Annex 1

Evaluation of tax incentives for R&D: an overview of issues and considerations

A paper produced by the Crest OMC Working Group "Evaluation and design of R&D tax incentives"

1 Introduction

This paper is part of the work of the OMC Working Group on the evaluation and design of tax incentives for R&D. The purpose is to give a general overview and considerations concerning the evaluation of tax incentives for R&D. More in particular, the intention is to highlight important stages and choices in the evaluation process that policy makers and operating agencies might find useful to take into consideration. The evaluators themselves will have more detailed and operational knowledge than what is covered here.

There is no single recipe for doing evaluations – evaluations of tax incentives for R&D are no exceptions. Although the primary objective most often is to increase R&D spending within firms, there may be a range of other objectives that evaluations should capture. In addition the national circumstances may vary and evaluations of tax incentives should be seen in the broader context of the national innovation systems and mix of government tools for innovation and R&D. Last – but maybe most important – access to various data sources varies across countries and can have a large impact on how evaluations may be carried out.

Evaluations give a stronger knowledge base for making policies, but evaluations in themselves do not create policies. There has to be a translation of findings from evaluations to policy making, and policy makers have to understand and be able to use findings from the evaluations in their work. Evaluations should therefore be seen as a tool for learning that should be used in a proper way to improve policy making. Since the political systems differ across EU/ CREST member states this makes for a wide range of approaches in the follow up of evaluations.

Evaluations do not give answers that are final. This goes for evaluations in general – and for evaluations of both direct and indirect R&D support measures. Evaluations can, however, be used both to develop a stronger knowledge base for targeting certain policy aims and objectives and in order to create a stronger empirical basis for research concerning innovation and R&D.

These are all themes that are addressed in the report of the OMC Working Group on the evaluations and design of R&D tax incentives. This paper is not meant as a detailed handbook, but shows important items to be aware of when planning and doing evaluations of tax incentives for R&D.

From the outset the Working Group had as its mandate to produce a report. During its work, the group was requested to expand its work on evaluation of tax incentives. Because of the need for policy makers to get a better understanding of how evaluations might be carried out, it was decided to do this through a paper with an overview of issues and general planning-advice on evaluation methods. In this process, the Working Group had a need to get on board more expertise, and it is greatly appreciated that Torbjorn Haegeland, Head of Research at Statistics Norway, and Bruno van Pottelsberghe, Director at the European Patent Office, agreed to take part in the final rounds of discussion on the report and the handbook. Torbjorn Haegeland is also the main contributor to the writing of the handbook.

This paper is not intended as a detailed handbook and should not be taken as such. Rather, it aims to highlight and explore important issues that might usefully be considered when planning and conducting an evaluation of tax incentives for R&D.

2 Considering evaluations when designing the scheme

Tax incentives for R&D are comprehensive support measures – and the tax foregone can be considerable. This calls for evaluations that might map results, show how desirable outcomes are met and whether the tax incentive causes increases of R&D investments. When designing the scheme it would therefore be sensible to have in mind that the scheme will be evaluated. The operating agency could then put in place solutions for databases that are adequate and could be used together with national statistics. Adequate resources should also be put aside to ensure that the evaluations are of high quality. This might be costly – but the benefits for policy making could be significant.

Tax incentives for R&D are often general in nature, based on a belief that the scheme is going to work in the desired way, and therefore should be available to all. However, policy makers usually want to or are obliged to evaluate the scheme because they are to some extent uncertain whether it works or not.

The main challenge for evaluators is that the more general a scheme is, that is the more equal the scheme treats different firms, the more complicated is the evaluation. The reason is that a higher degree of "generality" or "equal treatment" brings evaluations further away from the ideal setting where firms, under otherwise similar conditions, could be compared according to whether they are eligible for tax incentives or not.

Ideally, evaluations should be built into the design of the scheme itself. However, an evaluator typically comes in after the scheme has been designed and implemented, maybe even after the scheme has been in place for some time. Seen purely form the perspective of an evaluator the question of whether a specific measure works or not should be answered by carrying out a controlled experiment, randomly dividing the population of firms into two groups and giving just one group access to the scheme. This is however politically unfeasible in most or all countries.

The following chapters consider how this problem may be solved through proper collection of data and the use of a combination of evaluation methods. These different methods also require different data. This underlines the importance of the evaluation being considered as early as possible in the policy making process (and ideally before the policy is introduced) to ensure that the most appropriate approach can be taken and the necessary data collection procedures put in place.

3 Organising the evaluation

3.1 When should the evaluation be carried out?

Careful consideration should be given as to when the evaluation should be carried out. This will vary from one tax incentive measure to the next as well as between the specific questions within a single evaluation.

The timing of an evaluation should take into account a range of issues, such as:

- how long it will take before the tax incentives is taken fully into use by businesses;
- how long it will take for the expected policy effects to emerge;
- how long it will take for the necessary data to become available; and
- when the results are required for policy purposes (including any public commitments that might have been made)

One of the most important determinants of when an evaluation should be performed is when it will be possible to estimate the different effects of the policy. This will reflect both the period over which effects emerge (at least sufficiently to be quantified) and the time needed for sufficient data to become available to measure the effect. The additionality effects of R&D tax incentives, for example, can take a number of years to emerge and the methods often used to measure such effects can require a number of years of data. The risk of attempting such an evaluation too early is that misleading results might be produced, for example showing no policy effect when it is simply too soon for effects to show up.

