The Challenges of Impact Assessment

Working Group 2: Impact Assessment

ESF Member Organisation Forum on Evaluation of Publicly Funded Research

Table of contents

1. Introduction

Background

What is the purpose of and how can one assess impact from research? This is an important question as societies are dependent, and increasingly so, on the creation of knowledge to yield innovations and better policies to generate economic and social benefits. The assessment of impact from research is the attempt to establish to what degree research affects certain changes in science and society. What evaluation means for research funding organisations is discussed in a report from an earlier European Science Foundation (ESF) Member Organisation Forum (European Science Foundation 2009). This study identified five different types of evaluations, and considered different levels of evaluations: a) of the funding agency as an organisation; b) of funding policies; c) of research fields or scientific disciplines; d) of funding schemes; and e) of research grants.

There is not always a clear-cut distinction between impact assessments and other evaluation studies. In general evaluations of research the prime objective is to support and strengthen the quality of research, but impact assessments have a broader objective. Impact assessments can share some characteristics with a goal oriented type of evaluation. While the latter aims to evaluate how the goals and objectives of the intervention (for example, a research project) are being met, the impacts do not necessarily have to be formulated as goals for the intervention. One of the aims with impact assessment is to gain a more thorough knowledge of the relationship between actions and effects, but this can also be an aim for formative or process oriented evaluations. It is therefore important to make a distinction based on the purpose of the evaluation.

Ideally, the use of impact assessments could perform a dual task: demonstrating the value of research, and increasing the value of research through a more effective way of financing research in order for research to have impact. Firstly, publicly funded research has a responsibility to contribute something in return to society, and impact studies are one way of showing these returns and, in doing so, legitimise investments in research and serve as instruments to advocate for funding. Secondly, impact studies can strengthen returns to science and society by improving the instruments that are used to fund research. In this instance they may also provide a better understanding of transfer of scientific knowledge into practice. The improvements in instruments may be structural, i.e., improving the way funding schemes are constructed and how research environments interact with society, or improvements may also arise as the process of evaluation affects the individual behaviour of researchers and stakeholders, as they become more aware of how research affects society and vice versa. They can also be used to further discuss the question of the relationship between scientific excellence and research being beneficial to society.

The interest in understanding the links between research, innovation and societal development is not new and neither is the search for appropriate methods to assess the impact of research. In the last few years, however, there has been growing pressure from the political sphere to measure and achieve value for money from investments in research. The complex and interrelated reasons behind this recent development is not the focus of this report but a discussion of impact studies cannot disregard the relationship between the STI (science, technology and innovation) policy framework and an evaluative framework (Donovan 2007a and 2007b). It is argued that "research policy has been broadly reframed emphasising notions of 'value for money', democratic oversight and accountability" and that "[p]ublic research funding is … increasingly understood as a strategic

investment where state economic and regulatory strategies are oriented towards maximising returns" (Kearnes and Wienroth 2011, p. 157). In this context, public research funders are assigned a new role where they are players that transcend funding research, and research and the benefits of research are more closely linked to other sectors of society.

Most notable is the increasing pressure to demonstrate impact in basic research. The debate on the economic value of public research funding has therefore invoked "… a set of contested cultural values concerning the role of science in contemporary social life" and can be seen as "… a historic contestation over the standards by which the principles of creativity, autonomy, and diversity are to be judged" (Kearnes and Wienroth 2011, p. 155). The research sector has reacted by emphasising the long-term economic and social values associated with basic research and arguing for the importance of assessment of the quality of knowledge produced. That this policy shift has caused not only debate but also some anxiety is understandable as much might be at stake if assessments of impact become a predominant deciding factor in the distribution of public research funding (see, for example, the REF Pilot a 2009; b 2010; c 2010).

This raises the question if it is at all feasible to assess impact of all kinds of (basic) research. A NIST report from 2003 points to the general methodological problems associated with impact assessment (Tassy 2003, p. 1):

> In fact, a single 'manual, for impact assessment may never be achieved. The technology trajectories and economic outcomes that government programs and projects seek to leverage vary significantly, as do the complex economic structures that characterize a technology-based economy. Thus, no metric or measurement method can (1) address the diversity and complexity of an R&D agency's technological outputs, (2) describe the subsequent processes by which private sector impacts occur, and finally (3) accurately capture the resulting economic outcomes.

Resulting from an awareness of the shortcomings of the instruments for assessment of impact, a significant amount of energy has been directed towards the improvement of established methods and towards the development of new methods of assessing impacts. Special attention has also been paid to developing methods that aim to capture values other than economic outcomes. This development and the problems as well as advantages associated with different methodologies will be discussed in more detail in this report.

The aim of the report

For the member organisations of ESF, how to conduct ex-post impact evaluations is a burning issue. Demonstrating and achieving returns to society from research are high on the agenda. The prime instrument for this task – impact assessment – is difficult.

This report aims to highlight the development of different methodologies for assessing impacts. We will discuss the advantages and shortcomings of different methodologies. The focus is on impact assessment in relation to public research funding, and is especially concerned with the evaluation of funding schemes and research programmes. The main concern is to identify good practices of impact evaluations and to make recommendations on how to perform impact evaluations.

The structure of the report

We will begin by discussing the concept of impact; what does impact mean? The main challenges with impact evaluations will then be highlighted and further discussed. We will then proceed to discuss the various approaches to assessing impact that exist and we will highlight areas of special interest to the member organisations of the ESF. The report concludes with recommendations for the future.

