Approaches to Impact Assessment

The techniques appropriate to impact assessment necessarily depend, first, upon what the impact assessment is trying to achieve and, second, the data and other resources that can be assembled. As with most things, there is no universal and timeless “best practice”.

Since impact assessment is a policy-related activity, it operates under time and resource constraints to produce the best answer available within these constraints – not necessarily the best possible answer. As ever in policy work, if a decision has to be taken in six months, a two-year study will not answer the question. The obverse of that coin is that impact analysts could often be braver in describing – or sometimes even noticing – the limitations of their work.

This document discusses methods in relation to four types of impact analysis:

- Estimating or modelling impacts in money terms;
- Studies aiming to understand how impacts happen;
- Focusing on particular limited outcomes;
- Impact assessment for performance measurement.

1. Following the money

We traditionally divide economic analysis between macro- and micro-economic techniques and will discuss that distinction in this section. Macroeconomics tries to explain the behaviour of the whole economy, taking a rather aggregated approach to behaviour that tends not consider the level of the individual firm. Microeconomics, on the other hand, looks at the behaviour of sub-populations within the economy, normally at some level considering the individual firm.

Macroeconomic approaches used in impact assessment, however, are themselves of two broad types. One relies on national economic models of the type typically used by ministries of finance and national banks in monitoring and trying to manage the economy. These are relevant to **ex ante** impact analysis in the sense that they can be used to answer “what if?” questions about the **expected** effects of changed policies on the economy. What happens if we raise the interest rate? What are the effects of changes in wages? So they **model** the **expected** effects of policy change; but they do not **measure** the effects of such changes after they have happened. The do not help, in other words, with **ex post** impact assessment.

Other macroeconomic approaches are more focused, relying on various kinds of **production function**. A production function is used to decide how much of aggregate growth is the responsibility of changes in inputs (capital, labour…) and how much is due to changes in technology. In recent years, a lot of production function work has focused on total factor productivity, which expresses the effects of technology on overall productivity (rather than focusing on the productivity of individual factors, e.g. labour productivity). Production function approaches use case-specific data to estimate the effects of change, so they can be used or ex post impact assessment.

Microeconomic approaches do not attempt explanation at the level of the economy as a whole. Rather, they focus on inputs and outcomes for specific parts of the economy. Econometric techniques, both at the macro (e.g. production function) and the micro level can be useful for ex post impact analysis.

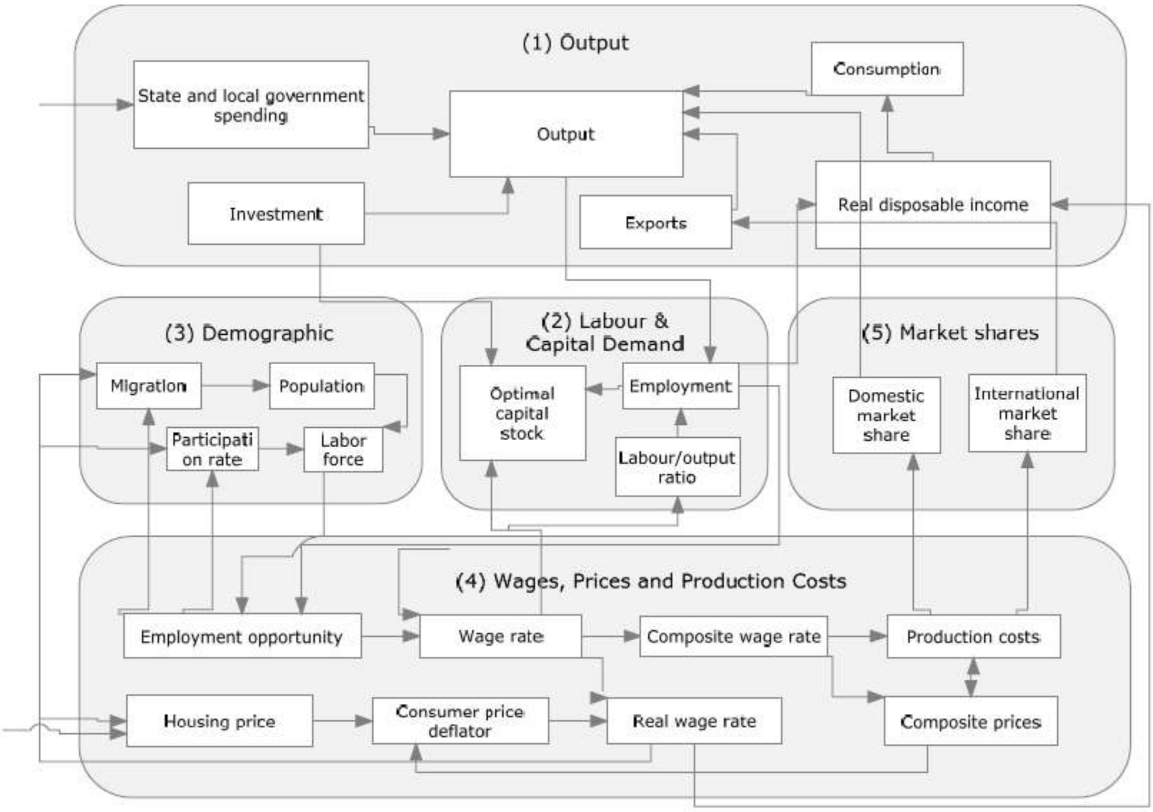
Computable general equilibrium models for ex ante impact assessment

Computable general equilibrium models (CGE) use equations to model the interdependence of the various parts of the economy. Typically modular (and therefore containing large numbers of sub-systems), they are based on economic theory and calibrated using observed economic relationships.