On the other hand, however, it is important for an evaluation plan to recognise that results are needed to inform improvements to the policy and so, as far as possible, evaluations should also aim to produce some results as early as possible. This might be in areas where effects can be seen and assessed more quickly, such as on the administration and delivery (for example, testing the awareness and take up of the policy and the ease with which firms can understand and claim support)

3.2 Planning the evaluation

Evaluation should be planned so that it is understood how the different evaluation questions will be tackled, which methods will be used, how the different methods will fit together and when results will become available.

Evaluation planning should start as early as possible, i.e. when the policy is being developed. This should help to ensure that any data requirements can be built into the policy from the start and any data on the pre-policy situation can be collected.

The planning of an evaluation might include factors such as:

- evaluation objectives and questions;

- potential evaluation methods to tackle these questions;
- consideration of the counterfactual and how it will be measured;
- data that will be needed and where it will come from;
- how the different methods will fit together;
- the expected outputs from the evaluation (such as reports);
- what results will be published, how and when;
- the timescale for the evaluation; and
- the cost of the evaluation

It may be helpful to draw up an evaluation plan or strategy setting these issues out. This will form a basis for reference when it comes to actually performing the evaluation and will provide a benchmark against which progress can be measured. An evaluation plan can also help to manage expectations amongst policy makers and politicians of when they can expect different results to become available.

It can also be useful to open the evaluation plan to peer review, such as from policy makers and other evaluation experts who can provide advice. This should help to test, for example, whether the evaluation questions will inform policy improvements, whether the methods are appropriate, and whether the timetable is practical and realistic.

3.3 Who should perform the evaluation?

For an evaluation to be effective, it needs to be robust, relevant, candid, credible and unbiased. In addition, it is important that the evaluation is seen to possess these characteristics in order for its results to be believed and thus have value. The way the evaluation is performed plays a central role in determining the credibility of the results and so this issue warrants very careful consideration.

There are a number of different ways that an evaluation can be performed. For example, it could be done entirely by the authority responsible for the policy, or done entirely outside of the authority responsible for the policy, or a mixture of the two.

There are advantages and disadvantages to internal and external evaluation and the decision should take account of a number of issues, such as: expertise and experience in performing such evaluations, knowledge of the policy area, history and traditions of evaluations in that country, and the independence of the evaluators and evaluation.

3.3.1 Internal evaluations

Internal evaluations can represent a more efficient use of resources, particularly where the evaluators already have expert knowledge of the policy and the data, as well as ready access to the data and other policy information. This may be cheaper than commissioning the work externally, where such knowledge may need to be built up.

However, internal evaluations may risk a favourable bias towards the policy – either unconsciously where the evaluators were involved in the design of the policy or more directly if policy makers are able to influence the evaluation. Crucially, even if internal evaluators are unbiased, they may be seen as being partial or biased simply because of

their association with the policy. Internal evaluation may also be of poorer quality methodologically or less efficient, if internal evaluators are less aware of the latest developments in evaluation methods in the area or where the absence of competition has allowed inefficiencies to survive.

Some of the disadvantages of internal evaluation can be resolved by exposing the evaluation to some level of public scrutiny. This could be through some form of peer review, either of the end product of the evaluation or at various interim stages whilst the evaluation is carried out. Making the results of the evaluation publicly available may help to raise the perceived quality of the work, although it may still leave some doubts over the quality of the methodology, particularly amongst those not experienced in evaluations.

3.3.2 External evaluations

External evaluation should be less at risk of suffering from bias either for or against the policy in question, although it is essential that the selection process tests the independence of external evaluators. Because of their expertise and experience, external evaluators should also be aware of the latest methods and techniques in policy evaluation and may bring new ideas to the evaluation of a policy. The process of competition should also promote efficiency in carrying out evaluations.

However, external evaluation might be perceived to be more expensive, perhaps reflecting the need to build up policy-specific knowledge. External evaluators may also be less aware of nuances of individual policies, which may mean they are less able to interpret results effectively and may be less able to ensure that results focus on specific interests of policy makers. In some areas, there may also be a limited supply of external evaluators with the necessary expertise.

Where all or part of the evaluation is carried out externally, the task or project should be awarded on the basis of a competition between potential evaluators. This competition should be based on a detailed description of what issues the evaluation should focus on, what outputs should be produced and to what timescale. The team of evaluators should demonstrate hands-on knowledge of relevant methods, practical experience of evaluation expertise in the area required and be independent. Close connections to academic research may also be an advantage.

During an external evaluation, there should be close contact between the evaluator, the agency administering the scheme, and other data-producing agencies. Key contacts should be identified to ensure the evaluation runs smoothly on a day-to-day basis, with regular contacts or meetings with a wider group of stakeholders to check progress at key stages. The evaluator should also present regular interim reports on specific parts of the evaluations. Regular updates and contacts are important both to check the evaluation progresses on time but also to ensure it will tackle the right evaluation and policy questions as well as testing the quality of the methodology. It also provides opportunities for the external evaluator to ask questions and gather any additional policy information required to interpret results correctly.