2. What do we mean by impact?

In the literature there are different conceptions about what impact is, but it is surprisingly often left without a definition. Impact can be described as consequences of an action that affects people's lives in areas that matter to them. The consequences would not have occurred without the original action. It can be added that not all consequences may warrant an assessment of impact. They must be of some importance to us, or seen as useful.

Impact does not necessarily have to beneficial. There can also be different perspectives as to whether a particular impact is positive or negative. Impact does not have to be intended. Nor should it be seen as final consequences of an action, as we can never know when the consequences of an action are final.

At an overarching level it is possible to make a distinction between *scientific consequences* and *societal consequences* of research. Scientific consequences are, for example, the advancement of knowledge and how the research landscape is influenced. For example, what consequences does the funded research have on the organisation of research (size, cooperation, human capital)? If there is investment in a certain field of research, how does this affect this field scientifically and how are related fields affected? And with regards to research questions within this field: To what extent are the research questions directed towards the needs of society? To what extent are the research questions risky?

Societal consequences include addressing questions, such as what does society gain in the form of better products, better services, healthier lives, better welfare, a sustainable development, etc? Hence, impact on society can include every important aspect of society. Early studies on assessing impact concentrated on what science contributed to technological innovations, improvements in health and the creation of national wealth. Yet these benefits are still most commonly associated with what science is supposed to deliver to society. There has been a considerable broadening of impact assessments to include, for example, cultural and social returns to society, although studies of academic and non-academic impact of arts and humanities, conceptually and methodologically, are still underdeveloped compared to other subjects (Levitt *et al.* 2010, p. 35). However, in the above examples of impact there is a difference in how tangible they are. Whereas some are very noticeable – economic wealth – some are less easy to measure, for example, changing attitudes and ways of life. Nevertheless, not all important consequences of an action need to be labelled impact, which will be discussed below. But, the increasing interest in the less tangible impacts have changed the methods used to assess impacts; a move from a heavy reliance on quantitative measures, often on a macro level, to a wider set of methods of used.

Outputs, outcomes, impacts

Sometimes, there is a distinction drawn between outputs, outcomes and impacts of research, all of which are consequences of an action. Publications, discoveries, patents are normally seen as outputs. These outputs can become outcomes: for example, policy guidelines, building competence and product development. In some research impact assessments there are other terms besides outputs and outcomes, but the logic remains consistent. Impact is then described as increased employment, improved health and/or wellbeing and increased productivity and reduced waste.

Sometimes a conceptual framework of outputs/outcomes/impacts is portrayed as a linear process. According to the linear model the research process is a one-way street. Knowledge from basic research transforms into applied research questions and finally into social and technical innovations, and then impact. A linear model of how research translates to impacts has been criticised for not corresponding well with reality. Impact sometimes also relates to final outcome, which is problematic as it is impossible to know when the outcome is final. Instead, in the conceptual framework above, it would be better to describe impact as a broad effect on science or society.

But the conceptual framework does not have to be portrayed in a linear fashion. The output/outcome/impact perception may be useful to give a structure to the analysis without presupposing a linear development from research to practice. However, the dividing line between outputs, outcomes and impact is not always that clear-cut.

Instrumental, conceptual and broad consequences

Another way to conceptualise impact is to see impact as something all-encompassing rather than as in the above distinction between outputs/outcomes/impacts. Consequences could instead be described as instrumental, conceptual or broad. The idea behind this theoretical model is that the interaction between research and society affects how different actors in society behave, either by direct influence or by influencing their opportunities to act. This could be applied in numerous ways, for example, through policy – a plan of action or a measure developed in response to a perceived need, in order to achieve a particular outcome; through public debate – in what aspects does science affect the public debate – topics, how the public debate is structured; or through medical treatments and new technical appliances.

Consequences of an action on a societal level could then be seen as the following examples (see Davies, Nutley and Walter 2005):

- Instrumental consequences: political decisions, official guidelines, new technology, spin-off companies
- Conceptual consequences: knowledge, understanding and attitudes in certain issues
- Broad consequences: economic welfare, improvements in health, sustainable environment.

And consequences on science can be seen as:

- Instrumental consequences: publications, organisation of science, collaborations, training of personnel
- Conceptual consequences: ways of thinking, new research questions

- Broad consequences: paradigm shifting knowledge.

Which of these consequences should be seen as impact? The question is not straightforward to answer, and it is difficult to give an authoritative answer. If transformed to the logic behind the outputs/outcomes/impact model at least the broad consequences are to be regarded as impact. But the other consequences or parts of them may also be seen as impacts.

The nature of impact

As already mentioned impact can roughly be divided into scientific and societal. Societal impact in turn can be divided into several categories (see, for example, European Commission 2010; Delanghe and Teirlinck 2010).

- *Scientific impact*: contribution to the subsequent progress of knowledge, the formation of disciplines, training and capacity building.
- *Technological impact*: contribution to the creation of product, process and service innovations.
- *Economic impact*: contribution to the sale price of products, a firm's costs and revenues (micro level), and economic returns either through economic growth or productivity growth (macro level).
- *Social impact*: contribution to community welfare, quality of life, behaviour, practices and activities of people and groups.
- *Political impact*: contribution to how policy makers act and how policies are constructed and to political stability.
- *Environmental impact*: contribution to the management of the environment, for example, natural resources, environmental pollution, climate and meteorology.
- *Health impact*: contribution to public health, life expectancy, prevention of illnesses and quality of life.
- *Cultural impact*: contribution to understanding of ideas and reality, values and beliefs.
- *Training impacts*: contribution to curricula, pedagogical tools, qualifications.