By solving large numbers of simultaneous equations, they aim to predict the state of the economy at equilibrium, i.e. once the shock of changing some of the economic variables has passed through the system and the economy has settled down. The analysis is therefore inherently **static**, i.e. it does not help consider development trajectories. CGE models are also completely context-dependent: they aim to model particular economies at particular points in time. There is no **general** empirical test of their validity, except with respect to economic theory. A positive aspect of the context dependence is that CGE models could be used to consider the potential effects of similar policy changes in different countries, though they appear to be little used in this kind of application.

Despite their complexity, the level of aggregation in CGE models is very high. In almost all cases, this means that there is no variable that represents “technology” (e.g. see Figure 1).

Figure 1. Key linkages in the RHOMOLO CGE model



Source: European Commission, Joint Research Centre

- **Causality**. CGE models are built on assumptions about causality, rather than generating results about causality. This is a key reason why they are useful primarily in ex ante analysis. At this high level of aggregation, little can be said about the mechanics of impact, which are either assumed or simply ignored. There is no particular expectation that the model corresponds in a literal way to causative mechanisms.

- **Skew**. Skew is a microeconomic phenomenon, not relevant to CGE.

- **Comparison**. CGE models are intended for ex ante comparison of policy alternatives within a single economic system. As with other economic comparisons, care must be taken to avoid assuming that interventions are alternatives simply because their effects can be compared.

- **Timing**. Time is not a factor in equilibrium models, which aim to predict the state of an economic system once shocks have worked through it.

- **Data availability**. At national level, countries tend to be well supplied with the major economic and financial statistical series needed to build and run such models. However, they can be hard to disaggregate to the regional level, owing to lack of data.

Macroeconomic approaches based on production functions

Like CGE models, other macroeconomic approaches work at high levels of aggregation. This is useful for considering impacts at the aggregate level of variables such as productivity and economy-wide effects such as technology and knowledge spillovers. Introducing R&D expenditures into macroeconomic models would allow a number of hypotheses to be tested, such as:

- Comparison of rates of return on public and private R&D expenditures;
- Estimation of social and private return to R&D investments;
- Quantification of spillover effects from foreign investments in R&D;
- Measurement of the impact of tax incentives or credits to innovative activity and productivity levels (Scott, 2013).

The text box illustrates some types of study that can be done using this approach.

- **Causality**. Endogeneity can be a problem because the high level of aggregation can lead to the use of metrics that may be simultaneously determined, e.g. GDP levels and investments in R&D. However, the good supply of statistical data at this level provides opportunities to test a range of instrumental variables. The fact that these series have generally been collected for long periods also means makes it possible to look for variations over time. By focusing on the aggregate relationship between inputs and main outcomes of interest, these approaches bypass the understanding of the exact process by which impacts materialise. These approaches implicitly assume uniform impact mechanisms across all economic actors and shed little light on the innovation process. The specific context within which economic agents behave and create feedbacks cannot be captured in an exact way. However, aggregate spillover effects, can be quantified at this level of aggregation.
- **Skew**. Skew is a microeconomic phenomenon.
- **Comparison**. While most of this style of macroeconomic work suggests that at the level of the economy the benefits of doing R&D are high, it can say little about what **particular** R&D to do. Nor, since key variables in the economic context are bound up in the formulation of the specification of the function lying at the heart of the analysis, is it easy to produce much comparison. Equally, the wide variation among rates of return produced in this style of analysis suggests that the calculations may be subject to wide margins or error or uncertainty.
- **Timing**. The availability of long statistical time series provides opportunities to experiment with a range of time lags in search of explanation as well as to observe differences among short-, medium- and long-term effects. Panel data provide opportunities to work with information collected under quality-controlled circumstances over time.
- **Data availability** is generally good, in so far as this technique relies on national statistical series.

Box 1. Macroeconometric-based studies on the impact of public R&D; some examples

Empirical studies of the macroeconomic effects of private or total R&D spending are more common than those that examine only public R&D. There is a reasonable consensus in the literature that the rates of return for total R&D stocks and that conducted by businesses are positive and statistically significant, even though the estimated rates of return vary considerably (see Hall and Mairesse, 2009; Shanks and Zheng, 2006; Comin, 2004). Evidence for international spillovers from private R&D is provided by Coe and Helpman (1995) and Coe et al. (2009). The literature on business R&D is relevant for public investment because it demonstrates the feasibility of R&D expenditure (which is small relative to the economy) explaining productivity growth, whilst providing a potential comparator.

