



Mutual Learning Exercise

Ex-post evaluation of business R&I grant schemes

Horizon 2020 Policy Support Facility



Research and
Innovation

EUROPEAN COMMISSION

Directorate-General for Research & Innovation
Directorate A — Policy Development and Coordination
Unit A4— Analysis and monitoring of national research and innovation policies

Contact (H2020 PSF):

Eva Rückert, responsible Policy Officer - Eva.RUECKERT@ec.europa.eu

Contact (H2020 PSF coordination team):

Román ARJONA, Chief Economist and Head of Unit A4 - Roman.ARJONA-GRACIA@ec.europa.eu

Stéphane VANKALCK, Head of Sector, Unit A4 - Stéphane.VANKALCK@ec.europa.eu

Diana SENCZYSZYN, Team Leader, Unit A4 - Diana.SENCZYSZYN@ec.europa.eu

RTD-PUBLICATIONS@ec.europa.eu

European Commission
B-1049 Brussels

Mutual Learning exercise

Ex-post evaluation of business R&I grant schemes

Horizon 2020 Policy Support Facility

Written by the independent panel of experts:

Paul Cunningham, MIOIR

Pim den Hertog, DIALOGIC

Viola Peter, Technopolis Group

***EUROPE DIRECT is a service to help you find answers
to your questions about the European Union***

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the internet (<http://europa.eu>).

Luxembourg: Publications Office of the European Union, 2017.

ISBN: 978-92-79-65260-8

KI-AX-17-003-EN-N

DOI: 10.2777/28966

© European Union, 2017.

Reproduction is authorised provided the source is acknowledged.

Cover images: © Lonely, # 46246900, 2011. © ag visuell #16440826, 2011. © Sean Gladwell #6018533, 2011. © LwRedStorm, #3348265. 2011. © kras99, #43746830, 2012. Source: Fotolia.com

Table of contents

| | |
|--|----|
| THE PSF MLE PANEL | 4 |
| FOREWORD | 5 |
| EXECUTIVE SUMMARY AND POLICY MESSAGES | 6 |
| 1. INTRODUCTION | 9 |
| 2. SYNTHESIS OF THE GROUP'S WORK | 15 |
| 2.1. Evaluation culture and infrastructure needed | 15 |
| 2.2. Diffusion of evaluation results | 18 |
| 2.3. R&D schemes' evaluations in the wider policy cycle..... | 19 |
| 2.4. Challenges evaluating R&D schemes | 20 |
| 2.5. Methodological aspects | 23 |
| 3. A SELF ASSESSMENT OF THE PROCESS | 25 |
| 4. CONCLUSIONS | 26 |
| APPENDIX A - ADDITIONAL MATERIAL | 28 |
| APPENDIX B - BIBLIOGRAPHY | 30 |
| REPOSITORIES OF EVALUATION STUDIES | 31 |
| APPENDIX C – SEMINAR'S AGENDAS..... | 32 |

The PSF MLE panel

Dr Agneta Bladh is an independent consultant involved in evaluations and other engagements in the field of higher education and research. Dr Bladh was Rector of the University of Kalmar (now part of Linnaeus University), Sweden, between 2004 and 2009. From 1998 to 2004, Dr Bladh served as State Secretary at the Swedish Ministry of Education and Science, responsible for Higher Education and Research. In this role, she was one of the signatories of the Bologna Declaration. From 1995 to 1998, Dr Bladh was Director General at the National Agency for Higher Education. Bladh served 1987-91 as well as 1993-95 in various positions in the Swedish national administration.

Dr. Paul Cunningham is senior research fellow at the University of Manchester, Alliance Manchester Business School. He joined PREST in 1985. Since then his work has covered a wide range of activities and his research interests have diversified greatly from his original interest in new marine technologies. Particular (albeit broad) areas of personal research interest now include: studies of the development and application of evaluation methodologies, and R&D evaluation; investigations of HEI-industry collaboration mechanisms and the role of HEIs in regional economies; studies of quantitative measures of R&D performance and the development of science and technology indicators; national and regional level studies of innovation and the identification of best practice in innovation policy and governance; studies of transnational scientific collaboration; and international science and technology policy development.

Dr Pim den Hertog is founding partner of Dialogic and active in the innovation domain for more than 20 years. Pim's research interests are in (management) of service innovation, open innovation, innovation policy-making (including evaluation) and innovation governance. Pim conducted many research projects, many of them in the role of project manager, for various Dutch Ministries, the European Commission and OECD. Prior to his work at Dialogic, Pim worked at TNO Strategy, Technology and Policy (1990-1998).

Dr Viola Peter is deputy director for the Brussels based office of Technopolis Group. She joined as senior consultant in October 2005. Viola has a background in the field of innovation studies including science and technology policy analysis, indicators, evaluation, knowledge-based technologies and innovation systems, Viola is the head bibliometric expert at Technopolis Group and provides bibliometric analysis to a number of projects. She is interested in indicator development and contributed to relevant studies such as the Public Sector Innovation Scoreboard or the ERA progress monitoring. Prior to joining Technopolis Group, Viola worked at the European Commission's DG-RTD in the area of economic analysis and indicators. Prior to that, she was a researcher at the Fraunhofer Institute for Systems and Innovation Research (FhG-ISI), Karlsruhe, Germany.

Foreword

This report from the Mutual Learning Exercise “Ex-post evaluation of business R&I grant schemes” summarises and analyses the discussions among representatives from five countries during late spring 2016.

This “MLE” has been characterised by engaged discussions among participants. It can therefore be seen as a true bottom-up exercise, as it was also intended. The experts and invited speakers with their deep knowledge and experience from the field contributed in both an enlightening and provocative way and thereby broadened the discussions.

The participants seemed to have found new colleagues to discuss approaches, methods and new tracks for problem solving. The two country seminars in Spain and Denmark were professionally prepared and organised from the respective country and we are grateful to them for that. Hopefully, this report will remind the participants of the MLE they took part in, but also give insights to others.

Thanks to Viola, Pim and Paul for writing this report. It has been a pleasure chairing an MLE with such engagement from participants.

Agneta Bladh

Chair, PSF MLE panel

EXECUTIVE SUMMARY AND POLICY MESSAGES

The purpose of this MLE was “to improve the exchange of information, mutual learning and identification of good practices between the participating countries and in this way contribute to put in place better systems for the ex-post evaluations of business R&I grant schemes”.

Stakeholders from Spain, Denmark, Sweden, Norway and Germany actively participated and shared views and experiences in this MLE as well as some countries with observer status (Turkey, Hungary, Romania). In addition to the first scoping workshop in Brussels (23 February), and, more significantly, the two country workshops or site visits to Spain (Madrid, 26 April) and Denmark (Copenhagen, 3. May) and a concluding workshop in Brussels (30 May), the process benefitted from additional views from mostly local experts and practitioners during the country visits and a rich discussion in all four workshops. These perspectives were complemented by a small survey among the participating countries to solicit input for the concluding Brussels workshop as well as additional views and comments by the two external experts and rapporteur to this MLE.

Below we briefly summarize the main insights gained from the MLE experience on the process of evaluation, generally and more specifically on issues arising in the context of the ex-post evaluation of business R&I grant schemes. For both we also point at challenges that were identified. Some lessons for MLEs as a tool for mutual policy learning and possible follow up action are given as well.

Insights gained on the evaluation process

During the first scoping workshop it became clear that it is not realistic to exchange views and experiences on the evaluation of business R&D grants without having an understanding of the context in which these evaluations take place. From discussing this context in greater detail we learned, *inter alia*, that:

- **Evaluation culture and the maturity of the evaluation process differ inherently between the participating countries.** At one end of the scale, evaluation systems form a well-integrated element of the policy cycle, with a clear division of labour in place and evidence of policy learning, with an understanding of the use (and its associated caveats) of the process of (ex post) evaluation. At the other end of the scale, evaluation is not yet a standard routine, nor is it well integrated into the policy cycle. Similarly, the required information infrastructures and budgets for the support of internal evaluation units and/or external evaluations are available to a limited degree only.
- **There is a high level of variation between the participating countries in the extent to which a division of labour between ministries, dedicated evaluation units within ministries, agencies, academia and consultants exists.** In the relevant ministries of the more ‘advanced’ countries, the role of dedicated evaluation units is relatively more important and there is a lower inclination (or need, possibly due to greater in-house resources for evaluation) to rely heavily on the use of external academic evaluators and consultants. However, the greater ‘professionalisation’ of ministerial evaluation agencies and units does not preclude the use of external evaluation agents, although it is often the case that government officials are better able to act in concert with the latter as ‘informed clients’.
- **There is also much variation between the participating countries regarding the presence of an evaluation community of practice.** In these countries the practice and study of evaluation appears to be well developed amongst the academic community and there seems to be at least some tradition for evaluators with diverse backgrounds to meet and learn from each other and to form professional societies. This community of practice often provides an external resource and input for evaluation advice or for the conduct of commissioned evaluations.
- **Policy-makers and experts are looking for intelligent combinations or mixes of more qualitative and quantitative approaches to evaluation.** There is a drive across all countries to seek evidence on the economic effectiveness and impact of schemes, often led by demands for financial stringency. This has placed an emphasis on the use of quantitative measures, particularly those employing econometric approaches. Such approaches are greatly assisted by the presence of extensive data-infrastructures (including provisions for safeguarding privacy), hence in countries that have access to these assets, there is a greater inclination to strive for the development of increasingly sophisticated econometric analyses to use in evaluations. However, at the same time in these countries there is a recognition that opening up the ‘black box’ of R&D and innovation in firms is a necessary goal in order to really understand how firms benefit from R&D and innovation schemes (and mixes thereof) and how this may affect firm behaviour and thereby to contribute to improved policy learning. Specifically, in the meetings, this

opening up was also referred to as 'understanding the innovation journey', a term that both better appreciates the temporal nature of companies' involvement with a sequence of innovation support instruments and the time it takes for effects and impacts to accrue.

- **The building up of information infrastructures for evaluating R&D and innovation schemes is a demanding yet important task and a prerequisite for more quantitative evaluations.** The participating countries are very different in this respect, which can be a consequence of historical factors quite unrelated to the level of evaluation culture. However, it was also concluded that a well-functioning scheme administration (perhaps including provisions for monitoring arrangements) can be a valuable feature. Having insight at firm level to the actual use of/participation in the various R&D and innovation schemes is also highly useful (i.e. the 'innovation journey', especially if this type of data can be coupled to regular R&D, innovation and business statistics. Again the participating countries are very different in the degree to which these information infrastructures are in place and the ease with which these can be accessed and interrogated for evaluation purposes.
- **Participating countries show a considerable difference in the degree to which evaluations are a logical part of a regular policy cycle** (and hence the degree to which policy learning is institutionalized and the results of evaluations fed back to the policy process). This is linked not only to the formal obligation to evaluate (which can sometimes be counterproductive), but also the way in which the results are communicated to policy-makers, parliament and the wider public. Here there is much scope for mutual learning among the participating countries.