These findings from the evaluation could also be discussed in peer review to subject the evaluation to expert testing of both the methodology and the results.

3.3.3 A combination of internal and external evaluation

It is also possible for an evaluation to be carried out partly internally and partly externally. This might involve the evaluation being managed internally, with some parts conducted internally and some parts commissioned to external evaluators.

The aim of such an approach would be to maximise the advantages of both internal and external evaluations, whilst minimising both their disadvantages. This is possible, but requires much greater planning and management to ensure that the evaluation remains well co-ordinated overall and it is the advantages rather than the disadvantages that are maximised. For example, the internal evaluators may be better able to ensure the evaluation meets the needs of policy makers whilst the external evaluators are seen to be independent. However, there may be a risk that the internal part of the evaluation reduces the perceived independence of the external evaluators.

The final choice of who carries out the evaluation will, inevitably, reflect the tradition and structure of policy evaluation in each country. Whoever carries out the evaluation, however, it is important that the process is managed to ensure that it is, and is seen to be, methodologically robust, policy-relevant, candid and independent.

3.4 Independence and publishing evaluations

One of the most important ways to promote the perceived independence of an evaluation is for the reports from the evaluation to be published (covering the data, methodology, results and conclusions). The evaluation reports could be published either by the evaluator carrying out the evaluation or the authority responsible for the policy, but it should make clear who carried out the evaluation and any role that the policy authority played in the evaluation to ensure the independence of the report is clearly understood.

As noted earlier, another way to promote both the quality and independence of an evaluation is for the work to be discussed in peer review. The review might include users of the scheme, the agency administering the scheme, other government representatives, and other external researchers and evaluators. A wide ranging review is helpful to test not only the robustness of the methodology and significance of results, but also how the results should be interpreted, the implications for the policy and any potential improvements they suggest.

If independence is the most important criteria for an evaluation, a different approach could be to make the evaluation data base accessible to all external evaluators and allow any of them who are interested to carry out their own evaluation free from government interference. This has the clear advantage that the evaluation would be entirely independent but also risks the evaluations tackling questions that are different to the ones identified by policy makers. Thus, this approach could promote wider research in a particular area in general but may not provide the specific answers that policy makers need. This could be tackled by the authority publishing the questions they are most interested in, although they would not be able to force researchers to tackle those specific questions. In addition, there may also be some legal constraints on making data widely accessible due to confidentiality issues.

4 Issues to be evaluated

The main task for evaluations of government measures is to find out to what extent the objectives of the measure is fulfilled, and whether the benefits to society generated by the measure are larger than the costs involved. Evaluations of tax incentives for R&D should of course have the same focus.

The main objective for tax incentives for R&D is generally the same across countries: to generate benefits for society as a whole by inducing an increased level of R&D investments, with returns exceeding costs involved. Though they share the same main objective, different schemes have different specific objectives. These specific objectives may be regarded as different strategies to achieve the main objective of the scheme. Examples of specific objectives may be to increase the level of R&D in SMEs, to stimulate collaboration between firms and R&D institutes or to stimulate the creation of knowledge-based, research-intensive firms. Since evaluations should focus on the objectives of the measure in question, the issues to be evaluated will vary between countries, although some of the main components will be identical.

In the following, an extended checklist of issues that should be included in most evaluations is presented. Following this issues that could be included, depending on the objective of the scheme is shown.

4.1 General evaluation issues

4.1.1 Input additionality: Does the scheme generate more R&D?

Input additionality is defined as the effect the tax incentive in question has on the R&D investment of its beneficiary firms. The question is whether the support from the scheme creates any *additional* R&D investment over and above the level of investments that would have been realized without the scheme, and if so, whether firms increase their R&D investments by more than the amount received in support through the scheme.

The main objective of any R&D-stimulating scheme is to increase the level of R&D. Analysis of input additionality will therefore be a key point in any evaluation. Other aspects and effects of the scheme either depend on or must be viewed in light of the effects on R&D investments. If the scheme fails in generating more R&D, the whole scheme should be called into question. One may therefore say that a positive input additionality is "a necessary condition for success".

Introducing fiscal incentives for R&D may give firms incentives to report more R&D than they did before, even without a change in the R&D level. This should be taken into account when analyzing input additionality.

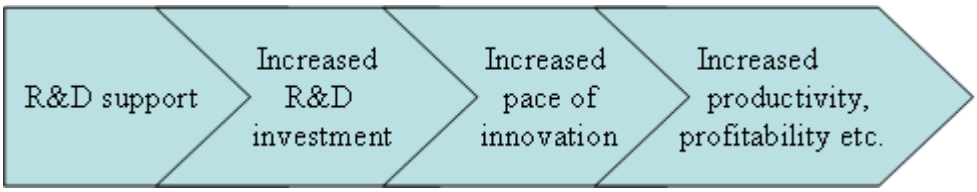
4.1.2 Result or output additionality: What are the effects?

Increased R&D investments are a policy objective in many countries. There is abundant theoretical and empirical evidence that the market alone provides too little R&D. Input additionality is therefore a central objective for the scheme. However, R&D is not an end to itself. The justification of the tax incentive depends on the return on investment.