Analysis conducted under the auspices of the OECD using panel data remains one of the most influential macroeconometric studies on public R&D expenditure. Guellec and van Pottlesberghe de la Potterie (2001) found that government and university R&D (GOVERD and HERD) had a statistically significant positive effect on TFP growth. Using a panel of fifteen OECD economies from 1990-1998, the long-run average elasticity of TFP with respect to all public R&D was estimated to be 0.17, with countries where university R&D had a higher share than government R&D demonstrating a higher elasticity. This compared to an estimated long-run elasticity of TFP with respect to domestic business R&D expenditure of 0.13. In addition, there was strong evidence for international spillovers from public investment – the estimated long-run elasticity of foreign R&D expenditure on domestic TFP was 0.46. The effect of foreign R&D was estimated to be higher in small countries and those with a higher level of domestic R&D.

Similar to this study, Kahn and Luintel (2006) estimate statistically significant long-run elasticities of TFP with respect to public R&D of between 0.11 and 0.30 for a panel of 16 OECD countries. Foreign R&D expenditure was also estimated to have a statistically significant positive effect on domestic TFP, with the estimated elasticity ranging from around 4-8%. Luintel and Khan (2011) split R&D knowledge stock into basic research and applied R&D across a panel dataset of 10 OECD countries from 1970-2006. The estimated elasticities of basic R&D with respect to output and TFP varies (up to a maximum of 0.10 for output and 0.15 for TFP) but is uniformly positive at the 1% significance level, although substantially lower than that for applied R&D. Musolesi (2007) and Eid (2012) also provide comparable estimates, though in the latter case government sector R&D expenditures (GOVERD) have a negative estimated effect on a country's growth (a point elasticity of -0.08), a result which the author attributes to crowding out effects.

Using growth accounting, the Australian Productivity Commission (2007) find, with some uncertainty, that from 1983/4 to 2002/3 around 5% of growth in Australian market-sector TFP can be attributed to R&D stimulated by public support, with around 10% attributable to business R&D and the remaining 85% to other factors.

Source: (Scott, 2013)

Microeconometric-level econometrics

The econometric techniques used to assess the impacts of specific research and innovation policy interventions tend to be microeconomic in character, focusing on limited groups of beneficiaries (together with control groups / non-treatment groups) and changes in their performance. Macroeconomic work looks at impacts at the level of the whole economy and can therefore look for effects not only on beneficiaries (typically private returns) but also on society as a whole, for example through spillovers. The micro-economic focus on those who have been directly affected by an intervention tends to keep the focus of impact analysis on private benefits – which are important, but are not ultimately the reason why the state invests in research and innovation interventions. Relevant variables include turnover, profitability or proxies for innovation activity.

Microeconometric approaches can focus on wages, industrial composition or patenting activity at a high disaggregation level and compare these to performance data. In many cases, they can follow the behaviour of specific economic actors over time (e.g. micro-econometric panel studies) by assessing the impacts of narrow(er) policy interventions on specific targeted populations.

Understanding the counter-factual in order to estimate the net effect of an intervention is often the most challenging part of the analysis. The simplest technique is to use a control group: a “non-treatment group” selected from the same overall population as the beneficiary or treatment group.

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. In Norway, such medical-style control-group methods can no longer be applied owing to the lack of viable control groups (Kvitastein, 2002). Most people 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 but certainly implies endogeneity.

A growing range of techniques is used to overcome the difficulty of finding simple control groups. These to varying degrees involve brave assumptions. The techniques can be used with non-economic variables (e.g. publications) as well as economic ones:

- **Difference in difference analysis** looks for a separate population that is showing similar trends to the treatment group. For example, if the GDP of countries A and B has historically been developing in a similar way and the impact assessor wishes to understand the net effect of an intervention on A’s GDP, then it is possible to use the trend in B’s GDP to approximate A’s GDP in the counter-factual case. The difference in the differences between the trend in A’s and B’s GDP then gives an indication of the net effect;

- **Before-after analysis**, where the behaviour and performance of the beneficiary group **before** treatment is compared with its behaviour and performance after treatment. Extrapolating pre-treatment performance provides a proxy counterfactual; the difference between that counter-factual and the observed performance of the treatment group is then the net effect;

- **Regression discontinuity analysis**, where the fact that funding programmes tend to fund only the “best” applicants or applications is used to create a control group. Typically, competitive subsidy programmes fund only the best x% of the applications they receive. By hypothesis, there is likely to be little difference in “quality” between those applicants falling just above and just below the line, so the technique uses regression analysis to compare the performance of these two groups, with those falling just below the line functioning as a non-treatment control group. The fact that the treatment group in these analyses is selected involves endogeneity. A second issue is that in many cases, those refused treatment under the scheme being investigated can obtain alternative treatments elsewhere. This technique has aroused interest as a possible way to analyse the effects of research funding on researcher performance. However, those refused funding from one source often obtain it from another.