Some of the main challenges identified regarding evaluation in general are:

- How best to combine various evaluation methods (and have the information infrastructure in order to do so);
- Ensuring that the results of evaluations are better fed back into the policy process;
- Improving the overall rationale or intervention logic of many R&D and innovation schemes (as these are not always clear when evaluations are commissioned and hence hamper a proper evaluation);
- How to fully understand why policies and schemes work (or not), including: the degree to which effects can be attributed to specific scheme characteristics; the degree to which effects can be attributed to individual schemes when they are used in combination; and what is the persistence of effects by the (semi) permanent or repetitive use of schemes.

Insights gained on the evaluation of R&D grants/schemes

The MLE has given extensive attention to the pros and cons of (and limitations and requirements when) performing advanced econometric-based analyses as part of the wider trend towards evidence-based policy-making. It discussed various examples of what some see as the "golden standard" in evaluation. Key observations are the following:

- **Evaluations using econometric analyses are far from standardized and quite complex type of analyses to perform.** Their results are quite difficult to convey to a non-specialist audience and vulnerable to discussions among experts on what methodology is most appropriate to use for what type of dataset. Finding the right control groups and preventing selection bias are key issues here and differences in choices made may affect the outcome of this type of analyses. Moreover, decisions on methodological approaches and techniques are inherently difficult to explain to a non-technical audience that can adversely affect the credibility of the results.
- **Econometric analyses are very demanding in terms of data-availability and quality.** Hence, countries that have good administrative business data and can couple this to R&D and innovation and more regular business statistics are much better positioned to be able to employ this type of analysis. However, even in those countries it is rare to find schemes where high quality data are available in abundance and where this type of analysis is possible. Moreover, there may be a risk in becoming overly reliant on econometric type evaluations, particularly where schemes may be too small to exhibit detectable impacts or data-availability is low.
- **The working procedures regarding access to data and data confidentiality are far from solved in only but a few cases.** Here both politicians and statistical bureaux could step in to increase the use of data that are in many cases available, but underutilized.
- **There is a trend toward this type of econometric analysis, but this needs to be balanced by recognising the simultaneous need to understand behavioural effects**

(i.e. the “innovation journey of firms”) of the use of R&D and innovation grants more fully. Therefore, when using this type of analysis there remains the need to open up the ‘black box’ of how firms decide on, benefit from and most importantly change their R&D behaviour (or not) following the use of these R&D grants. A development towards using a mixed approach to evaluation where this type of econometric analysis is used in combination with more qualitative evaluation methods is therefore preferred.

Some of the main challenges discussed include how to measure (and understand) behavioural effects, learning from the use of econometric type analyses as applied in other domains (such as education and labour market studies) and how to deal with the interaction of various schemes used in combination. Also more information was required on the way in which firms used support schemes, for example through being ‘repeat users’ of the same scheme, or using multiple schemes in combination or in progressing from one type of scheme to another. Finally, there seemed to be potential in exploring the use of ‘big data’ to analyse firms’ behaviour and their response to schemes.

Lessons learned from the MLE-process

Participating countries were *positive about the learning process* they experienced and indicated that a *MLE/workshop size of between 4-8 countries is best*. The *country visits* in particular were identified as very helpful exercises in terms of learning from other practices (bench-learning) and reflecting on one’s own practices. Too much heterogeneity (in terms of the experience of the participating countries) was viewed as a less attractive factor (and here peer review maybe a better instrument).

A number of points for improvement include:

- pinpoint the scope of the whole exercise early on;
- more upstream information for participants;
- more time between subsequent country visits;
- reduce the scope for those countries organising workshops to benefit fully from the learning process;
- provide a standardised format for a country visit/workshop;
- exchange of more practical aspects concerning evaluation processes.

However, possibly the best indication of the usefulness of this particular MLE experience is the *willingness of the participating countries to maintain the existence of this small learning community* and the idea to *involve each other during actual future evaluations* to encourage exchange and learning on the job.

1. INTRODUCTION

The purpose of this MLE is to improve the exchange of information, mutual learning and identification of good practices between the participating countries and in this way contributing to put in place better systems for the ex-post evaluations of business R&I grant schemes. It should be noted that the content of the MLE workshops was defined very much in a bottom-up fashion through the expectations and experience of the participating countries. The MLE is very much a process oriented approach and hence this report does not in any way set out to provide a prescriptive or generalised account of the best way in which the evaluation of R&I grant schemes may be conducted. Instead, it offers an insight into the challenges faced by the participating member states when considering such evaluations and is supported by some illustrative examples of the methodologies used. These are not presented necessarily as best practice examples, since these approaches are very much conditioned by the country-specific contexts in which they operate, but served as the basis for discussion, reflection and mutual learning.

Policy measures which provide direct support for R&D within companies is possibly one of the oldest and most established policy instruments, dating from the immediate post-Second World War period. Over time, the focus of these instruments has shifted from the direct support of single R&D projects within large individual firms, towards a focus on direct support to SMEs. This shift is largely based on arguments over the comparative efficiency of financing R&D activities in smaller companies, thereby giving access to an increased range of clients although at the potential cost of the size of spillovers that may be obtained from the support of larger firms. In parallel, there has been a growth of more generic support that aims to stimulate firm R&D expenditures, such as the use of R&D tax credits, for example. Arguments over the relative efficacy of these two types of scheme centre on the comparative ease with which the latter may be implemented and administered, versus the strategic opportunities offered by the former (for example in targeting specific growth sectors, or in supporting firms in disadvantaged regions or those threatened by the loss of credit liquidity resulting from the 2008 financial crisis) (Cunningham, et al. 2012).

The provision of direct support for R&D is founded on the rationale that R&D conducted within firms will, directly or indirectly stimulate innovation leading to the production of new marketable products, processes or services – a view that is predicated on the linear model of innovation and which explains the long history of this type of measure. This form of support follows the classical economic rationale for public intervention, linked to the capacity of firms to appropriate investments made and the relative importance of spillovers associated with their R&D efforts. They are thus intended to compensate for firms' propensity to under invest.

Although they offer relative simplicity in comparison with other innovation support schemes, the evaluation of direct measures also poses a number of particular problems. These include:

- The desired effects of the measure tend to emerge at various times, posing the question of when is the optimum time to conduct an evaluation, and how frequently. For example, issues concerning uptake and management will emerge soon after implementation, while, it may take months or years until prototypes are generated or new products, processes or services introduced to the market. Likewise, organisational and behavioural changes will take time to generate and become embedded, and their sustainability, along with that of other desired effects, will require even longer time frames.
- R&D expenditure, growth, profitability and employment, along with many other anticipated impacts of direct support measures are readily measurable and can lend themselves to the construction of easily obtained quantitative indicators. It is less easy to capture information on less tangible outcomes such as skills, innovation capabilities and capacities, and spillover effects, etc.
- As with many other forms of policy intervention, the types of outcome and impact arising is difficult to ascertain in the absence of counterfactual examples or benchmarks established prior to the establishment of the funding.
- Lastly, and this is particularly the case as the size of the target firm increases, the direct outcomes of public support may be difficult to distinguish from other forms of support (Cunningham, et al. 2012).

Furthermore, a range of data collection approaches and methodologies are available for the evaluation of the effects of innovation support instruments and which are not restricted to the assessment of those concerned with the provision of direct support to R&D. Those methodologies that are typically encountered include:

- Existing databases and monitoring data
- Surveys
- Interviews
- Focus groups/workshops/meetings
- Peer reviews (including stakeholder reviews)
- Formalised data on intellectual property (patents, etc)
- Publications data

Likewise, a range of data analysis techniques may be applied to the information collected. Those most frequently encountered include:

- Case study analysis
- Network analysis
- Econometric analysis
- Descriptive statistics (e.g. uptake analysis)
- Input/output, cost/benefit, return-on-investment analysis
- Intellectual property (IP) data analysis (including technometrics and other analyses)
- Publications data analysis (including bibliometrics)
- Qualitative or quantitative analysis of texts

The choice of data collection and analysis methods is largely determined by a number of factors, first and foremost of which is linked to the purpose of the evaluation. In very simplistic terms, if the purpose of the evaluation is to gain an insight into the implementation and management of the programme, then surveys and interviews of participants can be used to gauge stakeholder satisfaction and feedback, allowing policy makers to redesign relevant aspects of the programme's administration. If, however, policymakers wish to gain an insight on the economic effects and impacts of the programme, then surveys which seek information on participant performance can be carried out and subjected to econometric analysis. In broad terms, the choice of approach is dictated by whether the evaluation is designed to investigate the more qualitative process-oriented aspects of the programme ('how?' and 'why?' the intervention performs) or the more quantitative outcomes and impacts ('what?' and 'how much?' does the intervention deliver). The difference in methodologies between these 'soft' and 'hard' approaches formed the topic of some discussion in this MLE and is detailed later in this report.

Another factor influencing the choice of methods relates to timing: if an evaluation is carried out too early, many of the impacts and effects will have had insufficient time to manifest. Thus, if the evaluation is to adopt a 'hard' approach, it must be delayed until a sufficient body of relevant information has been accumulated.

A further factor is the type of policy intervention that is to be examined. It is beyond the remit of this report to explore this aspect in detail, but a good overview of a range of innovation support measures and the most appropriate evaluation methodologies that can be employed is provided in Technopolis and MIOIR (2012). Taking the specific case of direct support to R&D conducted by firms, this report suggests the following menu of evaluation approaches:

- "Use of structural business statistics or data from tax authorities (company accounts, etc.) to compare beneficiary performance over time with those of a comparison group of non-assisted enterprises. Such data can also be used to assess multiplier effects of the public funding (gross-value added, etc.).
- Bespoke surveys of beneficiaries (either a sample or the entire population) extended to a comparison group (or double comparison group, i.e. non-applicant and unsuccessful applicants) to allow for a counter-factual analysis.
- Counterfactual econometric analysis of micro-data (e.g. the Community Innovation Survey data from Eurostat) or national/regional panel data (e.g. the Mannheim Innovation Panel in Germany). An econometric analysis is generally only appropriate when an evaluation covers a large enough number of beneficiary firms for which a reliable and complete economic time series can be obtained
- In-depth analysis of a sub-set of beneficiary firms applying interview/case study methods to understand the synergies between innovation measures. This can be done by tracking over time companies that have received a 'package' of support (e.g. innovation voucher, grant for R&D, prototyping and follow on investment, training and export grants). By

covering the full 'project life cycle' from the firm's viewpoint, the evaluation avoids the risk of project fallacy (assuming that a grant, which may only cover part of the product development phase, leads to a direct and verifiable outcome for the firm)". (Technopolis and MIOIR, 2012).

In their review of the impact of direct innovation support policy measures, Cunningham et al (2013) encountered evidence of a range of evaluation methodologies which closely concur with those identified by Technopolis and MIOIR (2012). They note that these methods are generally employed to examine programme issues such as rationales, user characteristics, governance aspects, input additionality, output additionality and behavioural additionality effects. However, there is a significant difference between academic studies in this particular area, many of which tend to focus on the issue of input additionality and, to a lesser extent, output additionality, while (commissioned) policy evaluations tend to focus on the continued relevance of the rationale for the intervention and on its implementation performance.