Evidence suggests that the returns to such investments on average are high. However, the literature also points out that there is vast heterogeneity in the returns to R&D investments. This is partly due to general idiosyncrasies, but there may also be systematic differences between types of investments. Hence, there may be differences between types of government support with respect to the results of the investments they generate. Input additionality may be a necessary condition for the success of a scheme, but not a sufficient condition.

It is therefore necessary to evaluate output or result additionality: whether and how the R&D investments cause improved economic performance of the firms undertaking them and for society as a whole.

One may think of the following "chain of results" for a tax incentive:



The first effect is input additionality, while the subsequent effects relate to output/result additionality. In addition to increased pace of innovations this might be increased quality and quantity of innovations. There may also be considerable external effects, i.e. effects accruing to others.

It should be noted that result additionality is more demanding to identify and quantify than input additionality. The main reason for this is that effects are in general more delayed. In addition the effects of R&D may be difficult to distinguish from other factors that are important for profitability and enterprise – external effects are particularly difficult to ascertain. There may also be a range of other objectives for tax incentives that give reason for evaluations that cover other results than those described here.

4.1.3 Behavioral additionality: Do firms change their R&D strategy?

Introducing a tax incentive means changing the external conditions for firms. The two above points concern how firms respond to these changes in terms of R&D investments, and how these investments pay off.

To fully understand the effects of tax incentives, and to understand how additionality effects are produced, it is necessary to look into firms to see to what extent the introduction of a new incentive changes the orientation and perspectives within firms, thereby influencing their goal and strategies. This is called behavioral additionality, and the use of this approach in evaluating tax incentives is relatively new. There has been established a Working Group under the auspices of OECD/ TIP which try to develop further the concept of behavioral additionality, and the use of this in evaluating government support measures for R&D.

Specific questions to be analysed under this heading may include:

- Has the introduction of the scheme resulted in a change in firms' R&D decision process?

- Does learning about benefits of R&D lead to sustained higher levels of investment?

4.1.4 Administrative costs and efficiency

Tax revenue forgone or subsidies paid constitute the major part of the cost side of tax incentives for R&D. However, administrative costs, both for the government agencies administering the scheme and the firms participating in it, may also be substantial. An evaluation should therefore also look into administrative efficiency. One way of doing this may be to evaluate the different steps involved in transferring policies to measures used by the firms. In all parts of this chain there may be different types of costs according to how the tax incentives are designed.

In the administration of a tax incentive, there are some inherent conflicting interests that must be weighed against each other. One the one hand, governments wants to attract firms to the scheme, given the objective of increasing R&D investments. Since firms ideally take all costs and benefits into account when making decisions, an increased administrative burden associated with participating in the scheme may make firms decide to stay away. On the other hand, there is a need for control systems to avoid abuse of the scheme. The evaluation should focus on this trade-off and assess whether there is any potential for making the administration of the scheme more efficient.

4.2 Examples of more specific evaluation issues

4.2.1 How the scheme works together with other R&D measures

A tax incentive for R&D is seldom introduced in a vacuum. Most countries have other policy measures directed toward R&D. In the process of designing an "optimal R&D policy", it should be evaluated how different measures work together. Important questions may be:

- Are the different measures substitutes or complements?
- Are there differences between measures with respect to which firms they attract and their effect on R&D?
- Are there differences between measures with respect to what type of R&D is carried out?

4.2.2 The role of research institutes

A number of schemes provide extra incentives for co-operation between firms and research institutions – universities, public research organisations and institutes. The rationale for this differential treatment is an assumption of larger additionality, and that such co-operation projects may stimulate knowledge flows from R&D institutes. If such extra incentives are a central part of a measure, they should be included in the evaluation of the wider scheme.

4.3 Policy recommendations

The objective of the evaluation is to enable policy-makers to make decisions on R&D policy on a more informed basis, based on the results from all parts of the evaluation.

Policy recommendations must take into account that there are variations as to what extent different effects can be observed and identified. When effects occur may also vary and not all effects can be expected to be observed within the evaluation period.

5 Methods of evaluation

This section gives a brief introduction to different methods that are widely used in evaluations. It should be noted that what follows is far from a detailed "how-to-evaluate-tax incentives" guide. It is rather an attempt to give a short and fairly non-technical description of the different approaches, pointing out strengths and weaknesses of the different methods in addressing different aspects of the scheme.

What should be learned from this section is that there is no "magic bullet." There is no single method that can be used to throw light on all the aspects of the scheme that is to be evaluated. Different methods have their strengths and weaknesses with respect to the questions they are suited to analyse, and their costs, data requirements, etc. The design of the tax incentives themselves may also complicate evaluations when it comes to the identification of causal effects.

5.1 Different methods of evaluation

An authoritative survey article on evidence on tax incentives for R&D is Hall and van Reenen (2000). The focus in their article is on econometric evidence, but they also discuss other methods. In the following, some of these approaches are discussed, drawing heavily on Hall and van Reenen (2000).

5.1.1 Event studies

Event studies will typically consider the launching of the tax incentive as a sudden and surprising event for firms. Under this assumption, one can measure effects of the scheme by doing “before-after” comparisons.