Broadly, microeconomic and econometric techniques offer a lot of promise to give indications about impact. The more complex the context, the more sophisticated the analysis may have to be. Like other techniques, however, they involve significant assumptions and simplifications – a fact that the apparent precision of numbers should not be allowed to obscure.

- Causality. As at higher levels of aggregation, the need in econometric analysis to hypothesise simple relationships between inputs and outputs rather than following a more complex chain of effect logic means that analysis of causes can only be limited. Attribution is therefore difficult because the effects cannot be tied back to the impact mechanisms in simple ways. Micro data tend to contain bigger variations than data series at a more aggregate level, where differences in individual performance tend to be averaged out.

- Skew. Because microeconomic analysis treats data about individual research and innovation-related events, it is vulnerable to the effects of skew where samples are being used or where there is partial information.

- Comparison. Microeconomic analyses suffer from the same issues in relation to comparison as their macroeconomic equivalents. Differences in context can undermine comparisons, as it is difficult to define and capture relevant contextual variables.

- Timing. Micro data are often collected specifically for a particular study and this can mean the analyst has little choice about the time axis – irrespective of whether impacts are likely to build up in the short or the long term.

- Data availability. The growing availability of commercial micro-data about firm performance is making it easier to conduct such studies. Funding agencies increasingly hold databases with information about interventions, which can be supplemented with survey data as necessary. Data protection legislation permitting, it is therefore increasingly practical to investigate the link between intervention and micro-performance.

Cost-Benefit analysis

Cost-benefit analysis (CBA) is conceptually the simplest of all forms of economic impact analysis. It is:

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. (HM Treasury, 2003)

It is essentially an ex ante technique, but has also been used ex post in an effort to understand the absolute and relative value of interventions in research an innovation policy. Normally CBA focuses on direct costs and benefits. As with much other microeconomic analysis, it tends to focus on private benefits, though it can creatively be extended to encompass additional types of benefits. For example, impact analysis of an intervention to support company innovation could be extended to consider not only the benefits received by the treatment group but also to include an analysis of spillovers to other companies.

A key strength of CBA is that it is systematic. In order to make the needed calculations, the analyst has to be explicit about each step of the intervention logic and how to quantify costs and benefits.

The main activities normally carries out in a CBA are:

- 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, normally using discounted cash flow analysis or another technique to allow for the time-value of money. This is important because costs tend to precede benefits by a significant amount of time.

Figure 2 gives an example of the kinds of analysis that could be included in a CBA. However, the extent to which the analysts try to explore outcomes and impacts further "downstream" is limited largely by their imagination and the available budget for impact analysis.

Figure 2. 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 of R&D projects and programmes, but can be extremely important in other evaluation or project-appraisal contexts relating to technology, the environment, health and quality of life.

- **Causality**. CBA tends to use (explicitly or implicitly) a simple logic chart approach to causality, where the analyst attributes effects to the intervention. In the presence of "necessary but not sufficient" logics, attribution may therefore be arbitrary. Contextual factors tend to be ignored.

- **Explanation**. The chain-link logic used in CBA is simple to use and understand, especially ex ante where the complications of "necessary but not sufficient" logic do not have to be addressed.

- **Skew**. Ex post, CBA can easily be affected by skew. For example, when used to analyse the effects of innovation programmes, it has to address the question of how to capture the extreme cases, which may well prove the most crucial to the overall balance between benefit and cost.

- **Comparison**. CBA is poor at comparing impacts because there is no systematic way to ensure that it covers all the relevant steps in the intervention logic or that the time frames considered are relevant to understanding total impacts.

- **Timing**. Like other microeconomic techniques, CBA is vulnerable to analysing an inappropriate period of time from intervention to impact.

- Data availability. CBA normally has to be fed using data specifically generated for the purpose, either directly or via surveys. Improved storage of information about the cost and extent of intervention by innovation agencies is making this a little easier, but special-purpose surveys are often required in order to collect needed impact data.

2. Understanding how impacts happen

Tracing

Some of the earliest research impact studies are based on the idea of "tracing" – literally using expert knowledge supplemented by interviews to trace the inputs that went into particular inventions, discoveries or developments.

As noted earlier, DoD's Project Hindsight (Iserson, 1967) looked at twenty weapons systems and identified "RXD (research or exploratory development) events" that contributed to them. RXD events were ideas resulting from experiments, scientific insights or inventions. Two-thirds of the events occurred long before the system was in development. The median time between an R&D event and its use in development was nine years. Results from undirected basic research typically took about 20 years to implementation while those from directed basic research took about half as long. Some 85% of the RXD events were D0D-funded.

NSF's TRACES project (Loellbach, 1968) studied "critical events" leading to five innovations: magnetic ferrites; the videotape recorder; oral contraceptives; the electron microscope; and matrix isolation. The TRACES team looked further back than the 20-year horizon used by Hindsight. About 70% of the 341 critical events identified were in basic research. Universities accounted for a much higher proportion (one third) of the events than was the case in the Hindsight study. Applied research events were important in the twenty or so years prior to innovation while basic research events stretched up to 50 years back, with a peak in the interval 20-30 years before innovation.