All of these forms of impact may be interesting to investigate, but what forms of impact that are assessed do also affect what methods that are the most appropriate. Different methods and approaches to impact assessment will be discussed further in chapters 4 and 5.

Conclusion

The point of making all these distinctions is not to recommend that every one of these distinctions needs to be applied. But, when designing an impact assessment, it is advisable to conceptualise what impact is to be assessed in order to make a study that is fit for purpose.

To summarise, impacts are consequences of importance of an action, and the focus of impact assessments can be both impact of research and impact on society broadly defined. Before a study is undertaken it is useful to clarify the concept of the impact, and if and how it differs from other consequences of an action.

3. Impact studies – methods and methodological challenges

To assess impact of research poses serious methodological challenges and there are at least four crucial problems that have to be taken into consideration:

1) How to attribute the intervention to the observed effects.
2) How to determine counterfactual positions, i.e., would the observed effects have occurred anyway?
3) How to deal with the time lags between research and tangible outcomes, and the multiple stages inbetween.
4) Where to focus the assessment, ranging from a research project to research in general.

The answers to these questions have a profound effect on the design and methodology of an impact study.

Attribution

The general idea behind attribution is to link impact to a certain instrument or action (e.g., a research programme). If it can be established that an observed phenomena (e.g., economic growth) is more or less dependent on the instrument or action, we can state that the observation is attributed to the action taken. However, the means of convincingly establishing an attribution may be insufficient, and some paths to impact are more obscure than others. To isolate the role played by a single actor or action, such as a research funder, is consequently cumbersome or more likely virtually impossible.

Thus, from a methodological point of view, it is very difficult to establish attribution. It is also more difficult if the object of analysis is on a macro level, i.e., overall economic impact, general improved health, etc., as there will be a multitude of factors affecting the outcome. Some impacts appear many years, even decades, after the research is carried out. This makes it even more difficult to deal with attribution.

The recommendation is to focus on contribution or the value of the research (Levitt *et al.* 2010, p. xiii), rather than attribution, of a funder, programme or project, i.e., to say that a funder has played a role in the impacts of a research project it has funded, rather than determining the exact share of the impact that can be claimed.

There are two reasons to abandon attribution for contribution. Firstly, we know that the outcomes of research are very rarely the consequence of one singular research activity, and whether research transforms into impact is also dependent on factors and actors outside the research sector. Consequently, it is a false way to describe reality. Secondly, from a methodological point of view it is very difficult to attribute outcomes to inputs and it demands a number of assumptions (of which some surely will be over-simplified) (Cox 2010, pp. 10-15).

Research funders who wish to establish what difference their specific funding makes might still be interested in attribution, and rightly so. Attribution, rather than contribution, is also probably a better basis for making future strategic choices. But the urge to establish attribution should not stifle the possibilities of doing impact studies. So to raise attribution as a real obstacle for doing impact

assessments is in some way misdirected. To strive for attribution is understandable, but the possibilities of the methods that can be used are limited. To conclude, there is not a clear-cut distinction between attribution and contribution but more of a sliding scale. As far as possible one should try to establish the magnitude of the influence of an action. But, this should not be done at the expense of the rigour of the methodology.

The counterfactual argument

Related to the question of attribution is the challenge of defining an appropriate counterfactual position. To determine the impact of an intervention, one must also estimate what would have happened if the intervention had not taken place, as it may be possible that the outcomes we are tracing might have occurred anyway. Even more difficult is the question of opportunity costs. Would the same resources have provided more impact if they had been invested in another research project or programme (Go8 2011, pp. 6-7)?

To establish a counterfactual position is relatively easily done in a laboratory setting where essential factors influencing the outcome can be held constant. Another way to establish a counterfactual position is to use a randomised controlled trial (RCT), a method often used in medical science. The key feature of a typical RCT is that the study subjects are randomly allocated to different groups, which are distinguished by how they are affected by the intervention or not. Hence, the results will show one (or more) group that is affected by the intervention and another group that is not affected. The differences that can be observed between the two groups can be described as the impact of the intervention.

Generally, methodological discussions on how to establish a counterfactual position are somewhat underdeveloped in impact assessments of research. Ramberg and Knall (2012) do extensively discuss the challenge of establishing a counterfactual position, but from a theoretical perspective. There are very few examples of elaborating with counterfactual positions in impact assessments, and when assessing impact of research it is practically impossible to establish a counterfactual position based on experiments or quasi-experiments.

What other measures are then appropriate to establish a counterfactual position? There are various techniques available, which it is advisable to combine, to establish a satisfactory counterfactual position.

1) Interviews
   One way is to interview researchers who have been financed or in other ways affected by the action or stakeholders who are likely to be affected by the action. They are in a position to judge what difference the action has made. In this way, it is possible to compare those have been exposed to an intervention with the situation before the exposure. However, this is also a source that has its weaknesses. Those who are interviewed may have difficulties remembering and estimating the impacts in an impartial manner.