The most common outcome variable used in event studies is the market value of firms, thereby measuring how the stock market estimates the value of the scheme in terms of returns that accrue to firms. In many circumstances, this is probably not an adequate evaluation method. First, a scheme is seldom launched as a sudden event, but is often the result of a long debate and is announced in advance. This creates expectations that may affect the valuation of firms. Second, the method is relevant only for publicly traded firms, where information on market value is readily available. Third, the change in the private valuation of firms will provide limited information about the social returns to the scheme, and hardly anything about to which extent the scheme stimulates R&D investment (input additionality) or behavioral additionality.

One possibility is to do event studies with R&D investments as the outcome variable. In such studies R&D investments after the scheme is launched are compared to the investment level that was planned before the tax incentive was announced. (Several countries gather such information in their R&D surveys). Though the methodology used in event studies is intuitively appealing, it has its weaknesses. The most obvious weakness is that it is difficult to control for the effects on the outcome variable of other events or trends that appear simultaneously with the event the evaluator focuses on (i.e. the introduction of the scheme). In our setting, there may be other reasons for deviations

between planned and actual R&D than the tax incentive. Event studies are most suitable to study sudden events where effects materialize quickly. Neither of these criteria is likely to be fulfilled in the case of tax incentives for R&D.

5.1.2 Case studies and surveys

Case studies/surveys are an important tool for assessing respondents views or perspectives on a scheme, for example, whether they understand the scheme and how easy they find it to claim support. Such evaluation questions can only be tackled by these methods.

Hall and van Reenen (2000) consider case studies as "retrospective event studies". The method is straightforward; involved actors are asked whether the launching of a tax incentive had any effect on variables and factors that are of interest to the evaluator, e.g. R&D investments (input additionality), innovations, profits etc (result additionality), R&D decision processes (behavioral additionality). The major advantage of this approach is that respondents implicitly control for other external conditions when they answer the questions. In this context surveys refer largely to qualitative type surveys, rather than the surveys used to collect large datasets that can then be used to perform quantitative analysis. An advantage is that these types of data can be used to capture more contextual and in depth aspects of the tax incentives, which may not be easily identifiable through other data sources. In addition surveys and case studies can be used to follow the firms that use the tax incentives through different stages. In this way the evaluations will also get time series data and a better grip on the development within firms.

Case studies and surveys are very useful, and should probably constitute an important part of any thorough evaluation. However, they do have shortcomings and should be supplemented by other methods. As Hall and van Reenen point out, respondents may have the incentive to answer strategically. If they feel that their response may influence the continuation of the scheme, they may adjust their answer accordingly. For example, if they think that a finding of a large input additionality increases the probability that the scheme will be continued, they may exaggerate the effect on R&D investments. Even if the respondents do not answer strategically, it is far from obvious that they are able to isolate the effect of the tax incentive from other factors. This may not be a problem if the error is not systematically related to the real effects of the tax incentive. However, this may not be the case. For example, there might a tendency that managers with positive results overestimate the effects of their own effort, while those with negative results exaggerate the impact of external factors. Another potential error may be that survey respondents remember past events inaccurately. A challenge in this kind of approach is to determine who the best person is to speak to in a company about R&D tax incentives (particularly large companies).

5.1.3 Econometric studies

Econometric methods may be used in analyzing a wide range of issues relevant to an evaluation of tax incentives, but their highest relevance and most widespread use is in the analysis of input additionality and result additionality. The discussion is concentrated around these issues.

Input additionality

Hall and van Reenen focus their discussion on estimating the input additionality effect, through econometric estimation of demand equations for R&D. Within this framework, there are basically two approaches. The point of departure for both is a regression equation that predicts R&D investments at the firm level.

In the first approach, the demand equation includes a variable that indicates whether the firm had access to the tax incentive, in addition to other variables that affect R&D investments:

$$\ln(R\ \&\ D)_{it} = \alpha + \beta C_{it} + \gamma' X_{it} + u_{it}$$

This equation expresses the logarithm of the R&D investments of firm *i* in year *t* as a function of the presence of a tax incentive (C_{it} equals one if firm *I* had access to the scheme in year t and zero otherwise) and other variables, which are contained in the vector X_{it}. Such variables may be previous R&D investments, previous output and sales, expected future output, cash flow, product prices, etc. Whether the firms have been granted R&D subsidies through other channels will also be an important factor.

The β-parameter measures the expected growth of R&D investment following a firm getting access to the scheme. The basic framework assumes that this effect is identical across firms. This assumption is hardly innocuous. It may be relaxed but this often increases the data requirements.

Such models should be estimated on micro data, to utilize cross-section variations in access to the scheme across firms. Using macro data only, it is impossible to distinguish the effects of the scheme from unobserved macroeconomic shocks.

The second "demand equation" approach has very much in common with the one described above. The major difference is that instead of just including a variable indicating existence of or access to a tax incentive for R&D, one calculates the so-called "user cost" of R&D investments, i.e. a variable that reflects the price of R&D investments for the firm, on the margin, taking into account R&D tax incentives, other tax rules, interest rates and depreciation. The introduction of say a tax deduction scheme for R&D will reduce the user cost of R&D.

$$\ln(R\ \&\ D)_{it} = \alpha + \beta \rho_{it} + \gamma' X_{it} + u_{it}$$

The key advantage of the user cost approach compared to the first approach is that one may utilize variations in the generosity of the scheme between different firms, and also changes over time. Such variation may be very useful in identifying the effect of the scheme. In addition, variations in other components of the user cost (tax rules, interest rates, depreciation rates) may in theory help in identifying the effects of the tax incentive.