A particular focus of this MLE was the use of sophisticated, highly quantitative approaches in order to determine programme impact. These approaches are typically based on advanced econometric methods. As noted by Technopolis and MIOIR "the application of econometric methods to look at how the performance of recipients of funding compares to a counterfactual situation is a technique that is still used correctly in only a handful of evaluations. The advantage of a counterfactual approach is the (potential) rigour and accuracy of the results". This statement may explain why many econometric studies are conducted within the academic sector rather than the policy sector as noted above, although it is noted that many academic studies do not involve the application of econometric techniques to specific support programmes but generally tend to utilise existing (general) data sets such as those provided through the Community Innovation Surveys, for example.

The authors note, however, that the use of such methods are strongly dependent on the availability of sufficiently robust and complete (time series) data on the use of different sources of funding by a (statistically viable) sample or panel of participant and non-participant firms. For this reason, evaluations of this type often combine data obtained from a survey of beneficiaries and compare it with the 'average' performance of firms in their sector, etc. using standard economic statistics. The availability of suitable and robust comparison data thus becomes an absolute prerequisite for econometric studies: such data is rarely readily available to policy makers and evaluators, however. A further practical tip offered by the authors when using econometric counter-factual approaches was to ensure that data is gathered from a sufficiently large group of beneficiary firms and to check for the combined effects of different types of funding.

A further caveat that may be applied to the use of econometric modelling approaches in evaluation is that, even if designed and executed correctly, the interpretation of results of the analysis may require a very high order of technical expertise, which can limit their use in evaluation practice and subsequent policy-making (Technopolis and MIOIR, 2012).

While Cunningham, et al (2013) do not make any specific comparisons of the utility or appropriateness of the range of available evaluation methodologies, they note that the findings of several econometric surveys exhibit great variability in terms of their results: even when applied to the same or similar datasets, the selection of modelling approaches and parameters can deliver quite different outcomes. Clearly, such variation raises questions about the policy utility of these approaches.

It is clear (not least from the focus of the MLE workshops) that policy attention on the application of econometric techniques in policy evaluation is increasing, although progress in their development has arguably been limited. A review of innovation policy evaluation practice conducted by the OECD some ten years ago noted that "some of the most interesting recent developments in evaluation methodology concern the use of econometric techniques based on longitudinal micro-level data...The quality of results based on this approach is, however, conditional on the extent to which researchers can control for firm characteristics other than programme participation. Furthermore, this technique is only the first step in a full cost-benefit analysis: at its best it establishes the private benefits conferred to firms as a result of the programme; justification of a programme needs to account for social benefits against the total costs" (Papaconstantinou and Polt, 2007). Such caveats appear as relevant today as then.

A more focused examination of the utility of econometric approaches to impact assessment is made in a sister paper (Capron and van Pottelsberghe de la Potterie, 2007). The authors present a rather mixed message noting that, while extensively used as decision-making tools in economic policy, the credibility and usefulness of econometric methods as a tool in the field of S&T policy is controversial, if not contested. They note that quantitative economic methods can provide answers to questions relating to the efficiency of S&T policy, but that their use depends on the specific

question being addressed and that the results of such studies should be viewed as “providing an additional piece of information in the evaluation process”.

It is not within the scope of this report to detail the various econometric modelling approaches that may be employed, a discussion of the more salient considerations of their use in policy evaluation is provided by Athey and Imbens (2016). One interesting observation made in this paper is the statement that “the gold standard for drawing inferences about the effect of a policy is the randomized controlled experiment; with data from a randomized experiment, by construction those units who were exposed to the policy are the same, in expectation, as those who were not”. This approach makes it relatively straightforward to draw inferences about the causal effect of a policy and the difference between the sample average outcome for treated units and control units. It thus forms an unbiased estimate of the average causal effect. However, as the authors add, the luxury of being able to test policies in an experimental fashion is outside the financial resources of most policymakers, not to mention the inherent political risk entailed. Hence, evaluators must deal with the problem of creating robust comparison datasets, together with dealing with the assumptions these rely upon.

Many of the conclusions by the OECD Expert Group on the Evaluation of Industrial Policy (see Warwick and Nolan, 2014) were to a large extent echoed in discussions and presentations during the various workshops and in particular the two country seminars. The expert group has considered recent evidence from the evaluation of industrial policy, against the background of new approaches to evaluation. Industrial policy is of course much broader than business R&I grant schemes and ranges from R&D support policies, innovation-oriented public procurement, capital market interventions, sectoral approaches (including public-private partnerships), cluster and regional policies to national industrial strategies. Nevertheless some observations are worth mentioning here as many of them were discussed in similar terms in this particular MLE. One of the main conclusions of the expert group was that “rigorous and systematic evaluation evidence in many areas of industrial policy is scarce, and methodological challenges particular to this field are many” (p. 4). A series of causes was suggested for the relative limited use of especially experimental methods and rigorous control groups in the evaluation of industrial policy in comparison to other policy areas (p, namely:

- The feasibility of identifying control groups can vary from one instrument to another.
- Data availability. Researchers still complain about difficulties in accessing relevant data in the industrial and innovation policy sphere.
- Unit of analysis – businesses are large units of analysis making assessments of the effects of interventions on firms difficult to capture when looking at the impact on the business as a whole. Interdependence between the outcomes for economic agents.
- Multiple declared objectives which are hard to evaluate in combination.
- Time lags and long-run impact. The first may limit the usefulness of evaluation as a tool for policy development. The second refers to gains which are important in the long term but not easily quantified over the short-term.
- Context dependence refers to the idea that it may be hard to translate the results of one policy experiment to a different context.

It was further concluded that (p. 63) “while there are a few exceptions, there is a need for most OECD countries to increase the quantity and quality of industrial policy evaluation. Where appropriate, evaluators of industrial policies should make more use of evaluation tools used routinely in other spheres, including randomised control trials, quasi-experimental methods and well-grounded counterfactuals. For more complex industrial policy interventions, the challenges are greater. There may still be some scope for the use of experimental methods and for an iterative approach to evaluation, but there also needs to be realism about what can be achieved and about the difficulties of establishing counterfactuals and undertaking impact evaluations. Use of structured hypothesis testing, the intelligent use of carefully selected indicators and creative attempts to extrapolate from diverse sources of evidence might be the best that can be achieved.”

Some of the principles suggested as the basis of good practice industrial policy evaluation included the following (see p. 63-64):

1. Make explicit, at the highest possible level, the commitment to evaluation of industrial policy.
2. Consider mandating evaluations when public funding is provided.
3. Insist on the development of data and evaluation strategies as a pre-requisite for the commencement of programmes.

4. Choose the evaluation technique in the light of the size and nature of the programme concerned.
5. Evaluating industrial policy requires an eclectic approach mix of methods.¹
6. In the face of complexity, or when outcomes are uncertain, consider the approach of developmental evaluation.
7. Insist on full disclosure in evaluation reports.
8. Robust governance mechanisms are needed to ensure evaluation is objective and free of political influence.
9. Good mechanisms for policy learning are needed to ensure that the findings of evaluation feed back into future policy making.

Against this background, and in order to gain a picture of the context within which the stakeholders were operating in their home countries, a number of potential relevant and related questions were collected from the participants during the opening workshop in Brussels (see box below).

Issues regarding evaluation of business R&I grant schemes raised during scoping workshop

Organisation

- Organisation of the Impact Assessment unit (resources, staff,...)
- Cooperation with external experts and internal staff
- Practical issues of (fixed) contracted, external evaluations

Design of the instrument

- Setting realistic goals?
- Improve/change outputs, inputs, behaviour?
- Grants or loans?
- Targeting or spreading the budget thinly?
- Societal impact?

Evaluation process/method

- Is there an intervention logic?
- Are unintended effects included in the design?
- Measure at firm level or project level?
- When to measure results (timing)?
- Scope for econometric evaluation?
- Qualitative methods to be used (integrated and systematic surveys; customer surveys; interviews, case studies, ...)?
- How to combine and integrate qualitative and quantitative analyses?

Factors enhancing complexity

- Systemic interventions 'owned' by multiple intervention 'owners'?
- Attribution of impacts in the case of e.g. repeated grants, multi-goal systemic instruments and existing complementary instruments?

Diffusion/ feedback

- How to diffuse evaluation results into new or existing instruments?
- How to built-in policy learning?

¹ Here we cite the more detailed conclusion as they are very much in line with the observations made in this particular MLE, namely: "conclusion aptly described". At the level of industrial strategy, a mix of evaluation methods is likely to be needed – while state-of-the-art econometric methods may have their role in assessing components of the programme, but are less likely to be useful for the policy package as a whole. Tracking of macro or meso-level indicators, international benchmarking, subjective assessments via survey methods, narrative reporting, case studies and other techniques all have a role to play. More generally, there is merit in using a mix of methods and a need for triangulation across the results, which may sometimes conflict" (OECD, 2014, p. 64).

Following the discussions and a mapping exercise, the key issues emerged; they were then taken up during the two country seminars planned under this MLE. In Madrid, the focus was intended to be on qualitative issues while in Copenhagen, concrete examples on quantitative issues were to be discussed.

Organisation of the process

The process started with identifying the key issues, the participants wanted to learn about. Once these were identified and 'boxed' it was decided to organise the country visits as a sort of learning seminar. There were then two country learning seminars, one in Madrid, which was followed by one in Copenhagen.

The participating countries and their corresponding organisations were Spain (CDTI), Denmark (DASTI), Norway (Innovation Norway), Sweden (Vinnova), and Germany (VDI/VDE). Observer countries that joined for the opening workshop were Hungary, Turkey, Romania and Czech Republic. The Hungarian observer also joined the Madrid workshop. The group was chaired by Agneta Bladh and supported by the three experts Paul Cunningham, Pim den Hertog, and Viola Peter (rapporteur).

The country seminar

How was the country seminar organised and how were they run? What are typical topics for discussion and mutual learning?

The MLE came up with two broad topics – the trend of quantitative methodologies on the one hand, and the need for qualitative approaches on the other hand. These topics seemed to be fit for being the focus of one of the two planned country seminars each.

Once Spain and Copenhagen had volunteered to host the seminars, the participants went back home and discussed within their organisation, what would be useful to address and which external experts could be invited to provide an intervention. The MLE experts made also suggestions when appropriate.

The ideas were then discussed with the expert team by mail or phone and a draft agenda prepared. In the meantime, the country hosts spoke to potential external experts and invited them. The agenda was finalised once the interventions were fixed.

For both country seminars, the MLE participants, local invited experts as well as colleagues from the organisation spent either a few hours or the whole day to listen to the presentations and participate in lively discussions.

In particular the hands-on examples of evaluations, the difficulties the evaluators/researchers face and strategies to overcome these barriers were discussed. Examples of the topics included:

- obtaining the right sample or the right counterfactual group,
- introduction of sometimes very specific databases,
- general problems policy makers face such as
 - self-selection of participants,
 - the absence of a sound understanding of behavioural effects and
 - poor measurement of additionality or crowding out effects (in terms of multiple scheme participation).

While the problems seemed to be manifold, it was clear from the discussions that they are widely shared among the MLE participants. Since they are not shared evenly, the participants benefitted by gaining an understanding of how participants treat issues differently.

