysis	• Objective, credible • Applicable to variety of evaluation topics • Easily scalable • Results relatively easy to understand • Places no burden on people being evaluated	• Narrow view of outputs, limited to publications and patents • Narrow view of impacts, limited to citations • Publishing and citation patterns differ across fields • Weak in the social sciences and humanities • English language bias
Webometrics	• More timely than bibliometrics • Covers wide range of scholarly produced artefacts • Free access to web data, affordable	• Web not quality-controlled, hence results indicative rather than robust evidence • Difficult to separate out different types of publications/ sources of different value processed in webometric results • Web-data incomplete and arbitrary/ majority of academic articles not freely available online
Meta-evaluation	• Inexpensive, comprehensive • Increases the power of individual study • Can guide an improve future evaluations	• Limited to answering questions already addressed by existing studies • Hampered by quality of existing studies • Only as current as existing studies
Organisational review	• Useful tool for exploring the causes of good or bad performance in a=management and administration • Provides opportunities to study organisational options for performance improvement	• No basis in science or evidence-based study • Highly dependent on the evaluator's experience and understanding of the specific context of research and innovation funding
Benchmarking	• Useful tool for identifying performance standards where these cannot be deduced from first principles • Can trigger organisational learning and process improvement	• Poor availability of data in the field of research and innovation funding
History	• Identifies causes and behaviour drivers not necessarily obvious from a static analysis • Makes 'sense' of performance in ways that make sense to evaluators and evaluation readers	• Requires respect for historical methods and standards of evidence • Sources can be had to find

Figure 43 General relevance of evaluation techniques

Method	Ex ante	Process	Ex post		
			Outputs	Outcomes	Impacts
Macroeconomic modelling	Y				Y
Microeconomic modelling			Y	Y	
Cost-benefit analysis	Y			Y	Y
Statistical approaches			Y	Y	Y
Spillovers				Y	Y
Survey of beneficiaries	-	Y	Y	Y	-
Peer review	Y	Y	Y	-	-
Panel review	Y	Y	(Y)	(Y)	(Y)
Case studies	-	Y	Y	Y	Y
Historical tracing	-		Y	Y	Y
Network analysis	-		Y	(Y)	(Y)
Prospective studies (foresight and technology assessments)	Y	(Y)	-	-	-
Bibliometrics and patent analysis	Y		Y	-	(Y)
Webometrics	Y		Y	-	(Y)
Meta-evaluation	-	(Y)	Y	Y	Y
Organisational review		Y			
Benchmarking	Y	Y	Y		
History		Y			

A.20. Authors' Views and Assessment of Methods

The repertoire of techniques used in R&D evaluations has been evolving over the past 20-30 years. Scientometrics and social network analysis are important extensions to the repertoire. It is increasingly argued[179] that at least some of the techniques routinely used and the associated styles of evaluation have reached diminishing returns in terms of generating new knowledge. For example, repeated surveys of beneficiaries in pre-competitive, collaborative R&D programmes produce similar results.[180]

Evaluating R&D policies and interventions is complex – especially when objectives are pitched at the level of changing important characteristics of research and innovation systems (as is the case for the EU R&D Framework Programmes). The overall repertoire of indicators and evaluation methods used needs to be applicable at the systems level as well as informing us about lower-level phenomena. Indeed, there appear to be especially large opportunities to improve methods for evaluating effects at the meso-level of policies and systemic bottlenecks and at the macro level of the whole systems. More specifically, areas of potential innovation and improvement include

- Further articulation of social network analysis, with better linkage to research and innovation processes. At present, social network analysis in R&D – while interesting – is mainly descriptive. Linking meso- and macro-network topologies to micro-level behaviour and strategy is needed, if social network analysis is to become more explanatory and potentially predictive
- Better connection between macro- and micro-economic modelling and simulation, feeding back results from micro models to the macro level
- Longitudinal studies of institutions, 'knowledge value collectives'[181] actors and actor groups, as distinct from interventions, in order better to understand the implications of the 'project fallacy', to understand the choices faced by intended beneficiaries and how they behave in relation to programmes and policies
- Techniques of meta-evaluation and systematic review appear to be an area where genuine innovation is needed in R&D evaluation practice, rather than one where a great deal of 'good practice' exists to be imitated
- Methods to trace social, as well as economic, effects of policies and interventions. The Finnish VINDI project as well as the 'tracking-back case studies' we conducted in the humanities could show a way

Against the background of the discussion conducted above, we can formulate the following general recommendations:

- Meta-evaluation and other systematic reviews of evidence should be used to provide inputs to policymaking, especially at systems level.
- Since there is significant scope for policies and programmes to have unexpected outcomes, evaluation models should find a balance between focusing only on testing the programme logic set out by the policy or programme designers, on the

[179] Luke Georghiou (1997), 'Issues in evaluation practices in OECD countries', in OECD, 1997; Erik Arnold (2004), 'Evaluating research and innovation policy: a systems world needs systems evaluations,' *Research Evaluation*, Volume 13 Number 1, April

[180] Erik Arnold, John Clark and Alessandro Muscio (2005), 'What the evaluation record tells us about Framework Programme performance', *Science and Public Policy*, Vol 32, No 5, pp385-397

[181] Barry Bozeman and Juan Rogers, 'A churn model of scientific knowledge: Internet researchers as a knowledge value collective,' *Research Policy*, Vol 31, 2002, pp 769 - 794

one hand, and a results-oriented approach that tries to uncover all possible outcomes.