2) Establish a model of the intervention logic
   Intervention logic models are often used in impact assessments. In such a model the search for impacts follows a description of how impact is created in an ideal situation. It involves a description of how impacts are thought to materialise. If impacts can be observed that fit

into the intervention logic there is a good indication that the action has caused the impacts. A useful model must strike a balance between usefulness and realism. It cannot portray reality fully as that would obscure its purpose. At the same time it needs an adequate realism, otherwise it would be unlikely to produce interesting results. An intervention logic model needs to include expected outputs (or consequences) from research, and an idea of how diffusion of knowledge to society takes place and how actors can make use of the research, and finally the connection between these levels (Molas-Gallart, Tang and Morrow 2000, p. 174)

A much used intervention logical model is the payback framework (see Hanney *et al.* 2004), but there are also other examples (see Academy of Finland 2009, pp. 21–24; Molas-Gallart *et al.* 2000, p. 173). Using intervention logic gives the possibility to investigate if expected impacts do not materialise. And it makes it possible to distinguish if the lack of impact is caused by a flawed logic in how it is thought to be materialised or if there is a problem in the phase of implementation. There is really no contradiction between an intervention logic model and capturing impacts outside the expected logic of impact. In the design of the impact assessment it is advisable, depending on the aim of the study, to include elements where unexpected impacts can be captured.

3) Judgments by experts
   Identify different actions (inputs), including others that are of primary interest to the impact assessment, and consequences (that may translate into impacts). Can impact be established or are the consequences caused by something else? In this process it is possible to have experts to make judgments on what the situation would have been if the intervention had not occurred.

4) Establish control groups
   If possible, try to use some form of control group. One way of constructing a control group could be to compare those who have been exposed to an intervention to the typical outcome in a group that has not been exposed to an intervention. This is almost a quasi-experimental design, but not as strict and controlled. It could, for example, be a comparison between researchers who were funded in a certain programme and those who were rejected. There are very few examples of some form of control group being used in impact studies, and it would be interesting to see a wider use of this technique.

Time lags

All studies acknowledge that the contributions of research often occur and manifest themselves over long timescales. The short-term impact of research can, furthermore, differ significantly from the long-term impact from the same research (Go8 2011, p. 4).

When is then a good time to measure impact? For how long should we be trying to identify impacts? There is, of course, no authoritative answer to these questions. On the one hand, there is a need for a sufficiently historic time window to allow impacts to occur. It might take a very long time before broader impacts from research appear, not least from basic research. In many cases it is not possible

or even desirable to wait that long to conduct an assessment of impact. As impact evaluations are often intended to serve as a basis for strategic planning, the information may not be useful after a long period and is required immediately. On the other hand, collecting evidence is necessary, and this must be done without losing too much detailed information and data. What is the quality of records, the ability of researchers to recall their activities and stakeholders to remember how practice was influenced in a process that has proceeded for a number of years?

Ideally, an impact study with such a profile is best suited to be conducted alongside this process. But impact studies running for 10-20 years are not very likely, and there would be very few, if any, executed. Furthermore, in many cases, the question most likely to attract interest is what recent investments in research have led to.

How can this situation be dealt with? First of all, there is a need for studies that can be described as longitudinal, and which create a foundation for the future understanding of impact. These studies follow a long time frame and try to capture impacts in the long run as well as in the short run. This kind of study also describes the process from research to impact.

These longitudinal studies have, apart from their own aims, a certain purpose for impact evaluations that are not longitudinal. If you lay out the full process from research to impact you will hopefully notice that there are short-term indicators that are predictive of long-term impact. These early success indicators presuppose that you have detailed information about what causes impact.

The recommendation is to do more research and studies into the processes of how impact occurs. This would hopefully lead to good and solid models and theories of how research gives impact. And, even more hopefully, this could pave the way for finding short-term indicators that are predictive of long-term impact. This would in addition lessen the problem of the time lag, and thereby impact studies may play an even larger part in influencing the strategy of research funding. There is definitely some interesting work being done in this area, for example the SIAMPI model that will be elaborated in the next chapter (see also Canadian Academy of Health Sciences 2009; Luoma *et al.* 2011 for an analysis of indicators).

Another similar way to deal with the time lag is to see the impact in different phases, which will leave room for different approaches to analysing impacts:

a) Potential impact (short-term), for example research on knee surgery that has the potential to be more effective so that people can be rehabilitated faster and use less health care resources.
b) Action towards impact (medium-term), new guidelines on how to perform knee surgery.
c) Impact observed (long-term), operations are carried out according to the new guidelines which lead to benefits for the individual and society.

Micro/macro level

Another challenge in impact assessments is the object of analysis and what conclusions and recommendations that is possible form an impact assessment, and that depends a lot on what level the impact assessment is carried out.

A study from a macro perspective tries to answer the impact of research in general and the impacts of research are on a high level of aggregation, for example, economic wealth. A micro perspective on the other hand takes a specific research project as the point of departure. Of course, there are several approaches between the two extreme positions described above.

The obvious drawback with a macro approach is the difficulty of determining if the consequences observed really are impacts of the action assessed. A correlation is easier to establish with a micro approach, but the drawback of this approach lies in the difficulty of aggregating the results from a micro approach to a general level.

4. Impact study methods – as discussed in the literature

Impact assessment methods and methodological challenges will be further discussed in relation to the reviews of the selected accomplished impact studies. As an introduction to these reviews we will in this part present some methodological approaches that have been discussed in studies of impact assessment. It is possible to categorise the methods or models in several ways, for example, on the basis of the type of impact, the level of aggregation (national research funding or specific research programmes), the expected audience for the results, whether the assessment is quantitative or qualitative, etc. Some methods can furthermore be used for different types of impact studies and for different purposes, and many of the more complex models presuppose that various methods are used in combination. This is also the case with most impact assessments. The aim here is to give a brief overview and the different methodological approaches are presented according to a simple grouping, which includes a mix of methods and what could be called impact assessment models.[1]

Input measures

Input measures do not identify impact but they can give vital information about the kinds of impact one might expect. On a national level the information about the prioritising between basic and applied research, between different research fields, etc., as well as information about the balance between different sources of research funding can serve as indicators. To understand the boundaries for possible impacts in this way can be an important and relevant starting point also for individual research funders.