The learning seminars were attended by the MLE participants, EC colleagues, as well as internal and external experts – private sector as well as academics. The agendas of the seminars are annexed to this report.

A final workshop in Brussels in May was used to discuss the exercise and the outcomes.

2. SYNTHESIS OF THE GROUP'S WORK

While there were very clear and focused aspects dealing with issues on the evaluation of business R&D schemes, the discussion and identification of key issues suggested that a broader perspective is useful. Hence, where relevant or where contrasting examples were presented, the discussions ranged into the broader evaluation context and did not restrict itself to the narrower issue of R&D scheme evaluation. For reasons of clarity, we have not separated the following synthesis into two sets of issues, focusing on the narrow and the broader aspects of evaluation respectively but have dealt with them collectively. However, such a separation of issues has been attempted in the Executive Summary. Below we synthesize the group's work under three headings i.e.: evaluation culture and infrastructure needed; diffusion of results and R&D schemes' evaluations in the wider policy cycle.

2.1. Evaluation culture and infrastructure needed

Whether an evaluation takes place, its set up and role in policy making depends by and large on the existing evaluation culture within a country, ministry or other type of organisation respectively and also the tendency to use different approaches (based on existing resources – such as evaluation staff, in-house evaluation expertise or information/databases – both pre-existing or newly constructed). **Evaluation culture and the maturity of the evaluation process differ inherently between the participating countries.** At one end of the scale, evaluation systems form a well-integrated element of the policy cycle, with a clear division of labour in place and evidence of policy learning, with an understanding of the use (and its associated caveats) of the process of (ex post) evaluation. At the other end of the scale, evaluation is not yet a standard routine, nor is it well integrated into the policy cycle. Similarly, the required information infrastructures and budgets for the support of internal evaluation units and/or external evaluations are available to a limited degree only.

Creating an evaluation culture or making evaluations compulsory?

How could we increase the evaluation culture in the public administration and how to increase the innovation culture? How can legislation lead or help? While some participants were not in favour of imposing evaluations, the necessary change in mindsets is often impossible without factual pressure such as legislation. Pressure through accountability is needed otherwise 'culture' is very often used as an excuse for not doing evaluations. However, changing culture and mindsets is rather difficult in a more decentralised system compared to centralised ones.

Norway indicated that the country has a legal framework for evaluations (Financial Management Regulations for the government, Article 16.). In general there are few references to it since evaluations seem to be an integral part of the policy cycle. However, the legal obligation is a helpful backdrop.

A further caveat is that any 'compulsory' evaluation processes must be used for the true purpose of policy learning (with the appropriate feedback loops in place) and should not just become an administrative 'tick box' exercise: this would both waste personnel resources and fail to deliver useable policy lessons.

It is insufficient to impose evaluations but a country also needs evaluation experts which have the skills to learn and implement good programmes.

There is variation between the participating countries regarding the presence of an evaluation community of practice. In these countries the practice and study of evaluation appears to be well developed amongst the academic community and there seems to be at least some tradition for evaluators with diverse backgrounds to meet and learn from each other and to form professional societies. This community of practice often provides an external resource and input for evaluation advice or for the conduct of commissioned evaluations.

Example 1 - Evaluation of The Danish Innovation Voucher Programme (Denmark)

Programme - Support programme for SMEs and RTOs to use max. €15.000 for scientific

consulting or access to equipment.

Evaluation Methodology – Econometric analysis, random selection of grant recipients (= experimental model)

Causal inference - Since participants are non-randomly selected into the programme, simple comparison of participants and non-participants leads to a selection bias. In the 2009 selection round, a random allocation ('lottery') was used which was the basis of the evaluation.

Data basis: Data on lottery winners and losers was merged with financial data for the period 2007-2012, thus allowing for analysing pre-treatment and post-treatment effects.

What proved beneficial...

- Linking additional financial data;
- Analysis excluding selection basis

Findings: Funded participants are 20% more productive than non-funded ones; programme affects mainly micro-firms.

Source: Christensen, T.; Kuhn, J., Schneider, C., Sorensen, A.: Science and productivity – evidence from a randomized natural experiment. *SIMPATIC final conference, Brussels, February. 2015*

There is a high level of variation between the participating countries in the extent to which a division of labour exists between ministries, dedicated evaluation units within ministries, agencies, academia and consultants. In the relevant ministries of the more evaluation-prone countries, the role of dedicated evaluation units is relatively more important and there is a lower inclination (or need, possibly due to greater in-house resources for evaluation) to rely heavily on the use of external academic evaluators and consultants. However, the greater 'professionalisation' of ministerial evaluation agencies and units does not preclude the use of external evaluation agents, although it is often the case that government officials are better able to act in concert with the latter as 'informed clients'.

Evaluation, as such, is thus organised mainly at organisational level. This can be a department within a ministry, a funding or evaluation body, etc. At this level, the organisation of the evaluation from its planning to implementation and diffusion of results seems to differ less than one might expect.

If we want to know about the **purpose** of the evaluations, 'policy learning for future programme design' and 'policy learning for broader policy development' are key aspects. Evaluations are often but not always a legal requirement and at least in the participating countries, they are rarely conducted for financial accountability reasons. Having an evaluation can, however, be an asset when it comes to inter-ministerial competition of resources: here, it is much harder to attract funding if there is no positive evidence of effects.

As mentioned, evaluations can be done in-house by the implementing organisation, externally by a third party, or a mix thereof. For monitoring purposes but also as a potential input to the ex-post evaluation, the implementing organisation tends to collect the required monitoring data – either from the application phase or during the implementation of the scheme. Monitoring tends to be done in-house while the ex-post evaluation is quite often executed in collaboration with academics or consultancies. The use of external consultants has at least two rationales:

- internal capacity is limited to do it all in-house,
- use of external bodies adds credibility and independence.

In a number of ex-post evaluations, the internal staff resources collaborate with the external evaluators. For example, the commissioned analyses are followed closely and the organisation offers analytical advice and comments. However, collaboration with external experts – whether they are from academia or consultancies - has its pros and cons as included in Table 1.

Table 1 Pros and cons for external evaluations

| Pros | Cons |
|--|--|
| Possibility to provide internal in-depth knowledge and understanding of the evaluated instrument | Less knowledge on thematic fields of funding |
| Proper exploitation of the available internal data | Less knowledge on programme administrative processes |
| Get access to expert forums and methodological knowhow. | Possible loss of internal competences |
| Independence | Requires substantial coordination or even control |
| Adds credibility | Additional costs |
| Frees internal resources | Requires informed customer within department |
| Adds a fresh perspective | |

One of the key questions of the ex-post evaluations of business R&D support schemes is how the evaluation questions are addressed. Do organisations prefer qualitative or quantitative evaluation approaches?

The building up of **information infrastructures for evaluating R&D and innovation schemes** is a demanding yet important task and a prerequisite for more quantitative evaluations. The participating countries are very different in this respect, which can be a consequence of historical factors quite unrelated to the level of evaluation culture. However, it was also concluded that a well-functioning scheme administration (perhaps including provisions for monitoring arrangements) can be a valuable feature. Having insight at firm level to the actual use of/participation in the various R&D and innovation schemes is also highly useful, especially if this type of data can be coupled to regular R&D, innovation and business statistics. Again the participating countries are very different in the degree to which these information infrastructures are in place and the ease with which these can be accessed and interrogated for evaluation purposes. In this respect, participants indicated that limited knowledge about sophisticated econometric analysis within the organisations is a reason why many evaluating agencies favour qualitative methods. The use of quantitative methods is however not only limited to capacity issues at organisational level; there are also a number of pragmatic reasons for the choices:

- **Size of the programme** - If a programme is very small (i.e., having only a few beneficiaries), then quantitative analysis does not make much sense. Whether or not to use quantitative approaches depends also on the period where the programme or scheme is in its programme 'life-cycle'. In an early phase it is often not feasible to use more quantitative approaches since much of the qualitative results data will not have emerged. Thus in earlier life-cycle phases the focus would be on qualitative and performance aspects (i.e. comparative uptake analysis, scheme administration, participant satisfaction). When it comes to impact studies, there is a drive to use quantitative analyses more often.
- **Evaluation goals.** A stronger focus on input-output-impact may more often use quantitative methodologies while a focus on behaviour and intermediate outcomes requires qualitative information (since the latter are 'soft' outcomes and effects). Nevertheless, the use of qualitative information may be able to provide clearer insights into the 'innovation journey'.
- **Type of programme.** Those programmes that aim at inducing behavioural changes in the participating entities require more qualitative approaches in order to be understood properly. While they may have a financial impact focus, the desire to achieve behavioural effects conditions the evaluation methodologies required.

Participants also mentioned **specific constraints** to the use of quantitative approaches, namely:

- limited data availability,
- strict confidentiality restrictions, and
- difficulties to identify appropriate control group(s).

A good and policy relevant evaluation needs to provide answers to 'why-questions', not only 'If-questions'. The former need to be almost entirely based on qualitative information. Quantitative data are necessary in order to estimate the importance or size of the intervention and are therefore equally necessary. In order to have an evaluation that could generate policy learning, a mix is almost always necessary. However, depending on the precise methodology, purely quantitative data may serve to identify broad effects from a (statistically) significant sub-sample which tends to mask the fact that a few outliers perform very well while a long tail of the sample perform sub-optimally.

2.2. Diffusion of evaluation results

In a study conducted for the European Commission, Cunningham and Gok (2010) identified the main issues that affect the degree to which evaluations provide usefulness and utility to policy makers. These primarily concern the information sought by the evaluation, namely on effectiveness of design, management, implementation and of the evaluation itself, achievement of objectives and the broader impacts of the instrument.

This set of issues is particularly relevant to evaluations of policy interventions that have, typically, been commissioned by those directly concerned with sponsoring or implementing the policy and thus have a narrow set of concerned stakeholders. However, it may be the case that the set of stakeholders will be much broader and, at a minimum will include other national government agencies, government sponsors (finance ministries, etc.), parliament, programme participants as well as potential participants, other interested parties (including policy academics and policymakers in other countries), and the broader public.

Each stakeholder group will have specific interests and needs concerning the outcome of any evaluation. Thus, funding agencies may wish to learn about the efficiency of the policy intervention to ensure that tax payers' money is being used optimally, whilst higher level policy makers will be interested in the broader policy effects. Positive outcomes may be of interest to other potential participants who may be considering applying to the programme and there is a general accountability to the tax payer to justify the expenditure of public monies.

This varied set of stakeholder concerns will probably necessitate some tailoring of the evaluation outputs to the different needs of the evaluation audiences. However, if a logical framework or similar objectives hierarchy has been defined at a preliminary stage of the evaluation design, this can be used to distinguish between the various audience requirements since the policy objectives will reflect the information needs of the different stakeholders.

For these reasons, it can make sense to utilise a range of communication strategies for the different stakeholders concerned: not all stakeholders will be interested in receiving the full evaluation report (although all should be provided with access to it). Thus, it is suggested that specific parts of the report, relevant to the information requirements of different stakeholders, can be produced as stand-alone documents.