Both Hindsight and TRACES used "backward tracing" from known outcomes. 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 pathways leading forward from the project and contributes to a better understanding of the evolutionary processes of science and technology.

Box 2. 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.

Historical tracing studies have evolved. 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 organisations to investigate on the backward tracing path. As computerised 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 programmes to successful innovations.

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. It can reveal complex chains of events involving many organisations and researchers.

Establishing cause and effect can be 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 organisations, 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.

- **Causality** is easy to establish at the level of ideas, since tracing focuses on the interlinkages among them. It is more difficult to include non-scientific influences in the effect logic. Similarly, the comparative neglect of non-scientific factors means attribution is hard outside the scientific sphere.

- **Explanation**. Tracing provides intellectual explanation but is poor at handling the wider context.

- **Skew**. Since tracing focuses on the points of successful linkage, it effectively seeks out the "skew" and ignores points where linkage was not established or effective.

- **Comparison**. Tracing is a case-focused approach. Tracings can be compared in search of regularities in how linkages work but the individuality of the cases rules out performance comparisons.

- **Timing**. As the Hindsight and TRACES examples showed, timescales can be very long. Forward tracing is more likely than backward tracing to omit key relationships by adopting too short a timescale.

- **Data availability**. Early tracing studies used expert knowledge of the scientific literature. Increased availability of bibliometric databases allows the use of citation and co-authorship to increase the amount of information usable in tracing.

Surveys, interviews

In practice, most R&D evaluation and impact analysis has relied on of beneficiaries, usually coupled with interviews.

While economic analysts sometimes regard surveys as a poor alternative to collecting data by external observation (or through surveys done by others, such as statistical offices), they often provide considerable insight in their own right, as well as being one way to collect data for econometric analysis.

In R&D evaluation, surveys have typically been used to generate 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").

Surveys ask a uniform set of questions about activities, plans, relationships, accomplishments, value, or other topics, which can be statistically analysed. They are open to a range of sources of bias including response bias and biases introduced by the survey instrument or the people running the survey. In contrast, interviews are typically more exploratory. Semi-structured interviews are widely used, in many cases via purposeful rather than random sampling. Again, there are risks of bias through sampling, response and the behaviour of the interviewer.

Surveys and interviews also provide opportunities to explore counter-factual situations where it is not possible to use a control group.

Box 3. 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 Policy Research Unit (SPRU) at the University of Sussex and the Policy Research in Engineering, Science and Technology programme (PREST) at the University of Manchester during the evaluation of the Alvey Programme for Advanced Industrial IT Research in the UK (Guy, Georghiou, Quintas, Cameron, Hobday and Ray, 1991). 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.

- **Causality** can easily be established through the perceptions of programme participants. However, their perceptions are clearly influenced by their own role and the limits to what they can see. For example, the technologists who typically answer such questionnaires may have views on the likely exploitation of new knowledge but tend to have little contact with corporate strategy or marketing functions, which possess market knowledge and have a decisive influence upon commercialisation. Attribution can be done using participants’ perceptions, which are likely to be useful in relation to short-term events but will make less contact with more distant events.

- **Explanation**. Large-scale surveys cannot readily cope with a diversity of paths to impact, so they produce highly simplified, generic impact stories that may need to be supplemented by case studies and interviews, in order to learn about the real mechanics.

- **Skew**. Except at extremely large scale, random surveys risk missing outliers. In many circumstances that is useful, but in the case of innovation the positive outlying cases can generate the bulk of the direct economic benefits, so it can be useful to take steps to ensure their inclusion.

- **Comparison**. Programme performance characteristics (such as speed and friendliness of administration) can be compared using beneficiary studies but the limited grasp that respondents have of medium to long term impacts makes impact comparisons unreliable.

- **Timing**. Most beneficiary surveys are done during or soon after the intervention, so they tend systematically to fail to capture benefits accruing later and are prone to underestimate longer-term impacts.

- **Data availability**. Such surveys of course generate their own data, but benefit from the generally improving quality of innovation funders' in-house databases of beneficiaries and instruments.

Case studies

Case studies are in-depth investigations into the course of an intervention, normally at the project level. The case study approach uses multiple methods of data gathering and analysis, covering both quantitative and qualitative data, including interviews, surveys, content analysis and statistical analysis of secondary data.

Case studies are rarely used to obtain statistical representativeness. Rather they result from purposeful sampling, focusing on success stories, "typical" stories that illustrate aspects of the impact mechanism or fiascos that provide lessons for future action. Their use depends on the purpose of the impact assessment but – except where cases are being used to explain extremely positive impact outliers – they generally serve a formative purpose.

Case studies can also be used in an exploratory manner, to identify impact mechanisms and guide the detailed design of other impact assessment tools to be used in the study.

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. 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. (Kingsley and Klein, 1998)

Case studies are particularly suited for the identification of new products and services, new capabilities and network effects at the micro level as well as for the exploration of impact pathways at the micro level.