- Some methods are more appropriate at certain stages of the policy or programme life cycle than at others, so methods should be selected in part for their appropriateness. For example, cost-benefit analysis is a very useful prospective technique for thinking in a structured way about the potential impact of an intervention, but is difficult to operationalise ex post in order to obtain a reliable estimate of actual impacts. These difficulties also underlie problems in making comparative estimates of the effects of different interventions ex post. On the other hand, broad economic impact estimates – while partial – can be useful in explaining the effects of policies and interventions.
- Panel reviews – where respected people are asked to make judgements about things in which they are not necessarily expert – need to be carefully managed where they cannot be avoided, as
  - There is a difficulty of assembling panels that exclude people from major beneficiary organisations
  - They are liable to being exploited to promote disconnected political or policy agendas.
  - In contrast, peer review by scientists of scientific aspects of programmes and policies still has an important role to play in both ex ante and ex post evaluation.
- More generally, R&D evaluation methods in use are often individually not fully reliable, and should generally be used in combination so that the evaluator can 'triangulate' between different methods and look for convergence
- R&D funders should aim to incorporate opportunities for methodological experimentation and innovation in their evaluation strategies and allow for experimental funding in order to improve techniques because the opportunities for experimentation will be limited and the rate of innovation will continue to be slow, as long as evaluation budgets remain tied closely to the performance of 'operational' evaluations

Appendix B - Summary of the Innovation Strategy for Competitiveness

Figure 44 The Strategy

Challenges	Next steps
Human Capital • Establishing a system of flexible life-long learning covering not only initial education but also subsequent re-skilling and re-training • Developing an integrated system of higher education and qualifications, in order to allow people not only to acquire high-level skills and qualifications but also to navigate non-traditional courses through the education system and enable mobility • Develop a system of technical and vocational education based on skills relevant to work • Modernise university curricula so that they are relevant to national and international needs • Strengthen and increase the throughput capacity of the system of higher education and associated qualifications	1. Promote the development of a life-long-learning system a Establish a system for certifying work skills, based on industrial needs b Strengthen mechanisms that make the education system more transparent and coherent through a system of transferable credits open to all universities, consolidating systems for accrediting technical raining and skills, and encouraging interoagency coordination c Strengthen information systems to provide better feedback to the education system by implementing a higher education information system and providing information about the design and evaluation of public policy in the field of human capital 2. Ensure the quality and relevance of education and training a Strengthen the quality assurance system in higher education at the level of institutional, vocational and technical education and publish the results b Transform the AFD university funding system into one of performance contracts, over time making the AFD system competitive and in the short term identifying the part of the AFD that goes to research and postgraduate education and funding this through ad hoc programmes c Strengthen the role of competitive funding in ensuring the quality and relevance of higher education, by strengthening the Ministry of Education's evaluation capabilities, using performance agreements in all tpes of financing and opening competitions for all funding to all accredited institutions d Improve the quality of training institutions through the competitive provision of standardised training modules and gradually making public funding conditional on the provision of modules based on sectoral standards and training outcomes e Ensure that the different levels of technical education and training correspond to the needs of the labour market 3. Increase the provision of training in technical subjects and to lower-income groups a Establish a system of credits applicable to the entire higher education system in order to provide access for those from low-income groups b Provide more grants to students from low-income families c Focus support for training on low-income groups by reorientating training towards the needs of the poorest and making more effective use of the SENSE tax
S&T System • Finding a new balance between curiosity-driven and "relevant" research • Deciding on which great questions to focus publicly funded research • Designing a funding system to meet the challenges – tackling the full costs of research, assuring quality through peer review and respecting national priorities while being internationally networked and multidisciplinary • Strengthening research capacity	1. Increase and strengthen the amount of fundamental – especially strategic –research undertaken a Increase funding for fundamental research and provide incentives for it to be done in strategically relevant areas b Communicate strategic priorities to the scientific community 2. Change the system for funding science a Consolidate the role of CONICYT as the agency responsible for developing the science base in Chile b Strengthen public funding of the science base c Fund broad, interdisciplinary research networks d Provide funding that meets the full costs of doing science 3. Ensure there is adequate scientific manpower to meet national needs a Ensure sufficient researchers are trained