It should be noted that the study referred to was not able to determine any positive link between the dissemination of an evaluation and its overall utility. However, since evaluation is (or should be) primarily undertaken as a policy learning exercise, the more widely the results of an evaluation are shared, the greater the potential learning impact.

According to the MLE participants, evaluation results tend to be diffused. The participants stated that the results of evaluations tend to be made publicly available on the web. That does not mean that all reports as such are made public - sometimes there is a shorter 'public' version. The form of the report varies and may exist as 'working papers', 'formal journal publication', or the rather 'raw' report without any conclusions or recommendations. The results are often in the national language only, although an English abstract may frequently be provided. In some cases, the organisation publishes it on its own website or there may be specific evaluation portals².

Beside the general public, evaluation results are also diffused to other departments within the organisation or disseminated to policy makers. For funding schemes that are based on a specific parliamentary decision, the evaluation reports are submitted to parliament as well.

² Norway for example has a portal with <http://evalueringsportalen.no>, Austria has one which also serves as repository of any evaluation with www.fteval.at

Results are also disseminated among potential recipients of R&D aids. Positive evaluation results, especially in the form of 'success case studies', are seen as an instrument for promoting further actions.

Example 2 Evaluation of the 'Win Nu' programme (Sweden)

Programme: Awards up to ~€32.000 to 24 new ventures annually. Recipient companies must have a proof of concept and be less than one year of age. Winning provides the companies with media attention, and financial resources while the subsidy works in a similar way to a certificate.

Evaluation methodology: Econometric study focusing on outcome additionality. Two groups were evaluated: selected ones and a control group of (final stage) non-selected ones. This approach overcame selection and endogeneity biases and thus resembled a randomized controlled experiment. Additional company (register) data was used. Ordinary least square (OLS) and propensity score matching were applied.

What proved beneficial...

- Available register data to link to survey data

Findings: Causal structure of subsidies influencing access to finance and human resources, additionality effect much higher than the small subsidy suggests. "The way the subsidy is designed, how attractive and competitive it is and how prestigious it is perceived is likely to be important for the extent to which it serves as a certification of the quality of new ventures" (p. 1510)

Source: Söderblom A., et al. (2015) : Inside the black box of outcome additionality: Effects of early-stage government subsidies on resource accumulation and new venture performance.³

2.3. R&D schemes' evaluations in the wider policy cycle

Participating countries show a considerable difference in the degree to which evaluations are a logical part of a regular policy cycle (and hence the degree to which policy learning is institutionalized and the results of evaluations fed back to the policy process). This is linked not only to the formal obligation to evaluate (which can sometimes be counterproductive – see above), but also the way in which the results are communicated to policy-makers, parliament and the wider public. Here there is much scope for mutual learning among the participating countries.

Policy-makers and experts are looking for intelligent combinations or **mixes of more qualitative and quantitative approaches to evaluation**. There is a drive across all countries to seek evidence on the economic effectiveness and impact of schemes, often led by demands for financial stringency. This has placed an emphasis on the use of quantitative measures, particularly those employing econometric approaches. Such approaches can be greatly assisted by the presence of extensive data-infrastructures (including provisions for safeguarding privacy), hence in countries that have access to these assets, there is a greater inclination to strive for the development of increasingly sophisticated econometric analyses to use in evaluations. However, the mere presence of data infrastructures may not necessarily be structured in such a way, or contain appropriate data formats, that makes them accessible to quantitative approaches. Frequently, proxy data may be used or the information may require extensive cleaning. As in bibliometric data, the original purpose for which the data is collected (to bring it into the wider research domain for the elucidation of other researchers) may not be directly appropriate to the purposes for which it is used in evaluations. This can also introduce unintended and perverse practices into the production and collection of data. In addition, in these countries there is a recognition that opening up the 'black box' of R&D and innovation in firms is a necessary goal in order to really understand how firms benefit from R&D and innovation schemes (and mixes thereof) and how this may affect firm behaviour and thereby to contribute to improved policy learning. The notion of the 'innovation journey' which traces the path of a firm between various innovation stages, knowledge and funding sources was repeatedly raised at the Workshops as a cautionary note.

³https://www.researchgate.net/profile/Johan_Wiklund/publication/278964456_Inside_the_black_box_of_outcome_additionality_Effects_of_early_stage_government_subsidies_on_resource_accumulation_and_new_venture_performance/links/558827d408aead25f0ed042b.pdf

When asked how do evaluations feed back into the actual policy design of the existing and possibly new R&D grant schemes and so induce learning, it seems that there is quite a gap and much room for improvement.

There is the view that an evaluation has to be designed for such a purpose, and thus, it needs to start with future-oriented policy questions (i.e. a clear statement of the policy's original intended rationale which prompted its introduction) and to be integrated in real policy processes. Such rationales do not merely reflect the purpose for which a policy measure has been introduced, but also the context within which it was launched and the barriers or problems that it was foreseen that it would overcome or address. Such background contexts are also highly likely to shift within complex innovation ecosystems and hence the original rationale should also be revisited through subsequent evaluations. Timing and the involvement of policy-makers is critical for future decision-making. Dialogue and interaction along and after evaluation is important.

However, so far, many evaluations are performed for specific units within ministries which use the results for improving their work. But spreading the knowledge gained within agencies and ministries or beyond, seems to require rules and formal channels through which this information and knowledge could reach out to policy-makers. Clearly, there is a need to strengthen the strategic use of evaluations.

Transparency can be quite beneficial: for example, in cases where programme designers and policy developers are involved as observers in the evaluation process and they are able to obtain the (publicly available) evaluation results. This transparency provides a degree of external pressure to address negative results and to modify the programme or by making failure part of the learning process in a non-stigmatic fashion. However, in a well-functioning policy cycle process, there should be optimal feedback loops between the programme designers, those administering or implementing the programme and those responsible for coordinating the evaluation, in order to facilitate policy learning.

Example 3 - Innovation incubators (IM) programme (Denmark)

Programme – IMs provide loans and equity finance to new firms in high-tech industries. Between 1998 and 2014 about 1.000 firms were supported with ~€250m (~€250.000 on average per firm).

Prior evaluation: The programme was evaluated externally twice previously with slightly contrasting results: i.e. from the one evaluation: "No detectable indications of positive effects" while the second concluded that "if one compares the financial returns since 2008 with the public capital invested, the ROI is 1:5."

Caveats: Data availability, reliability, and bias (companies may simply restart existing business, many IM supported companies are missing from the business register, others may have gone bankrupt before they get evaluated, presence of outliers...). Leakage of 20-30% of firms from the database. Descriptive results provide numbers but no assessment – hence a need for a benchmark or counterfactual.

Evaluation – variables: Survival rates, performance variables (employment, turnover, value added, annual earnings). Use of multivariate regression does not show significant differences to non-treated firms.

Conclusions: Low survival rate of participating firms and large heterogeneity in outcomes requires some sort of rethink of the link between programme design and the programme's potential measurable growth effects.

Effect on policy making: Not established.

Source: Evaluation available (in Danish): <http://ufm.dk/publikationer/2015/econometric-analysis-of-the-danish-innovation-incubator-programme>

2.4. Challenges evaluating R&D schemes

What are likely challenges the evaluating organisations are facing? For the final workshop in Brussels, participating countries were asked to indicate what type of issues they saw as most pressing in relation to evaluation of R&D-schemes. This is not a full-fledged large scale survey, but it provides indications what the participants in this mutual learning exercise envisage as important or less important.

As can be seen from the synthesised responses presented in Table 2, there was a high degree of consensus on the relative importance of the various types of challenges encountered. Especially the five challenges listed first, were seen by the participants as very important or important challenges.

While starting from the **rationale for a scheme** should be a given for evaluators, many schemes are not clear about their rationale. This makes it difficult for evaluators since the precise goals and framing of the scheme needs to be reconstructed (i.e. using a logical framework). Having a clear understanding of the rationale at the beginning of an evaluation helps in deriving a high quality evaluation. It was also pointed out that apparently, there still is a considerable need to improve the link between feeding back evaluation results into the policy-cycle. Here participants saw a clear responsibility of policy-makers and politicians to make sure these two are firmly linked. Policy-makers should use the programme logic charts to demonstrate the bigger picture of the policy cycle rather than using some instant numerical result (such as return on investment and leveraging) for their messages.

The participants also expressed, yet again, the need to better understand **why policies work** rather than simply measuring their effects. This can be seen as a plea to open up the black box of many schemes by talking to beneficiaries and to obtain a detailed understanding of how they use it, how it affects their R&D and innovation behaviour, and choices made (earlier referred as to better understand the 'innovation journey' of firms and their use of schemes).

Participants also considered the **use of a mix of evaluation methods** as highly important. Although there are still several adherents to either pure quantitative methodologies (mainly based on econometric modelling) or pure qualitative evaluation approaches, most respondents welcome a sensible combinational use.

Another important challenge is how to deal with the **complexity of programme logics** and multiple goals, which often require sophisticated sets of evaluation methodologies. This is seemingly a challenge linking to the need to involve external, specialised experts and the need for linking various data sources.

One can conclude that most participants share the view that evaluations of policy mixes need to be analysed more thoroughly. There are complex programmes as well as co-existing schemes with a greater variety of participants and/or different objectives but a meta-analysis and thus overview of how the various schemes relate to each other is by and large missing. For evidence-based policy-making systemic evaluations, looking at the integrated set of policies and support schemes in a given innovation system are needed as in reality, these schemes rarely exist in isolation.

Several aspects were much less uniformly ranked in terms of their importance. This may be due to the rather different roles, aims, expectations, etc. within each organisation as well as vis à vis outside stakeholders. For example, if one is not required to communicate with policy-makers or politicians, there may not be a need to balance detail and sophistication. The question of in-house versus external evaluations and the evaluation of generic versus more targeted R&D grant schemes did not appear to be regarded as major challenges either. Slightly surprisingly, there was a mixed view on the persistence of effects. This topic was discussed during the country seminars and it was referred to two kinds of persisting effects, namely those impacting directly on the performance of the firm and those that influence the longer-term policy 'behaviour' of the firm. Acknowledged as being a black box to programme designers as much as to evaluators, the lack of a clear understanding of firm rationale and firm behaviour – do firms apply for the same type of scheme over, do they apply to a variety and growing number of support schemes, or do they move along a chain of programmes? - prevails. This 'shifting' or 'stapling' behaviour is to our understanding hardly assessed yet. Insights into the combined use of (various) schemes over time is needed for sound programme development and better policy-making.