The equations above implicitly assume that the effects of the tax incentive are homogeneous across different firms. A relevant evaluation question is whether the effects differ with respect to observed firm characteristics, e.g. previous R&D experience, the education level of the staff, industry and location. If one should find that the scheme has larger effects for some types of firms than for others, it may influence the future design of

such schemes. In principle, it is straightforward to extend the framework described above to allow for such different effects. However, the number of observations may limit how many different additionality parameters that may be estimated with a reasonable degree of precision.

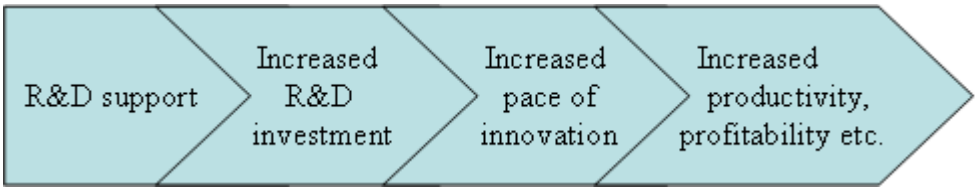
Some schemes have differential treatment for different types of R&D projects, e.g. a higher maximum tax deduction for joint projects with R&D institutes. The rationale behind this is a supposed higher effect of such projects. It is straightforward to extend the framework above to test this assumption.

The short-term input additionality effects may differ from the long-term effects. The long-term effects may be larger because there may be substantial adjustment costs in R&D investments. Such costs are due to the fact that it takes time to adjust the use of inputs to new external conditions. In addition, the long-term effect may be larger because an increase in R&D investments adds to the firm's knowledge base, thereby increasing the marginal payoff of future R&D investments. Evaluations should take these possible differences into account.

Result additionality

Econometric methods may also be useful in trying to evaluate the results (result additionality) of the R&D investments generated by the scheme. In some sense, the econometric framework for analyzing the results is similar to the framework for analyzing input additionality, but with some modifications.

As discussed in the previous section, one may think of the effects of the tax incentive that accrues to the firm as coming in the following steps:



The first step belongs to input additionality, while the next steps belong to result additionality. Hence there are many candidates for variables that may be indicators of result additionality. The number of patents is one obvious indicator. Productivity (labour productivity, total factor productivity), total sales, profitability, number of employees etc. are all variables where effects of increased R&D investments in principle should show up. Hence, they are candidates (among others) as dependent variables in the analysis of result additionality.

It is challenging to identify the effects of the tax incentive on R&D investments. Identifying the effects of the R&D investments they are in many respects even more demanding. The further along the result chain above that the evaluation goes, the more external factors may affect the result and it may be difficult, or even impossible, to control for these factors.

Results will probably come with delays of variable and unknown length (more on the timing issue below). This also makes it harder to identify the "result additionality". In

other words, the hope for clear and definitive answers is smaller the further away from the tax incentive the analysis is done.

The analysis of result additionality should take into account that one cannot distinguish clearly the effects of R&D investments generated by a tax incentive in a given year from other R&D investments, within the scheme or not, made by the firm. To account for this, the concept of R&D capital may be useful. The idea is to define R&D capital as the sum of past R&D investments net of depreciation, and include this variable in the econometric analysis. Then all the firm's R&D investments are included in the analysis, with a greater emphasis on recent investments.

A number of studies have found that the distribution of returns to R&D investments is skewed. Behind the average returns reported in many analyses, there is often a majority of projects with zero or low gross returns, thus failing from a commercial point of view, and a small number of projects with very high returns. This possible skewness should be taken into account, using methods that allow for heterogeneity in returns.

One important rationale for government support to R&D is the presumption that R&D investments generate positive external effects, i.e. effects that accrue to someone else other than those undertaking the investment. The standard econometric approach is to extend the models of result additionality with variables reflecting R&D investments outside of the firm, and to estimate to what extent such variables have an effect on different measures of firm performance. Identifying and quantifying external effects of R&D in general is a daunting task. To isolate external effects of R&D investments triggered by a specific tax incentive is even more challenging. Therefore, if analysis of external effects is a part of an evaluation, one should not expect any definitive findings.

5.2 The absence of random variation: A problem for all methods

In evaluations, the primary objective is to learn more about the effects of the measure or scheme in question. Does it work in the intended way? Should it be continued, modified or abandoned? The evaluation should give answers to such questions by uncovering *causal* effects. The situation *before* the scheme was introduced is known through the established data-sources. The present situation with the scheme in place is also known. However, finding the causal effect implies answering the counterfactual question: What would the situation have been now in the *absence* of the scheme. The methods briefly described above are in principle suited to do so, given that certain assumptions are not violated. The most important assumption is that firms' access to the scheme, or variations in generosity of the scheme over time and between firms are not related to factors which cannot be controlled for in the analysis, that are themselves related to the level of R&D investments – i.e. there must be some randomness in treatment. However, such variation may be rare, or in some cases even absent.