Output measures

[1] For this part a combination of several articles and reports has been used and they are all included in the list of references. Direct references will be given only when we are referring to specific information.

Output measures can be used to assess research productivity, and they can be seen as a stage in the road from research to impact. The most obvious output measure is publications, but processes or tools used to disseminate research can also be considered as a type of output indicator. Both types will be discussed in the following section.

Bibliometrics

Today publications and especially citations tend to be looked upon as indicators of impact rather than merely output measures. Bibliometrics has experienced a rapid and immense development and the methods have become more sophisticated and more commonly used. The use of bibliometrics has, however, also been subjected to severe critique. The critique has highlighted several problematic issues. One shortcoming with these quantitative indicators is that they do not grasp the qualitative aspect of research excellence and although citations say something about the impact one must remember that the citations can be both positive and negative. Another problem is associated with the differences in publishing cultures that disadvantage some disciplines compared to others. In order to overcome this, field sensitive citation indexes have been constructed. Not all are convinced that this will solve the whole problem and in a critical essay on research evaluation, Claire Donovan (2007b, pp. 591-592) raises the question if the novel metrics should not be looked upon as palliatives. Still, bibliometrics seems to serve a certain function in impact assessment, at least for assessing research impact. When it comes to impact in a broader sense, other types of output measures should be used as indicators, which we will discuss next.

Dissemination of research and interaction between science and society

Different ways of disseminating research can be seen as output measures in that they can potentially lead to impact. “These can encompass the use of technology transfer mechanisms such as industry seminars, industry secondments … field days … participation in government committees and policy development processes, participation in industry and academy meetings and seminars, preparing popular publications, research consultancy work …” (Go8 2011, p. 12). In the reports from the impact pilot exercise for the REF 2014 in UK these kinds of indicators are presented as possible instruments that the universities can make use of for assessing impact - defined broadly to include social, economic, cultural, environmental, health and quality of life benefits (REF Pilot a 2009; b 2010; c 2010). Another UK example is discussed in an article by Matthew Kearnes and Matthias Wienroth on the Engineering and Physical Sciences Research Council (EPSRC) and its way of responding to the contemporary policy discourses concerning the impacts of public research funding. In order to supplement the methods of assessing research impact the EPSRC has introduced the notion of “pathways to impact”. Kearnes and Wienroth interpret this shift in framing as an “… alternative theorisation of the relationship between research and socio-economic impact in which basic science is cast as ‘underpinning’ long-term social impacts and an attempt to generate new metrics that can quantify the cumulative and non-linear effects of a broad portfolio of publicly resourced research” (Kearnes and Wienroth 2011, p. 167).

The assessment of the dissemination of research can be seen as a type of formative or process oriented evaluation, in contrast to summative or outcome oriented evaluation. It is important to distinguish between these two types as they serve different purposes and require different

methodological approaches. In a strict sense, process oriented evaluations do not measure effects of an intervention (Ramberg and Knall 2012).

The interaction between research and society has been the focus also for the FP7 SIAMPI project (see Spaapen and Van Drooge 2011). The SIAMPI approach is based on the concept of productive interactions as a way of assessing the social impact of research. The model assumes that for social impact to take place there needs to be contact between researchers and non-academic stakeholders. When this contact leads to an effort by the stakeholder to engage with research a productive interaction is considered to have taken place. When the stakeholder does something new or in a different way based on these productive interactions, research can be said to have had an impact.

Another similar approach is developed in an article on "public value mapping" by Bozeman and Sarewitz.[2] The aim is to "… provide an alternative to 'market failure' thinking that has been so powerful in science policy-making" (Bozeman and Sarewitz 2011). They argue that to assess the capacity of research programmes to achieve (non-economical) social goals we need to map public values. Several methodologies can be used for this purpose, for example, the application of a set of criteria that makes it possible to identify public values failure, i.e., "… when neither the market nor public sector provides goods and services required to achieve public values" (Bozeman and Sarewitz 2011). The model has largely a case-based approach.

Expert reviews

One way of assessing the impact of research is to obtain information from groups of people that have special insight into the field in question. This form is often used in combination with other methods. The information can be gathered in different ways.

Expert panel

An expert panel with relevant experience can both contribute to estimating what difference research has made and to giving feedback on how possible pathways have been used. As in the case with peer review there are several methodological problems that have to be considered.

One specific method using experts that is interesting in impact assessments is the Delphi method. It is a communication technique that uses experts, and the purpose is to get informed predictions. Opinion collection is achieved by conducting a series of surveys using questionnaires. The result of each survey will be presented to the group and the questionnaire used in the next round is built upon the result of the previous round.

Anecdotes

Even if anecdotal evidence cannot be quantified it can help to identify some aspects of research impact. Anecdotes that relate how a certain research project has benefited the society in some way also illustrate the variety of possible impacts.

Surveys

[2] The article is published in a special feature issue of Minerva where the other articles present examples of implementation of "public value mapping".

Stakeholder surveys

As stakeholders by definition represent certain interests, there are shortcomings with surveying to determine impact. However, there are also positive features: "… stakeholders' perceptions of impact and the willingness of a university to work with outside agencies can be valuable … Such surveys can also help identify ways of increasing impact by improving linkages, changing perceptions and removing impediments to the flow, exchange and use of knowledge." (Go8 2011, p. 15).