- **Causality** at the micro level can readily be explored using case studies. The diversity of methods used within a case study approach increases the chances of understanding the context, compared with interview and survey techniques. Similarly, short-term attribution becomes easier to identify and establish.

- **Explanation** is often the central purpose of doing cases.
- **Skew**. Cases provide an opportunity specifically to explore skewed examples, notably the most successful.
- **Comparison**. While cases can be used to compare impact processes, they are not representative so they are not useful for quantitative comparisons.
- **Timing**. Case explorations allow the analyst to explore timescales. Given the long run nature of many research and innovation impacts, however, there is a constant risk of under-estimating the time needed for impacts to rise to their highest levels.
- **Data availability** varies but can be a useful criterion in choosing what cases to do.

3. Focusing on particular outcomes

Human capital

Human capital approaches have been little used in impact analysis relating to research. People tend to be regarded as vectors rather than repositories of knowledge, in line with the focus of the linear innovation models on the effects of new knowledge. This contrasts with the current perspective on the role of knowledge in innovation (Figure 3) and the idea of absorptive capacity (Cohen and Levinthal, 1990), which is inherently a human-embodied characteristic of the firm.

Recent Swedish long-term impact studies (Arnold, Good and Segerpalm, 2008; Stern, Arnold, Carlberg, Fridholm, Rosemberg and Terrell, 2013) analyse the build-up and distribution of "knowledge bearers" associated with particular interventions, focusing particularly on PhD students and their subsequent career development. There is clearly scope for much more extensive exploration of human development capital in research impact – and in considering human capital **as** an impact in itself.

- **Causality**. Better understanding of the role of human capital in impact generation should lead to more representative impact logics.
- **Explanation** will therefore be improved through describing and delineating not only specific but also generic roles for human capital.
- **Skew**. It is likely that there is skew here too: some people are much more effective at generating impacts than others. However, this dimension has barely been considered in impact analysis to date.
- **Comparison**. It is not clear that human capital considerations would increase the comparability of impact studies.

- **Timing**. Improved knowledge of the role of human capital in obtaining impacts will improve our understanding of the timescales that need to be addressed in impact analysis.

- **Data availability**. Generally poor in the sense that registers of people exist in many countries but there are significant barriers to accessing them in impact analysis. Improved bibliometric database and Research Information Systems may mitigate this to some degree.

Scientometrics

The field of scientometrics – the quantitative study and analysis of the outputs of research (normally publications and patents) – is comparatively young but beginning to mature.

While the production of bibliometric data remains dominated by two companies – Thomson-Reuters and Elsevier – prices have been falling and the availability of analytical software has improved, making it increasingly feasible to integrate bibliometric analysis into evaluation and impact analysis. In the past, bibliometric analysis has been carried out at the levels of scientific fields and countries; now, it is increasingly possible to use micro-data to explore behaviour and performance at the level of research groups, individuals and R&D programmes.

The simplest application of bibliometrics is to count research outputs in the form of publications. A more refined usage is to use the extent to which publications are cited as a proxy for their influence or quality – despite the fact that there is no well-articulated theory of citation at this stage. The limitations of bibliometric databases are well known and include:

- Different fields have different publication traditions, so numbers of publications and citations are not necessarily comparable across disciplines;

- Broadly, bibliometrics work much better in the hard sciences and economics than in the social sciences and humanities;

- There is a strong English-language bias in the bibliometric databases;

- The databases' historical coverage is poor; at the same time, journal coverage is continuously expanding, so comparisons over time have to be made with care;

- They are self-referential, so for example citations to books (which are not listed in the databases) are ignored;

- Open access introduces distortions, with open-access articles tending to be more frequently cited than non-open access ones.

Increasingly, analysts can work around or take account of these limitations.

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.

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.

Patent data are freely available on the Internet. Both the U.S. Patent and Trademark Office and the European Patent Office offer online and offline search capability for patent citations. Patent analysis is hampered by the uncertain value of an individual patent. Scientific articles and patents are not cited for the same reasons; so far, analysis of patent citations is therefore problematic.

- **Causality**. Links from bibliometric databases to funding sources barely exist, so bibliometric analysis is largely internal to the world of publication. New generations of Research Information systems are being built that make the connection with funding, but there is still a lot of work to do to relate their rather patchy coverage of funding to bibliometrics.

- **Explanation**. Co-publication and citation links provide important information about how knowledge spreads.

- **Skew**. Skew is important in publications as in innovation, with a minority of authors being most productive and producing the bulk of the highly cited papers.

- **Comparison**. Easy within the closed world of publications. Additional data are needed for performance comparisons, for example.

- **Timing**. Different fields and different journals have varying publication ages, complicating analysis.

- **Data availability**. For those able to pay the fees, bibliometric data availability is good. Google Scholar may eventually provide something of a free alterative but differences in data collection and coverage as well as the lack of tools for using Scholar data mean this is not going to happen immediately.