Not all "other factors" related to R&D investments cause problems. Many of them are directly observable and available, and can be controlled for in the analysis, e.g. by including them in the vector X in the equations above. Estimating the econometric equations in differenced form may also control for unobserved differences between firms that are constant over time. The main problem arises when firms' access to the scheme are correlated with unobserved factors which vary over time. This point makes it particularly

problematic to look at firms who apply for support through the scheme, using those firms that do not apply as a comparison group. Firms that get a good research idea are more likely to apply for support through the scheme, but they would also be more likely to carry out the project in the absence of the support. Hence, firms "within" the scheme and outside the scheme are likely to differ with respect "research ideas", a highly unobservable and time-varying variable.

Ideally, the question of whether a specific measure works or not should be answered by carrying out a controlled experiment, randomly dividing the population of firms into two groups, giving one group access to the scheme. This would provide the exogenous variation needed, and give a basis for comparing a treatment group with a control group using the above framework. This ideal situation is almost never feasible, cf. Jaffe (2002).

Schemes and measures are often general in nature. This creates great challenges for evaluators. The more general the scheme that is, the more equal similar firms are treated in the scheme, the more complicated is the evaluation. The reason is that a higher degree of "generality" or "equal treatment" brings evaluations further away from the ideal setting. When all comparable firms either have access to the scheme or not, it is impossible to construct a control group providing information about the counterfactual situation.

The challenge in a non-experimental setting, without a formal control group, is to deduce from historical data, what the situation would have been if the scheme had not been launched. In the absence of a controlled experiment, one needs to look for so-called quasi-experiments built into the scheme. A quasi-experiment might involve, for example, using variations in the scheme that may be regarded as "random" at least at the margin. In our setting, randomness implies that the variations are not systematically related to (unobserved) variables that affect firms' R&D decisions. A potential quasi-experiment could use variations in the generosity of the scheme with respect to firm characteristics that are relatively fixed in the short term, e.g. number of employees. Assuming that firms around the border of the size restriction are comparable, and that it is in a sense random whether they were eligible for support through the scheme or not, this discontinuity creates a quasi-experiment, and one may study the effect by comparing firms just above and just below the threshold.

The absence or rareness of random variation thus makes it very difficult to identify the true, causal effects of tax incentives for R&D. The problems due to non-randomness are most visible in the econometric evaluation approaches, but are also present in other approaches.

5.3 Summing up: Different questions require different methods

As discussed above, there are advantages and disadvantages with respect to all the methodological approaches described in this section. However, it seems clear that the different methods have clear comparative advantages in addressing different issues of an evaluation.

Econometric techniques are well suited to analyse issues such as input and result additionality, although case studies and surveys may be useful supplements. Econometric techniques have the greatest advantage when there is a lot of data available, so the model

to be estimated can be correctly specified. Data collection through existing sources is also relatively cheap. In situations where data on many relevant variables are lacking, and the preferred model cannot be estimated, an interview-based approach may partly substitute for econometric analyses. As described above, in an ideal situation the interviewee implicitly controls also for factors that cannot be included in an econometric model when they give their answers on e.g. input additionality. However, there is no guarantee that the implicit model is the same across interviewees. Data collection through interviews is relatively costly.

While econometric techniques are well suited to capture effects that may be quantified in a sensible way, they are not equally suited to identify behavioural additionality, i.e. changes in the way firms understand R&D and how R&D decisions are made. Data on these issues are seldom readily available from independent sources, and must be collected separately. Coding information on such issues in a way that they can be included in econometric model, often removes important parts of its content. Here, case studies and interviews are more relevant methods.

6 Data issues

Good data are crucial for evaluations. The lack of appropriate data severely limits the potential of an evaluation.

Before starting the design of the evaluation project – and even before implementing the scheme, governments or operating agencies should go carefully through the data requirements for the evaluation and compare them to the data actually available. Some of the key questions to ask are:

- What is the ideal situation?
- What are the possible data sources?
- What are the shortcomings of the actual data situation?
- How can they be remedied - at what cost?

Below, these questions are discussed in turn. Our general principle in the discussion of data issues is that one should, as far as possible, use already existing/available data to avoid extra red tape for companies and to be able to have long time-series of data at a modest cost. In some countries, there are few relevant pre-existing data sources, and the time-series dimension may be short. The advice to such countries is to rely more on data gathered specifically for evaluation purposes, and to start establishing a data infrastructure for future evaluations.

6.1 What is the ideal situation?

6.1.1 Creating an evaluation database

A proper evaluation has to rely on many different data sources, by combining different types of information gathered with different methods. Not all variables will be used in all parts of the evaluation, but many analyses require information from different sources. The ability to link data from different sources is therefore crucial. Consistent use of common identifiers across data sources is very helpful. Using the idea of creating an integrated evaluation database will be helpful as a guiding principle in the process of gathering the data. If possible it would be useful to include data on firms that apply for and receive R&D.

The unit of analysis should be the entity that faces the incentive inherent in the scheme. This will in most cases be the fiscal legal unit, implying that any data on plant or line-of-business level should be aggregated to this level.