Commercialisation surveys

"… commercialisation surveys normally collect quantitative data relating to matters such as the number of staff devoted to technology transfer, spin-off companies, patents and other IP rights applied for or granted, and licensing income." (Go8 2011, pp. 15-16). Although this kind of survey provides quantitative data, it has to be critically analysed due to different types of biases. "Moreover, commercialisation as measured by surveys provides a very narrow perspective on research impact even in the confined context of achieving impact through business." (Go8 2011. p. 16).

Case studies

Case study as a concept is defined and used in several ways. Looked upon as a method, case studies in themselves often encompass a combination of different quantitative and qualitative methods. What characterises case studies is the detailed analysis of individual research projects, programmes or individual research institutes, etc. In this way case studies, like some other non-experimental methods, allow for relevant contextual factors in contrast to RCT designs. Case studies have the advantage of giving detailed information in a process oriented manner. That the level of aggregation is low can, however, be a limitation of its usefulness. There is also much to gain especially as case studies often provide us with insights that can help us to develop a better understanding of the process of research impact. In other words a methodology well suited for formative or process oriented evaluation.

A general challenge with case studies is the selection of cases. If the aim of the study is to use the cases (e.g., research projects within a research programme) as representatives for a full research programme, care has to be taken so that the cases really mirror the full programme. But case studies can be used for other aims. For example, if the aim of the impact assessment is to investigate if excellent scientific research has more or less impact than good scientific research, cases should be selected that represent this. And if the aim is to get a better understanding of how research transforms into impact (or not) it is appropriate to select cases that are known to or are expected to generate impact.

In this context the relatively well-known "payback model" could be mentioned. The model is built around case studies, and is a tool to facilitate data collection (surveys, interviews and document studies) and provides a common structure for each case study, and thereby facilitates a cross-case analysis. It was originally designed to capture socioeconomic impact of health services research, but has been adapted and applied in a number of studies outside health and medical research. It consists of two elements. Firstly, there is a model of the research process (from research idea to impact on people and society) indicating when impacts can be expected. Secondly, there are categories of benefits from research in which paybacks or impacts can be classified. They include both benefits

associated with the academic world (knowledge production and research capacity) and wider benefits for society. These paybacks exist in somewhat different categories depending on the focus of the impact assessment. In general terms they can be seen as (see Klautzer *et al.* 2011, compare with Hanney *et al.* 2004 where the payback categories are directed towards the health sector):

a) Knowledge (explicit and codified knowledge)
b) Impact on future research (capacity building, new methods, career development)
c) Impacts on policy (impact on policy making at national level within professional bodies and organisations)
d) Impacts on practice (individual behaviour)
e) Wider social and economic impacts.

Hindsight studies

Hindsight studies are a very special type in that they aim to trace the links backwards from the identified impact to the research that contributed to the impact. Although hindsight studies can appear to be less problematic than foresight studies, the critical points – attribution, time frame and the counterfactual argument – are as relevant for these studies too.

Economic Models

One way to approach the interest in assessing value for money is different types of economic models. Public policy evaluation from an economist's viewpoint is concerned with the goals for public policy and priority settings. The computing power available today and the amount of statistics collected makes it possible to use complex systems modelling in studying the intricate ways in which science has an impact on society. While a regression-based quasi-experimental method is highly attractive for economists, agent-based and network-based models might provide a new and better way to assess and guide research and innovation policies (Ramberg and Knall 2012, pp. 13-16).

Econometric analysis

Econometric analysis is primarily used for assessing impacts of research on a macro level. Advanced numerical analysis techniques based on large amounts of data give information about the overall economic impact that can help justify government funding of research. It does not, however, capture the more intangible impacts of research and it is not applicable for assessing the performance of individual research agencies or research programmes.

Cost-benefit analysis

Cost-benefit analysis is suitable to assess impact from projects and programmes, but can also be used in a broader context. The objective is to compare the costs of investment in research to the estimated economical benefits of the research. There are many methodological challenges with cost-benefit analysis, such as the time frame and the attribution problem, and the quality of different studies can vary a lot. There is also the question of what kind of assumptions have to be made in order to define indicators and a model of causality. It is crucial that attention is paid to this kind of methodological challenge as the assumptions that the model is based on will highly influence the conclusions that can be drawn.

Questions asked and appropriate methods

Which purpose the impact study aims to fulfil affects the choice of method, and makes some approaches more appropriate than others. For example, the payback model is more oriented towards showing value for money. The SIAMPI approach is more directed towards the improvement of understanding the process of how research transforms into impact. For example, in one study using the SIAMPI model, one of the results was that researchers whose research was assessed changed their future research practice taking the perspective of stakeholders into account and stakeholders increased their contacts with researchers (Molas-Gallart and Tang 2011). Consequently, it is difficult on a general level to make recommendations on which method to use. The table below gives some indication of which method to use for a certain question, although it should not be seen as an exhaustive list.