Challenges	Next steps
	b Strengthen the grant system, focusing it on CONICYT c Quality-assure postgraduate education d Stimulate and support the development of doctoral programmes in strategically important subjects e Increase the number f new research posts in the system f Attract foreign researchers in order to meet needs, especially in strategically important areas
Business innovation • Incorporating new knowledge into production by raising the level of R&D • Encouraging the diffusion of new technologies • Supporting innovative entrepreneurship	1. Strengthen the development of clusters a Establish cluster route maps b Stimulate the creation of research capacity relevant to the clusters c Attract Foreign Direct Investment into high-potential clusters 2. Incorporate new knowledge into the productive process a Support the development of R&D consortia b Increase the number of firms and types of innovation that may be subsidised from public funds c Increase the input of innovators into company management through training and promoting mobility into companies d Provide better tax incentives for business R&D e Strengthen the protection and use of intellectual property 3. Strengthen the diffusion of technology a Make a network of technology institutes available to support industry b Provide a technology and R&D brokerage service c Create world-class centres for the diffusion of company-relevant technologies d Establish a technology diffusion programme significantly to increase the number of companies that innovate 4. Support innovative entrepreneurship a Establish more efficient business incubators b Promote spin-offs from existing companies c Encourage greater availability of seed capital d Encourage greater availability of venture capital e Develop instruments to encourage business angels f Reduce the way bankruptcy currently tends to prevent entrepreneurs from trying again with new business ideas
Institutions • Reduce fragmentation within the set of state institutions that promote innovation • Create or strengthen the links between advisory and management boards and the national innovation strategy • Strengthen intermediate institutions for implementation and monitoring of innovation policy	A National Issues 1 Establish institutional roles based on the principle of division of labour 2 Improve the governance of the state institutions involved with innovation, making a clear distinction between policymaking and implementation 3a Consolidate and strengthen the role of the CNIC a Define it as a permanent organ of the state with the task of setting the framework for public policies for innovation, research and the development of human capital b Define the CNIC's responsibility to cover all the national innovation system, not solely programmes funded by the FIC mining tax c Charge the CNIC with designing and implementing a system of evaluation for the national innovation system d Charge the CNIC to support the development of regional innovation strategies e Require the CNIC to report annually on its activities and progress f Require the Executive publicly to develop an innovation strategy, using that of the CNIC as a point of departure

Challenges	Next steps
	g Reduce the number of people who are members of the CNIC and increase the use of specialists h Require the work of the CNIC periodically to be reviewed by a panel of international experts 3b Consolidate and strengthen the management of the national innovation system within the Executive a Clarify and elevate the juridical role of the inter-ministerial committee on innovation b This committee should comprise the ministers of economy, education and finance – plus any others the President may see fit to appoint c Give the committee a policymaking role in relation to the whole innovation system, not solely to the FIC, in order to promote holistic policymaking d Form public-private partnerships to implement cluster policies e Provide the committee with a strong secretariat 2c Give the regions a role in the definition of innovation policy 3 Strengthen the agencies responsible for executing the strategy a Clearly define the role of each agency b Their boards should be appointed on the basis of high level skills – these should not be political appointments c The agencies should have permanent consultative committees to advise on their programmes d The boards should hire the agency directors, according to the principles of the system of 'High-level Public Management' e Agree objectives for administrative efficiency with the agencies, in the light of the epected workload f Promote transparent decision-making by the agencies g Undertake a thorough evaluation of each agency every four years h Increase evaluation capacity within the agencies so that they systematically evaluate all their programmes i Arrange cross-membership of agency boards, to increase coordination 4 Order and strengthen the sub-system of human capital and science a Strengthen the leadership role of the Ministry of Education and create a position of Sub-secretary for higher education and science b Differentiate clearly within the Ministry of Education between policymaking and implementation c Redefine CONICYT's governance to become an agency of the Ministry of Education 5 Order and strengthen the sub-system of business innovation a Assign the sub-secretary of economy the lead role in setting innovation policy b Redefine CORFO's governance to become an agency of the Minstry of Economy c Provide a system of Technology Institutes to provide an adequate set of technological services d Reform the intellectual property system B The Regional Dimension a Support the development of regional innovation strategies, within the framework of the national one b Ensure the balanced presence of both science and innovation in the regions, based on agreements between the regions and the national levels c Develop regional institutions in a way consistent with policy d Implement differentiated policies based on the knowledge needs of the regions e Build institutional capacity in the regions

Technopolis Ltd
3 Pavilion Buildings
Brighton BN1 1EE
UK
T +44 1273 204320
F +44 1273 747299
E info@technopolis-group.com
www.technopolis-group.com