6.1.2 Data describing firms

Put simply, the purpose of the evaluation is to find out what happens to firms after they are exposed to an R&D-stimulating scheme. To analyse the effect of the R&D incentive, one needs to control for other variables "known" to affect R&D. In addition, there is often a need for separate analyses of whether effects differ between different types of firms. This implies that one needs detailed information on firm characteristics, both prior to and after the introduction of the scheme.

One set of variables relates to the *effects* of the scheme, both first and higher order effects. The relevant variables will of course depend on what the evaluation specifically focuses on (see below), but key variables are likely to include:

- R&D activity: Own R&D, purchased R&D, R&D personnel, financial sources, partners etc.
- Result variables: Revenue, value added, profits, gross investments, innovations, patents, change in decision processes for R&D etc.

A second set of variables relates to other more *basic characteristics* of the firm, such as size (number of employees), education level of the staff, localization etc.

Note that the classification of variables into one of the two sets is not always obvious.

6.1.3 Data describing the subsidised R&D

In some countries support through the tax system is based on R&D project applications. The application process, the financing and the follow up reports are here important information sources. Key variables are:

- Classification of the project: Whether it be research, development, process or product oriented and what type of technology which is promoted
- Budget: Actual size and level of support, time span
- Cooperating partners: Other firms, research institutes
- Evaluations: "Grading" of projects made by the government agency in the applications processing.
- Self-assessments: Intermediate and final reports.

Firms applying for support through project-based measures, have to submit a substantial amount of information about themselves and their research project. Through the processing of the application at the government agency, a lot of information relevant for the evaluation is produced. Firms may also have to send in their final reports with information of how the project actually was carried out, and with a self-assessment of the results. In addition, the application and reporting procedures represent an opportunity to collect data that are necessary for the evaluation, but not available from other sources. Thus, the administrative routines surrounding the scheme should be designed also with evaluation in mind. This source may provide valuable data at a modest cost. However, one disadvantage is that the data only cover those who apply for support through the scheme.

In other countries tax incentive works on a self-assessment basis and so there is often no information on the R&D other than the amount of qualifying spending. This might not be a major problem for evaluations which have a primary focus on the increase in R&D spending and not on the type of R&D supported.

To study the effects of a tax incentive scheme for R&D, one needs a fairly long time horizon because the total economic effects of R&D, or even the effect on R&D itself, cannot be expected to be visible in the short run. The data should cover a sufficiently

long period both prior to and after the launching of the scheme. What should be regarded as "sufficiently long" can in principle not be established ex ante.

In practice, one has to make a compromise based on the availability of historical data, the cost of gathering new data, and the time horizon for the evaluation project.

6.2 What are the possible data sources?

Evaluations need to combine data from many different sources. Some data sources exist prior to and independent of the evaluation at hand but, in many evaluations, there will be a need for collection of new data.

Below, some of the main types of data sources are described, with their strengths and weaknesses.

6.2.1 Administrative registers

A key administrative source of data is likely to be from the administration of the tax incentive itself. In addition many countries increasingly rely on data from administrative registers for the production of official statistics. A number of countries also make such data available for research and analysis.

Examples of relevant administrative registers are, accounts statistics, tax statistics, business registers and employee registers, and there are several advantages concerning the use of these:

- They generally cover the whole population
- They contain a large set of variables
- The marginal cost for making the data available and useful for research and evaluation is small
- They generally use standardized identifiers, making linking easy

However, register data also have their disadvantages:

- They are generally gathered for other purposes than research and evaluation. Not all variable definitions may be ideal for evaluation purposes
- There may be legal obstacles for using the data, due to confidentiality issues

6.2.2 Regular censuses and surveys carried out for other purposes

As for register data, this type of data is often used for the production of statistics, but is suitable for research. Relevant data sources of this kind include manufacturing censuses, R&D surveys and innovation surveys. The advantages and disadvantages are mainly the same as above, with the exception that the surveys do not cover the whole population.

The representativeness of the sample in surveys may create a challenge. Often, surveys limit the sample to firms of a certain size, e.g. 10 employees or more. If the scheme is targeted towards – or turns out to be more popular among – small firms, then this data source is not ideal in the sense that is does not cover a main user group of the scheme of the scheme.

6.3 How can shortcomings be remedied – at what cost?

After reviewing the possible data sources, in light of the questions to be evaluated, one should "take stock" and see what is missing.

After this evaluation, one must decide how to try to remedy the data deficiencies. As mentioned above, the main ways of getting new data are through the administrative routines surrounding the scheme, and through specialized surveys, etc.

"The perfect" world can seldom be achieved with a limited budget. There is inevitably a trade-off between data quality, coverage and the number of variables.

Clearly, some variables are more important than others in an evaluation. A key question in any evaluation of an R&D-stimulating tax or subsidy scheme is simply whether R&D investment has increased due to the introduction of the scheme. A crucial piece of information for evaluation purposes is the level of R&D at the firm level prior to the launching of the scheme. To what extent is this information available? This question should be asked at a very early stage. In many countries, R&D surveys are the main source of information for this variable. However, R&D surveys often do not cover small firms, and so there is a need to look at other possible data sources.