Relevant Questions	**Methods for Answering Questions**
How much has been spent thus far? Does the progress achieved thus far match expectations based on those expenditures?	• Cost-benefit analysis
How are resources to be transformed into desired outputs and outcomes?	• Peer review/Expert judgment • Case study • Econometric studies
Is the programme's research of high scientific quality? Is it relevant, productive and well managed?	• Peer review/Expert judgment
What relationships are developing? Is the programme strengthening the research network?	• Network analysis • Before-and-after applications
What additional project-related relationships have developed among researchers?	• Network analysis • Before-and-after applications
How are programme mechanisms, processes, and/or activities working? How can they be strengthened?	• Monitoring activities • Case study - descriptive/exploratory • Econometric studies
What are the programme's codified knowledge outputs?	• Bibliometrics
How does the programme's output productivity compare with similar programmes?	• Benchmarking
How noteworthy are the resulting patents? What are the hot trends? Are there important regional impacts?	• Hot-spot patent analysis
To what extent have the programme's outputs	• Indicators

been commercialised?	• Technology commercialisation tracking
What factors are influencing industry's adoption/lack of adoption of the programme's technologies?	• Case study - descriptive/explanatory
What are the realised benefits and costs of the technology to date? What share of net benefits from the technology is attributed to the programme?	• Benefit-cost analysis
How is the programme working thus far?	• Case study - descriptive/explanatory
Are there one or more noteworthy consequences that can be shown to link back directly to the research?	• Hindsight studies, Historical tracing (including citation analysis)
If we had it to do all over again, would we have launched the programme or initiative?	• Peer review/Expert judgment supported by multiple retrospective evaluation methods
What benefits are there for society in general? Is the action (e.g., a research programme) showing value for money?	• Economic models (on an aggregate level) • Case studies (based on research projects, for example, the payback model)
How does research affect policy makers?	• The SIAMPI approach
What health gains does research create?	• The payback framework
How is the knowledge produced used in policy?	• Interviews, document analysis

5. Recommendations

It is clear to see that impact is difficult to assess, and there are several ways to conduct an impact evaluation. There is a wide discussion about "best practice" or "good practices". In this chapter the working group will summarise the recommendations on when and how to perform impact studies. It should be noted that these recommendations are primarily written for research funders and research performing organisations.

Assessments of impact are an important element in demonstrating, understanding and facilitating the value of research. Furthermore, impact should be seen as a broad concept, both in regard to in what areas impact can occur and what is to be considered as impact. This was discussed in detail in chapter 2.

However, an assessment of impact faces several methodological challenges and we have above – on a general level - dealt with how to cope with the most demanding ones. On a more specific level – when a certain assessment of impact is to be performed – there is no single method to be preferred. Instead it is a matter of finding an appropriate method, taken the recommendations on the general level into account, depending on what kinds of impacts are assessed. Furthermore, it is also, considering the difficulty and complexity to assess impact, advisable not to stick to one method in an impact assessment but to use different approaches.

It is also advisable to see the potential in impact evaluations as a tool to create better informed policies by research funders, better opportunities for the creation of scientific knowledge and the facilitation of the process where scientific knowledge produces benefits for the wider community. This does not mean that impact evaluations are an appropriate tool for showing value for money, but it is advisable that impact evaluations are constructed so that they have something to say about how benefits for society are better achieved. Another conclusion and recommendation is that the first thing to do when conducting an impact assessment is to carefully specify and select what is to be assessed rather than how to assess. This is not to say that how to measure is not important, it is very important. But it plays second fiddle to the question of why and what to measure.

A current trend in impact assessments is to find indicators of future impact. This is an area that is fruitful to develop further, because if there are stable indicators that serve as a proxy for longer-term impact assessments will be an even more useful instrument to monitor programmes and make better informed policies. However, as argued earlier, it is still important to continue to do studies that employ a longer perspective on how research transforms into impact.

There is also the question of cost for the impact evaluations. Obviously, the cost should never exceed the benefits that can be attributed to the impact study. It may be difficult to estimate the benefits, but this perspective must be taken into account. Also, the costs of the activities being assessed must be put in proportion to the cost of the study. Consequently, as in all forms of evaluations, before an impact study is carried out its usefulness must be put in perspective to its cost and the cost of what is being assessed.

There are also some further recommendations that the working group would like to highlight:

- Continuing cooperation between funding agencies and researchers on impact evaluation, as this will strengthen the activities both of research funders and researchers.
- Funding agencies should prepare for impact evaluations that are long term and also include a wide range of different impacts. A monitoring of impacts will provide valuable knowledge to further impact assessments and to the process of how research turns into gains to society.
- More focus on how to establish the counterfactual position, as this is one of the areas that can be stronger in impact assessments of research.
- Perform an impact evaluation that involves different countries and research agencies. It could be similar programmes that are assessed with the same approach. This could be the next step for European cooperation within the area of research evaluations.

References

Academy of Finland (2009). *Impact Evaluation of Finnish Programmes for Centres of Excellence in Research 2000–2005 and 2002–2007*.

Bozeman, B. and Sarewitz, D. (2011). Public Value Mapping and Science Policy Evaluation, *Minerva* 49, pp. 1-23.

Canadian Academy of Health Sciences (2009). *Making an Impact – A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*.

Davies, H., Nutley, S. and Walter, I. (2005). *Assessing the impact of social science research – conceptual, methodological and practical issues.* A background discussion paper for ESRC Symposium on Assessing Non-Academic Impact of Research, May 2005, 24 p.

Delanghe, H. and Teirlinck, P. (2010). *Optimizing the Policy Mix by the Development of a Common Methodology for the Assessment of (Socio-) Economic Impacts of RTDI Public Funding*, Idea Consult.

Donovan, C. (2007a). Introduction: Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact, *Science and Public Policy*, 34(8), pp. 538-542.

Donovan, C. (2007b). The qualitative future of research evaluation, *Science and Public Policy*, 34(8), pp. 585-597.

European Commission (2010). *A vision for strengthening world-class research infrastructures in the ERA.*

European Science Foundation (2009). *Evaluation in National Research Funding Agencies: approaches, experiences and case studies - A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes*.

Exceptional Returns: The value of Investing Health R&D in Australia II (2008). Prepared for the Australian Society for Medical Research by Access Economic, Canberra 2008.