Table 2 View on relative importance of various challenges re the evaluation of R&D schemes

| Challenge and its importance | Very Important | Important | Neutral | Minor importance | Not important |
|--|----------------|----------------|----------------|------------------|---------------|
| Framing the rationale of the scheme (e.g. through the use of Programme Intervention Logic Diagrams, or related approaches) | Very Important | Important | | | |
| Maximising the input to policy learning (needed to help in better policy designs) | Important | Very Important | | | |
| Better understanding why policies work rather than simply measuring their effects | Important | Important | Neutral | | |
| Combined use of various evaluation methodologies | Neutral | Very Important | | | |
| Measurement of behavioural effects induced by scheme participation | Neutral | Very Important | | | |
| Increasingly sophisticated evaluation methodologies | Neutral | Very Important | | Minor importance | |
| How to evaluate schemes with multiple goals | Neutral | Important | Neutral | | |
| Balancing detail and sophistication versus the need to communicate with policy-makers, politicians, etc. | Very Important | | | Minor importance | |
| Evaluation of policy mixes (could be all R&D grant schemes together or broader ranges of support) | Important | Important | Neutral | | |
| Balancing internal requirements for the evaluation versus influencing the view of external parties (e.g. Treasury) | Very Important | | Neutral | Minor importance | |
| Importance of investing in ways to better communicate the results of complex analysis to a wider audience | Neutral | Important | Neutral | Minor importance | |
| Persistence of effects: optimal length of participation in scheme (temporary use or permanent use) | | Very Important | Neutral | Minor importance | |
| Use of big data in the evaluation of R&D grant schemes | | Important | Neutral | Minor importance | |
| Evaluation of generic vs more targeted R&D grant schemes | Neutral | | | Minor importance | Not important |
| In-house versus external evaluations | | | Very Important | Minor importance | Neutral |

Example 4 - Measuring the economic impact of research joint ventures supported by the EU Framework Programme (Spain)

Aim: Analysing research joint ventures (RJV) supported by the EU R&D Framework Programme (FP) and their impact on the labour productivity of private participants

Evaluation methodology: Based on the following structural model: (i) firms decide whether or not to apply for a FP cooperation project; (ii) the proposal is awarded or rejected by the European Commission; (iii) the innovation activities involved in the cooperative R&D project succeed through the generation of new knowledge (represented by intangibles); (iv) the addition of this new knowledge to the production process results in productivity growth. The self-selection bias exists and the impact equation has been estimated by alternatively using the predicted and the observed probability of being granted. The CDTI-PM database includes information about all the proposals, eventually granted or not, in which at least one Spanish firm participated between 1995 and 2005. This information has been complemented with the SABI database that contains the company accounts of more than 1,000,000 Spanish firms.

| What proved beneficial... | What proved challenging... |
|---|--|
| <ul style="list-style-type: none"> • Available data coming from complementary sources: administrative and economic performance databases. • Collaboration between innovation agencies and university experts. | <ul style="list-style-type: none"> • R&D performance is not available in commercial databases coming from accounting sources. • Access to micro data on innovation. • Methodologies to avoid self-selection bias. |

Findings: FP cooperation increases the ratio of intangible fixed assets by employee and also positively affects labour productivity. These effects come out stronger in high-tech manufacturing and services. Medium-tech manufacturing firms have a higher potential to generate technological outputs. This result supports an indirect effect of participating in the FP on firms' performance, captured throughout labour productivity.

Source :

https://www.researchgate.net/publication/244993167_Measuring_the_Economic_Impact_of_Research_Joint_Ventures_Supported_by_the_EU_Framework_Programme

2.5. Methodological aspects

While the evaluation of R&D schemes needs to be set in a broader context as we have summarised above, the MLE dealt with a number of hands-on problems that some evaluating bodies are facing. Other evaluation bodies manage these problems but may be confronted with other issues.

Requirements & limitations of advanced econometric types of analysis

In this MLE, the trend towards more quantitative/econometric type of analysis was discussed extensively in both the Madrid and Copenhagen workshops. In the Madrid workshop Professor Bart Verspagen (UNU-MERIT) discussed the econometric evaluation of R&D incentive schemes and in particular the Dutch Innovation box that was evaluated in 2015 (see den Hertog et al, 2015). The following messages deserve highlighting:

- This type of evaluations benefit considerably from high quality administrative data that can also be coupled to micro-level firm statistics;
- Finding the right control groups is key and at the same time, limited data availability can limit analytical options. Especially in smaller countries the number of firms operating in a given sector may be too small for these types of analyses;
- There is a considerable discussion among scholars on the "how" of econometric analysis and there are many decisions to be made before an evaluation design is ready. Scholars make different methodological choices thus reducing the possibilities for comparing individual evaluation results;
- There are only a few schemes where high quality data are available in abundance, data availability differs considerably among countries;
- Working procedures regarding access to data and confidentiality rules are far from being solved everywhere;
- The communication of econometric results is a challenge: typically only the results in terms of additionality are communicated but rarely the assumptions and choices made;
- A development towards mixed approaches i.e. using quantitative and qualitative methods is preferable.

Both the Madrid and Copenhagen seminars included presentations building on econometric analyses of particular policy schemes (or even policy experiments in the form of a randomized natural experiment). This type of work is increasingly published and extending econometric type of analyzing on R&D schemes which to date is biased toward R&D tax credit schemes (see e.g. Huergo et al.; Barajas et al., DASTI (2014), Christensen et al. 2016

One of the main factors that seem to distinguish interventions addressing private sector companies from those aimed at the public sector is a lack of information about the benefitting companies and measuring the impact of the intervention - unless of course, as noted above, the country has access to extensive and well-structured data infrastructures which are suitable for the application of quantitative approaches. During recent years, micro-economic counterfactual impact evaluations

gained a wide acceptance for assessing the impacts of treated vs. non-treated firms. Standard practice is the use of econometric methods such as regressions or matching techniques. Much less standard yet are randomised control trials or natural experiments. These econometric methods need to take a number of factors into account, in particular the fact that firms select themselves into programmes: they decide to apply or not. Another important aspect is that many programme owners want to pick winners – there is therefore a selection bias. This needs to be taken into account for establishing a *correct* control group for providing answers on typical questions such as job creation, innovation investments and activities.

Additionality effects can equally be addressed; for input additionality, the literatures as well as plenty of evaluations have dealt with these issues and it seems a reasonably straightforward analysis is provided. The Community Innovation Survey (CIS) as well as other innovation and R&D surveys can be stated as examples for readily accessible data sources for analysis. When it comes to output additionality, the design of the econometric analysis becomes more complex and the evaluators need to address data characteristics in conjunction with the evaluation questions addressing additionality. The effects of the use of various or multiple instruments (policy mix) can to some extent be addressed through matching techniques, again, always under the premise of available data, they can answer a number of questions by comparing the treated with the control group firms and explain additionality effects for the (non-) participation of one or more programmes but also conceivable crowding out effects due to multiple treatments etc. (Czarnitzki/Debackere 2016). While there may be some examples, the majority of evaluations cannot address these factors precisely due to the lack of available comparative data.

While individual methodologies were discussed in greater detail during the country visits, we refer to existing toolboxes (see Annex) for the discussion of various methods and approaches. As noted in the introduction, it was not the purpose of the MLE nor of this report to provide a broad ranging prescriptive set of evaluation practices or guidelines. These exist elsewhere and require careful consideration of the specific country context prior to their adoption. Annex A Figure 1 gives an idea about the complex decision-making process about the right tool for a given dataset - based on essential objectives of the analysis.

While the econometric modelling gained in use in evaluations, its practical value to policy makers may be somewhat limited - the insights the modelling conveys may only be partially relevant and many questions remain unanswered. While there is still room to improve the econometric modelling (see Annex A, Figure 2 - the modified Maryland Scientific Methods Scale (SMS)), the participants in this MLE openly expressed a preference for a mixed-methods approach which uses quantitative and qualitative techniques.

Example - 5 System evaluation of KMU Innovativ (Germany)

Programme: As part of the Hightech-strategy, the programme aims to leverage participation of innovative SMEs in specific thematic project based funding. Beside the actual funding, it includes several tools to help SMEs to apply. On average a funded SME received 110.00€ per year, about 20% of its total annual R&D expenditure. The evaluation covered the first 3.5 years of the programme (2007-2010).

Evaluation goal: The system evaluation aimed to provide a holistic view on adequacy, goal attainment, effectiveness, efficiency of implementation and positioning within the public R&D funding landscape.

Evaluation methodology: Mix of qualitative and quantitative means for each of the main aspects: a benchmark was established, interviews, document analysis, analysis of application data was used for the implementation analysis while goal attainment was analysed using a comparison to a sample of innovative SMEs; effectiveness used a control group approach. A survey of applicants was performed. The information was coupled with existing company databases and panel data owned by the service providers.

| What proved beneficial... | What proved difficult... |
|--|--|
| <ul style="list-style-type: none"> • Available, private company databases incl. panel data to link to survey data; • Mix of qualitative and quantitative methods | <ul style="list-style-type: none"> • Classification of SME as such • Necessity to merge different data bases • Quality of data in databases of different funding agencies • Comparison of different funding schemes within one technology field due to non-existing common thematic classification |

- Identification of interview partners due to lack of necessary personal information in some databases

Findings: Even during the financial crisis 2008/2009, each euro spent leveraged 1.5€ of private investment of the participants. In total, the programme leverages the annual R&D investments of SMEs by 30% leading to an increase of R&D intensity from 21% to 27% during the funding. Control group companies who received funding elsewhere obtained similar results.

Source : <http://ftp.zew.de/pub/zew-docs/gutachten/KMU-innovativ2012.pdf>

3. A SELF-ASSESSMENT OF THE PROCESS

- The MLE brought together five countries which were willing to participate in the exercise and a small number of observer countries which participated in the opening workshop. In the view of the participants, **the number of participating countries would seem to be ideally in the range of four and eight**. While the group recognized that the challenges were probably shared by more countries and thus more countries could benefit from the exchange, the group size should not be too large in order to enable an open, mutually trustful atmosphere and which could maintain focus and momentum throughout the entire series of workshops. The element of 'buy-in' among all the participants was an essential ingredient
- This MLE started with a **kick-off workshop** in Brussels open to participating and associate countries and for which the experts had prepared interventions intended to highlight a number of the key issues and some of the state of the art concerning the evaluation of R&D support schemes targeted at companies. The workshop was followed by **two consecutive one day country seminars** (within ten days at the end of April, early May), and a final workshop (May 2016). In retrospect, to several participants, the *speed and intensity of the meetings* was very high and a slightly slower pace would have benefitted the organisation of the country seminars, allowing more space and time to reflect on the outcomes of the workshops and to highlight specific problems, issues and challenges of interest.
- In terms of **formal and informal outputs**, several contributions have been prepared by the host countries for the seminars. These presentations were provided either by external experts invited by the hosts or by the hosts themselves. In the latter case, the hosts (and MLE participants) often explained how evaluations are planned and executed by their organization. Invited experts then provided more insights on specific evaluations of business schemes. The different presentations and following discussions captured a large number of methodological challenges.

The experts to this group aimed to provide ideas and insights with their interventions – written and oral – which were equally integral parts of the meetings and this final report.

In addition to the country seminars, it was felt that a general survey of the participants should be taken to gain an oversight of the different evaluation contexts pertaining in each of the participating country. These insights were provided in a structured form, which fed into this final report and which was also disseminated amongst the participants, thereby allowing them to identify those issues that were shared, or which had already been addressed by others.

- The schedule of this MLE – in particular the time between the two country seminars was set too narrowly. It was felt by the participants that a gap of two to three weeks between the seminars (in Madrid and Copenhagen) would have allowed a better alignment of the content of the seminars.