Altmetrics and Webometrics

"Altmetrics" – i.e. devising alternative or supplementary sets of metrics to the traditional bibliometric indicators to "measure" research dissemination, visibility and impacts – is generating increasing attention. It has its own manifesto on the Web (www.altmetrics.org) and a significant number of research articles in the subject are hosted in the Altmetric Collection of the PLOS open-access journal at www.plos.org. Objects of study in altmetrics are continuously expanding, but include:

- Social media like Twitter and Facebook;
- Online reference managers like CiteULike, Zotero, and Mendeley;
- Collaborative encyclopedias like Wikipedia;
- Blogs, both scholarly and general audience;
- Scholarly social networks, like Research-Gate or Academia.edu.

The use of indicators of social media is relatively well established in advertising and marketing. Some organisations are starting to collect altmetrics on a commercial basis in relation to research publications. In this context, little is known about what individual altmetrics **mean**. This appears to provide a rich set of opportunities for research – though we are probably some time away from being able to think in a precise way about concepts such as a "field-normalised tweet".

For the time being, it is clear that altmetrics do not provide alternatives to established bibliometric measures for impact analysis. They can tell us about the sociology of knowledge production and dissemination – and they may be especially interesting in the field of understanding in more nuanced ways the nature and scope of impacts within research communities and with the "users" of research. In the meantime, the significant problems of "noise" in the data and the ease with which certain social media communication indicators can be "gamed" coupled to the immaturity of the field (Priem, Piwowar and Hemminger, 2012) suggest that altmetrics will complement rather than replace existing indicators.

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. 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. However, lack of uniformity in the way scholarly information is published and its presence is maintained on the Web undermines its use as a serious impact assessment technique.

- **Causality**. Unclear.
- **Explanation**. Unclear.
- **Skew**. Unclear.
- **Comparison**. Unclear.
- **Timing**. Unclear.
- **Data availability**. Highly available.

Social network analysis

Networks are important in the study of research and innovation. Social network analysis is a method of mapping and measuring relationships and linkages among researchers, groups of researchers, laboratories, or other organisations. This can identify routes through 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. Network shape, size, and density can serve as indicators of the strength of communities of practice and signal relative roles and relationships.

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 and density of the network or extent to which a network member is directly connected with others. Another important measure is centrality, based on the number of direct links one node has to other nodes.

Social network analysis can be utilised:

- 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;
- 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.

So far, network analysis has been used at a small scale in evaluations of phenomena such as networking in health research and, experimentally, at a much larger scale in understanding some of the co-operative networks associated with the Framework programmes.

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 earlier) 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.

- **Causality**. A key weakness of social network analysis in research and innovation impact analysis is a lack of well-articulated theory about what different kinds of network relationships **mean**. This is because almost no work has been done to connect network patterns to micro-behaviour.

- **Explanation**. This lack of theory means that the explanatory power of network analysis is often rather weak.

- **Skew**. In principle, network analysis can identify skew – and it would be useful if research effort were devoted to understanding the meaning and effects of unusual behaviour, in terms of networking.

- **Comparison**. Network properties are mathematically described and can be compared.

- **Timing**. So far, most network analysis is static. A new generation of dynamic visualisations offers promise for the future understanding of dynamic but so far there is little work that goes beyond visualisation to developing theories of dynamic network behaviour in research and innovation

- **Data availability**. There is a range of sources, notably bibliometric databases. Data for the study of specific interventions tend to have to be collected at first hand. A problem is that these data normally need to be **complete**: missing data undermine their analytic usefulness.

4. Impact assessment for performance measurement: assessing the social impacts of university research

Since the mid-1980s, when the UK introduced the "Research Assessment Exercise" to allocate institutional research funding to the universities, a growing number of countries have been introducing performance-based research funding systems (PRFSs) to allocate institutional funding, based on quality control and competition. These systems reallocate some of the institutional funding from organisations seen as performing badly to those performing well.

In principle, a PRFS has two components:

- A performance assessment system, which evaluates research performance;

- A funding system, which takes the results of assessment and uses them to determine the allocation of institutional funding.

Sörlin (Sörlin, 2007) identified two "waves" of PFRSs. The first wave PFRSs were peer review based, with the UK Research Assessment Exercise (RAE) as the archetype. They tended to focus exclusively on scholarly outputs from research. From around 2000, the second wave was more strongly indicator-based. The third wave is represented by the extension of previous systems to tackle non-academic outputs, most commonly patents. In effect, this represents a rather belated reflection of the change in the "social contract" between research and society, into which has increasingly been introduced the expectation that science is done not only for its own "sake" but is increasingly expected to deliver positive social impacts through economic development, improvements in the environment and quality of life, and so on (Elzinga, 1997, Hessels, van Lente and Smits, 2009).

In practice, PRFSs have not attempted a heavily metrics-based treatment of impacts. The Research Excellence Framework (REF) that succeeded the RAE has introduced a requirement for "impact statements" or case studies of societal impacts of past research, despite great uncertainty about how such impacts can be compared or measured. The REF leads the field in terms of trying to integrate social impacts into research assessment – but the outcome remains uncertain.

The REF allocates 65% of funding on the basis of outputs submitted to the expert panels, 20% on the basis of impacts and 15% based on the "research environment".