Go8 (2011). *Measuring the impact of research – the context for metric development*, Go8 Backgrounder 23, 20 p. Australia, The Group of Eight. www.go8.edu.au

Grant, J. *et al.* (2009). *Documented Briefing, Capturing Research Impacts - A review of international practice.* RAND Europe: DB-578-HEFCE Prepared for the Higher Education Funding Council for England. http://www.hefce.ac.uk/media/hefce/content/pubs/2009/rd2309/rd23_09.pdf

Hanney S.R., Grant, J., Wooding, S. and Buxton M.J. (2004). Proposed method for reviewing the outcomes of health research: the impact of funding by the UK's Arthrtis Research Campaign, *Health Research Policy and Systems*, 2004, 2:4.

Kearnes, M. and Wienroth, M. (2011). Tools of the Trade: UK Research Intermediaries and the Politics of Impacts, *Minerva* 49, pp. 153-174.

Klautzer, L., Hanney, S., Nason, E., Rubin, J., Grant, J. and Wooding, S. (2011.) Assessing policy and practice impacts of social science research: the application of the Payback Framework to assess the Future of Work Programme", *Research Evaluation* 20 (3).

Levitt, R. *et al*. (2010). *Assessing the impact of arts and humanities research at the University of Cambridge*, RAND Europe.

Langfeldt, L., Brorstad Borlaug, S. and Gulbrandsen, M. (2010). *The Norwegian Centre of Excellence Scheme – Evaluation of Added Value and Financial Aspects*, Rapport 29/2010. NIFU STEP, 2010.

Luoma, P. et al. (2011). *Better results, more value – A framework for analysing the societal impact of Research and Innovation*. A Tekes Review 288/220.

Medical Research: What's it worth? Estimating the economic benefits from medical research in the UK (2008). Health Economics Research Group (HERG), Brunel University, Office of Health Economics (OHE), RAND Europe.

Molas-Gallart, J., Tang, P. and Morrow, S. (2000). Assessing the non-academic impact of grant-funded socio-economic research: results from a pilot study, *Research Evaluation* 9(3), pp. 171–182.

Molas-Gallart, J. and Tang, P. (2011). Tracing 'productive interactions' to identify social impacts: an example from the social sciences, *Research Evaluation* 20(3), pp. 219–226.

Ramberg, I. and Knall, M. (2012). *Challenges measuring effects of research and innovation policy interventions in ex-post impact evaluations. A synthesis report.* 4 March 2012. Oslo: NIFU.

RAND Europe (2004). *The returns from arthritis research - A report prepared by RAND Europe for the Arthritis Research Campaign (arc*).

REF Pilot a (2009). *Identification and dissemination of lessons learned by institutions participating in the Research Excellence Framework (REF) bibliometrics pilot*. Results of the Round Two Consultation – report to HEFCE by Technopolis. 2009.
http://www.hefce.ac.uk/media/hefce/content/pubs/2009/rd1809/rd18_09.pdf

REF Pilot b (2010). *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions*. Final report. www.technopolis-group.com. 2010.
http://www.ref.ac.uk/media/ref/content/pub/refresearchimpactpilotexerciselessons-learnedprojectfeedbackonpilotsubmissions/re02_10.pdf

REF 2014 c (2010). *Research Excellence Framework impact pilot exercise: Findings of the expert panels. A report to the UK higher education funding bodies by the chairs of the impact pilot panels. 2010.*
http://www.ref.ac.uk/media/ref/content/pub/researchexcellenceframeworkimpactpilotexercisefindingsoftheexpertpanels/re01_10.pdf

Spaapen, J. and Van Drogge, L. (2011). Productive interactions as a tool for social impact assessments of research, *Research Evaluation*, 20(3).

Swedish Research Council (2010). *Evaluation of Impacts of Medical Research*, Stockholm, Swedish Research Council: Report Series 9:2010.

Tassy, G. (2003). *Methods for Assessing the Economic Impacts of Government R&D*, USA, National Institute of Standards & Technology: Planning Report 03-1.

UK Evaluation Forum (2006). *Medical Research: assessing the benefits to society*.

List of MO Forum Working Group Members

List of Members of the WG

Chair:	Per Janson	Swedish Research Council (VR)
Members:	Mustafa Ay	The Scientific and Technological Research Council of Turkey (TÜBITAK)
	Frank Bingen	Fonds National de la Recherche (FNR), Luxembourg
	Sarah Chen	The National Research Council, Italy (CNR)
	Helen O'Connor & Graham Love	Science Foundation Ireland (SFI)
	Christian Fischer	Austrian Science Fund (FWF)
	Marianne Gauffriau	Danish National Research Foundation (DNRF)
	Isabelle Henry	Institut national de la Santé et de la Recherche Médicale (Inserm)
	Inger Jonsson	Swedish Council for Working Life and Social Research (FAS)
	Paul reeves	BBSRC
	Jaana Roos	Academy of Finland
	Mike Mugabushaka	European Research Council, Observer
	Briony Rayfield & Liz Allen	The Wellcome Trust, Observers

Acknowledgements

This report has been compiled by Per Janson (VR). ESF is grateful to him – the Co-Chair of the MO Forum Working Group - and to the significant input from working group members who contributed to lively and informative discussion in the ESF Member Organisation Forum over the last four years as well as for its management to Laura Marin (ESF). In particular, the drafting of this document relied upon the contributions of Sarah Chen (CNR), Inger Jonsson (FAS), Mustafa Ay (Tübitak), Helen O'Connor (SFI), Marianne Gauffriau (DNRF) and Isabelle Henry (Inserm).