What can be improved

This MLE started on the basis of a rather pre-defined subject matter. However, the kick-off meeting in Brussels had to deal with the difficulty in pinpointing the scope of the whole exercise. For the participants, the absence of advance, 'background' information, which set out the framework of challenges and issues with which the participants could align themselves, was missing. Thus, there is a need for *careful focus and greater clarity on the specific topic or focus and of the roles and expectations of the various participants* (Commission, experts, facilitator, country participants).

More time in between the country visits were expected to be more helpful. The limited time between the two country visits proved to be a slightly weak point. While certainly the memory

about the past workshop is fresh, the organisers would have preferred to take into account the discussions and the results of the first seminar in planning the second. In this MLE, both seminars were planned in parallel and even if they were intended to be different in scope, there were similarities that could have been avoided with a greater time span in-between. This can be rectified by allowing more time for country hosts to prepare for workshops and for all participants to reflect on their expectations from the meetings and what they can also bring to the table.

The host countries feel a limited scope to learn when they host and organise the seminars. Thus, *more than two country seminars would be welcome*. A common structured format for a country visit/workshop could be addressed for future MLEs.

What proved beneficial

The **participants were positive about the learning process** they went through, acknowledging a *high level of 'buy-in'*. Their active participation was key for assuring the efficient information exchange and all participants invested considerable efforts in the MLE. It was also felt that the choice of participants was appropriate as these were the actual participants that would benefit directly from increased mutual learning. The country visits in particular were identified as exercises that were really helpful in learning from other practices and reflecting on one's own practices. During the country visits and various workshops new 'challenges' were identified during the MLE. These can possibly be followed-up in the next cycle of this MLE (opening up to new countries as well).

Future of this particular MLE

Possibly the best indication of the usefulness of this MLE is the **willingness of the participating countries to extend this small learning community** and the idea to possibly **involve each other during actual evaluations** to encourage exchange and learning on the job (peer/sparring on individual evaluations). The participating countries have expressed a willingness to remain as a group beyond this current setting.

BENCHLEARNING AND NEXT PRACTICES IN EVALUATION

Sweden proposed to exploit further the concept of bench-learning. Going to a country is more time consuming than just reading a paper but it is very useful since one can learn things. Organisations in advanced countries have a better chance for learning – thus visiting other countries offers a good opportunity to learn and to implement the learning in the home organisation. An evaluation report can only present the process and outcome of the evaluation i.e. what happened but rarely explains why things happened or didn't happen. Denmark added that context and country specific aspects are key factors for learning opportunities. In addition, only by talking with the policymakers responsible for the evaluation or for commissioning it, does a real picture emerge as to why the evaluation took the shape and form that it did, what prompted the decision to use certain approaches and methodologies, what were the anticipated impacts of the evaluation, etc.?

Norway suggested to start discussion on upcoming developments in the practice of evaluations and suggested another workshop in Oslo.

4. CONCLUSIONS

The MLE participants recognized that despite being in different countries, many of the challenges are shared. Thus, several of the topics that were presented and discussed have direct relevance in relation to analyses others are doing. One can for example see that the absorptive capacity for political learning is generally underdeveloped in most countries and in this respect, national challenges are similar.

What countries will 'take with them' is taking into account the mix between "hard impact data/results" and "soft impact/process related impacts". The relevance and complementarity of qualitative and quantitative methodologies have clearly been acknowledged. Qualitative methodologies should be applied (such as interviews) since they have an explanatory power that quantitative methodologies lack.

A relevant aspect that will equally impact the day to day work is a more holistic view on the policy cycle, starting with an assessment of the rationale for the policy intervention (i.e. the purpose for which it is being created), programme design, implementation, monitoring to measuring impact. It is certainly challenging to come from 'simple' evaluation and impact analysis to policy and policy

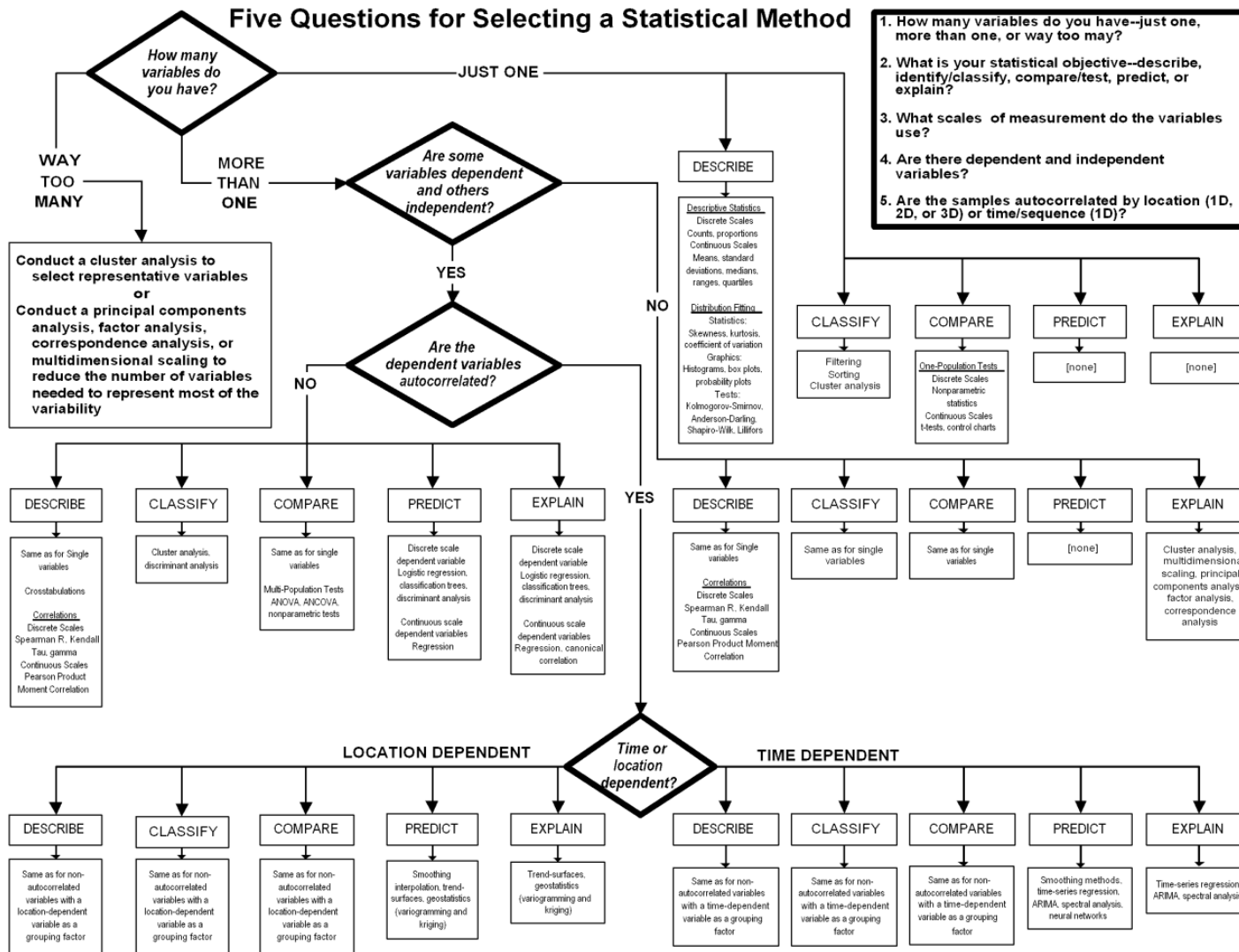
learning. Evaluations should be a strategic issue at the policy level: the use of evaluation as a 'bolt-on' afterthought completely undermines its value as a policy learning mechanism.

Behavioural effects are somewhat less well analysed and the holistic approach can help understanding firm participation behaviour and identify changes in the innovation behaviour of companies. In this context, it is also useful to address the effects of simultaneous multiple participation in schemes. Thus, it is important not just to understand what firms get from their participation in policy support instruments, but also what they expect to gain, their motivations for participating and how such participation influences their subsequent behaviour.

A more forward-looking suggestion is that future MLEs would benefit from developing a format for "bench learning" (i.e. direct experiencing the full policy-cycle analysis in country visits) exercises and subsequent codification of lessons.

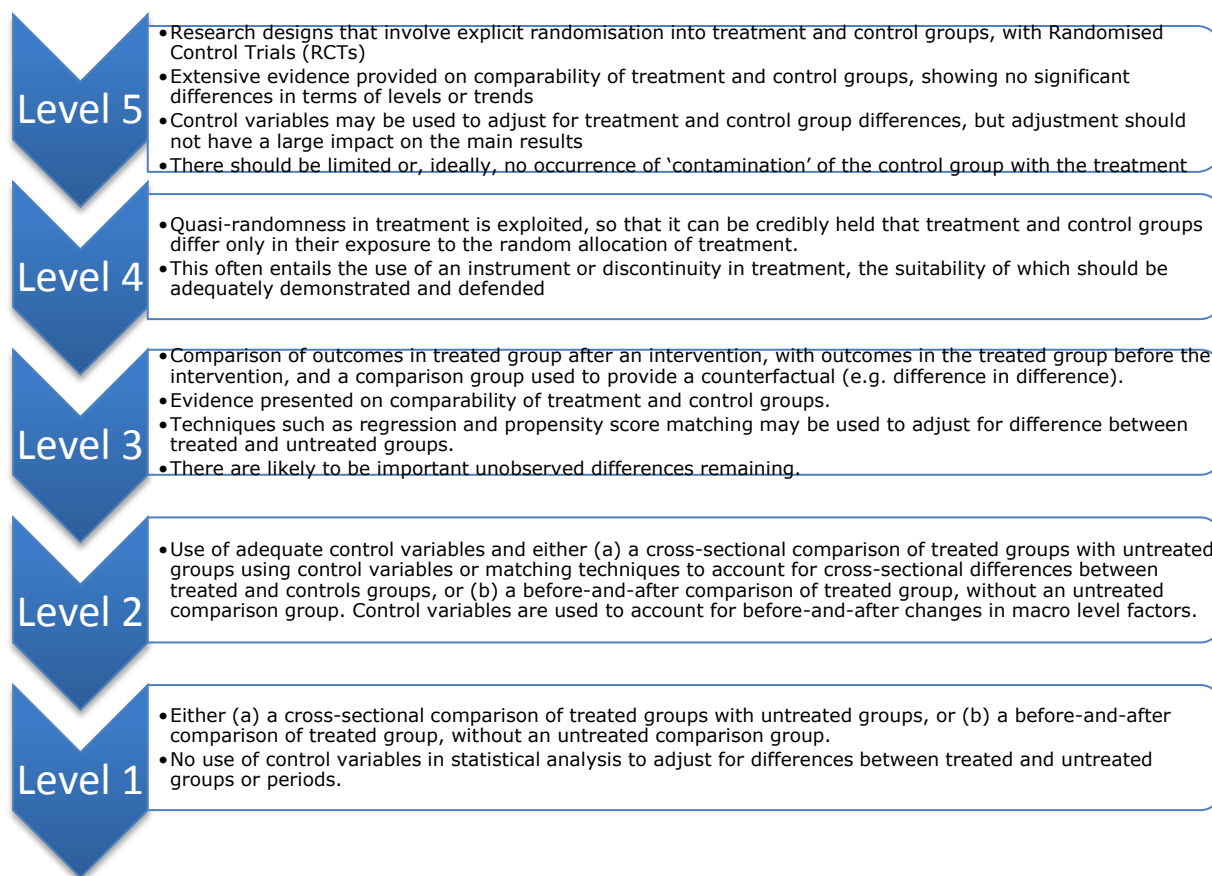
Appendix A - Additional material

Figure 1 Decision tree for statistical tools and techniques



Source: <https://statswithcats.wordpress.com/2010/08/27/the-right-tool-for-the-job/>

Figure 2 Modified Maryland Scientific Methods Scale (SMS)



Note: The modified SMS is used as a to screen evaluations on their level of robustness. According to Czarnitzki and Debackere (2016), current standard practice is level 3. Source: adapted from: <http://www.whatworksgrowth.org/resources/the-scientific-maryland-scale/>

Terminology used in the discussed evaluations

Crepon-Duguet-Mairesse (CDM). A structural model which uses simultaneous equation models, allowing to measure input and output additionality effects of treated units.