REF submissions were required to include:

- A completed impact template (REF3a): describing the submitted unit's approach, during the assessment period (1 January 2008 to 31 July 2013), to enabling impact from its research;

- Impact case studies (REF3b): describing specific impacts that have occurred during the assessment period (1 January 2008 to 31 July 2013) that were underpinned by excellent research undertaken in the submitted unit. The underpinning research must have been produced by the submitting HEI during the period 1 January 1993 to 31 December 2013. (Guidance on Submissions, p23).

The excellence of the research upon which the impacts were based had to be documented by the provision of a scientific article of at least "two star" quality.

The REF is so far the only PRFS that has adopted qualitative impact statements. Its assessment of the statements will not be published before 2015, so at the moment there is no well-documented experience of using such a technique.

Error! Reference source not found. shows the rating scale to be used by the REF panels. It relies on the field-based nature of the REF panel system to enable the panels to make relative judgements among the research groups they assess but leaves the peers with no concrete criteria for making this judgement. Panels are responsible for using the norms of their field in making all judgements, so impact judgements are likely to represent conventional wisdom in the respective field. The result is likely therefore to be consistent within fields but **not** comparable **across** fields.

Figure 3. REF Rating Scale for Impact

	Impact sub-profile: criteria and definitions of starred levels
Four Star	Outstanding impacts in terms of their reach and significance.
Three Star	Very considerable impacts in terms of their reach and significance.
Two Star	Considerable impacts in terms of their reach and significance.
One Star	Recognised but modest impacts in terms of their reach and significance.

Unclassified	The impact is of little or no reach and significance; or the impact was not eligible; or the impact was not underpinned by excellent research produced by the submitted unit.

UK interest in pursuing impact in research assessment antedates the REF by a number of years. The UK research councils increasingly have been demanding of applicants that they explain the anticipated social impacts of their research when applying for funding.

- **Causality**. Explanations of causality are case-based. In this type of system there is little opportunity to consider attribution.
- **Explanation**. Ditto.
- **Skew**. Those submitting cases for analysis will tend to identify and submit the positive outliers.
- **Comparison**. Criteria are to date seriously under-developed.
- **Timing**. In the REF, restrictions on timing appear to undermine ability to report longer-term impacts.
- **Data availability**. Depends on those being assessed to obtain data.

This document is based on: **OECD Directorate for Science, Technology and Innovation (2014), "Assessing the Impact of State Interventions in Research – Techniques, Issues and Solutions", unpublished manuscript.**

References

Arnold, E. (2013). A Trace of Hindsight: Evaluation and the Long Term Impacts of R&D (Presentation). OECD Evaluation Conference. Paris: OECD.

Arnold, E., B. Good and H. Segerpalm, (2008). Effects of Research on Swedish Mobile Telephone Developments: The GSM Story, VA2008:04. Stockholm: VINNOVA.

Cohen, W. M. and D.A. Levinthal, (1990). Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly , 35 (1), 128-152.

Elzinga, A. (1997). The science-society contract in historical transformation: with special reference to 'epistemic drift. Social Science Information , 36 (3), pp. 411-455.

Guy, K., L. Georghiou, P. Quintas, H. Cameron, M. Hobdy and T. Ray. (1991). Evaluation of the Alvey Programme for Advanced Information Technology. London: HMSO.

Hessels, L. K., H van Lente and R. Smits, (2009). In search of relevance: the changing contract between science and society. Science and Public Policy , 36 (5), pp. 387-401.

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

Iserson, R. S. (1967). Project Hindsight (Final Report). Office of the Director of Defense Research and Engineering. Washington DC: Department of Defense.

Kingsley, G. and H. Klein, (1998). Interfirm collaboration as a modernisation strategy: a survey of case studies. Journal of Technology Transfer , 23 (1), 65-74.

Kvitastein, O. (2002). Offentlige evalueringer som styringsinstrumenter: Kravspesifikasjoner og kontrollprobleme. Bergen: SNF Rapport N 20/2002.

Loellbach, H. (1968). Technology in retrospect and critical events in sciences (TRACES). Vol. I. Illinois Institute of Technology Research, Contract NSF-C535. Chicago: Illinois Institute of Technology.

Priem, J., H.A. Piwowar and B.M. Hemminger, (2012). Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact. ACM Web Science Conference 2012 Workshop. Evanston, IL: ACM.

Sörlin, S. (2007). Funding Diversity: Performance-based Funding Regimes as Drivers of Differentiation in Higher Education Systems. Higher Education Policy , 20, 413-440.

Scott, R. (2013). Linking science to Innovation, Economic Growth and Welfare: Scoping Paper. Paris: RHIR, OECD.

Stern, P., E. Arnold, M. Carlberg, T. Fridholm, C. Rosemberg and M. Terrell, (2013). Long Term Industrial Impacts of the Swedish Competence Centres. Vinnova Analysis, VA2013:10. Stockholm: Vinnova.

Young, J. (2008). Strategies for Influence and policy Relevance,. London: Overseas Development Institute.