Matching is a statistical technique used to evaluate the effect of a treatment by comparing the treated and the non-treated units in an observational study or quasi-experiment (i.e. when the treatment is not randomly assigned). The goal is to find for every treated unit one (or more) non-treated unit(s) with similar observable characteristics against whom the effect of the treatment can be assessed.

Ordinary least squares (OLS) is a method for estimating the unknown parameters in a linear regression model.

Propensity score matching (PSM) is a statistical matching technique which attempts to reduce the bias due to confounding variables.

Appendix B - Bibliography

Athey, S. and Imbens, G. (2016): The State of Applied Econometrics - Causality and Policy Evaluation: Available online at: <https://arxiv.org/pdf/1607.00699.pdf>

Barajas A., E. Huergo and L. Moreno (2016): SME performance and public support for international R&D, in: Journal of Small Business Management, Vol. 54 (4), pp. 1206–1228.

Boekholt, P., Arnold, E., Giaracca, F., Ploeg, M. (2014) : Evaluation reference model in search for a benchmark of impact, effectiveness and efficiency of innovation instruments. Report for TAFTIE. <http://www.taftie.org/sites/default/files/Task%20Force%20BIEE%20%20Evaluation%20Reference%20Model-140114.pdf>

Capron, H., van Pottelsberghe de la Potterie, B. (2007). Public support to R&D programmes: an integrated assessment scheme, OECD. Paris. Available at: <http://www.oecd.org/sti/inno/1822568.pdf>

Crespi, G., Maffioli, A., Mognen, P., Vazquez, G. (2011) : Evaluating the impact of science, technology and innovation programs : a methodological toolkit. Impact-Evaluation Guidelines. Inter-American Development Bank

Christensen, T.A., J.M. Kuhn, C. Schneider and A. Sørensen (2016), Science and Productivity. Evidence from a randomized natural experiment, *SIMPATIC final conference, Brussels, February*. 2015.

Cunningham, P. and Gök, A. (2010): Usefulness of Evaluations. In: INNO-Appraisal: Understanding Evaluation of Innovation Policy in Europe, Final Report, Pro-INNO Europe.

Cunningham, P., Gök, A., Laredo, P. (2013) : The impact of direct support to R&D and innovation in firms. NESTA Working Paper 13/03 (see also Repositories below)

Czarnitzki, D., Debackere, K. (2016): Towards a portfolio of additionality indicators. Paper presented at the OECD Blue Sky conference, Gent 19-21.09.2016

DASTI (2014): Short-run Impact on Total Factor Productivity Growth of the Danish Innovation and Research Support System. <http://www.fivu.dk/publikationer>

DASTI (2015), Econometric Analysis of the Danish Innovation Incubator Programme, Danish Agency for Science, Technology and Innovation, København . Available at : ufm.dk/publikationer

Fahrenkrog, G. (ed.) (et al) (2002): RTD Evaluation Toolbox – Assessing the socio-economic impact of RTD policies. IPTS/Joanneum

Hertog, den P., Vankan, A., Janssen, M., Minne, B., Korlaar, L., Erven, B., Verspagen, B., Mohnen, P. (2015). Evaluation Innovationbox 2010-2012 (in Dutch). Report for the Dutch Ministry of Finance, Dialogic, Utrecht.

Huergo, E., M. Trenado and A. Ubierna (2015) "The impact of public support on firm propensity to engage in R&D: Spanish experience". Technological Forecasting & Social Change.

Miles, I. and Cunningham, P. (2006) SMART INNOVATION: A Practical Guide to Evaluating Innovation Programmes. ECSC-EC-EAEC Brussels-Luxembourg. Available at: https://cordis.europa.eu/innovation-policy/studies/pdf/sar1_smartinnovation_master2.pdf

OECD (2015) : OECD Data and statistics on R&D tax incentives. Highlights. <https://www.oecd.org/sti/RDTaxIncentives-Data-Statistics-Scoreboard.pdf>

Papaconstantinou, G. and Polt, W. (2007). Policy evaluation in innovation and technology: an overview, OECD, Paris. Available at: <http://www.oecd.org/sti/inno/1822393.pdf>

Ploeg, M., Van der Veen, G., Arnold, E (2015) : Measuring innovation policy across Europe – Common indicator framework. Report for TAFTIE
<http://www.taftie.org/sites/default/files/Common%20Indicator%20Framework%20for%20publicati%20on.pdf>

[Technopolis Group & MIOIR \(2012\): Evaluation of innovation activities. Guidance on methods and practices. Study funded by the EC, DG-REGIO. Available at: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/eval2007/innovation_activities/inno_activities_guidance_en.pdf](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/eval2007/innovation_activities/inno_activities_guidance_en.pdf)

Warwick, K. and A. Nolan (2014), Evaluation of Industrial Policy: Methodological Issues and Policy Lessons”, OECD Science, Technology and Industry Policy Papers, No. 16, OECD Publishing. Available at: <http://dx.doi.org/10.1787/5jz181jh0j5k-en>

Repositories of evaluation studies

Nesta, Manchester Institute of Innovation Research: Compendium of Evidence on Innovation Policy: <http://www.innovation-policy.org.uk/>

Repository of the Austrian platform fteval : http://www.fteval.at/en/evaluation_studies/

Website of the LSE in cooperation with Centre for Cities and ARUP:
<http://www.whatworksgrowth.org/policy-reviews/innovation/>

Science and Innovation Policy Evaluations Repository (SIPER) – operated under the EC RISIS project: forthcoming (anticipated opening November 2016).

Appendix C – Seminar’s agendas

How to benefit from evaluating R&D grants/schemes I: strategic approaches and challenges related to data management.

The example of Spain

Date: 26 April 2016

Place: CDTi, Calle Cid, 4, 28001 Madrid

Draft agenda

| Timing | Sessions | Item by host | Q to be addressed |
|--------------------|---|--|---|
| 9.15 – 9:30 | Welcome by Host (TBC) and Chair <i>A. Bladh</i> Tour de table | Introduction session | |
| 9:30-13:00 | MORNING SESSION | | |
| 9:30-10:30 | Evaluation as a strategic issue | <ul style="list-style-type: none"> How to design a complete evaluation strategy: the case of the CDTI. Spain will present the Evaluation Strategy of the CDTI: objectives, conceptual framework, actions, resources and timing. Speaker: CDTI representative and other participants | <ul style="list-style-type: none"> Other participants may share their strategic approaches and how the evaluation tasks are internalized in their organizations. Similarities and differences between countries could be discussed. |
| 10:30-11:00 | Coffee Break | | |
| 11:00-11:45 | Monitoring – an important tool for evaluations. | <ul style="list-style-type: none"> The case of the results monitoring system implemented by the CDTI. Speaker: CDTI representative | <ul style="list-style-type: none"> How to take advantage of administrative data and operational procedures to implement a monitoring system. |
| 11:45-13:15 | Why econometrics? | <ul style="list-style-type: none"> Bart Verspagen, Maastricht University | <ul style="list-style-type: none"> When is the use of econometric methods for evaluations of R&D schemes recommended and when not? (requirements both in terms of data, capabilities, costs) What are the type of results that no other method can provide? What are the major difficulties encountered using econometric methods? |
| 13:15-14:15 | Lunch Break | | |
| 14:15-16:30 | AFTERNOON SESSION | | |
| 14:15-15:00 | Methodological challenges: How to manage and disentangle the joint effect of different aid schemes? | <ul style="list-style-type: none"> Lourdes Moreno, UCM Lourdes Moreno will focus on three types of instruments: the programme of low-interest loans provided by the CDTI, the national scheme of R&D subsidies; and the European system of R&D grants. | <ul style="list-style-type: none"> Which difficulties are faced when we try to measure joint effects of different programmes? |
| 15:00-15:15 | Coffee Break | | |
| 15:15- | The international | <ul style="list-style-type: none"> Elena Huergo, UCM | <ul style="list-style-type: none"> Which lessons could be |

| | | | |
|----------------------|---|---|---|
| 16:00 | dimension of evaluation - learning from peers in other MS | <ul style="list-style-type: none"> Based on Simpatic Project, Elena Huergo will present which steps have been taken regarding the international joint evaluation of similar R&D programmes | learnt from international evaluations of similar programmes? |
| 16:00 - 16:30 | 10 ways to spoil an evaluation | <ul style="list-style-type: none"> Pim den Hertog | <ul style="list-style-type: none"> how the interaction between principal and evaluation team and other stakeholders can be counterproductive |
| 16:30-16:45 | Wrap up | <ul style="list-style-type: none"> A. Bladh | |

Agenda 3rd May 2016 Copenhagen seminar

Venue: Danish Agency for Science Technology and Innovation, Bredgade 40, 1260 Copenhagen – Meeting room 24

| Timing | Sessions | Item by host | Q to be addressed |
|--------------------|--|--|---|
| 9.15 – 9:30 | Welcome by Host (Director, Andreas Graversen) and Chair A. Bladh Tour de table | Welcome | |
| 9:30-13:00 | MORNING SESSION | | |
| 9.30-10.00 | Strategic approach towards choice of evaluation framework | <ul style="list-style-type: none"> Short introductory presentation by Thomas Blomgren-Hansen on Danish choice of evaluation. Why Denmark is primarily focused on quantitative evaluations. Strategic considerations? Discussion in the group of the approach | <ul style="list-style-type: none"> We should be careful avoiding to much overlap with the discussion in Madrid on “Why Econometrics”, but participants are invited to provide their views on the right balance between quantitative and qualitative approaches and the reasoning behind. |
| 10:00-10:45 | Introduction to the Danish R&D data infrastructure | Presentation by Henrik Barslund Fosse and Hanne Frosch, DASTI, of: <ul style="list-style-type: none"> The InnovationDanmark database collecting information on participation in all our innovation programs Research Access to micro level register data in Statistics Denmark Constructing/collecting new types of data Development of new data (budget and trade data) Loss of data – addressing the challenge concerning loss of data during data wash or set up | Mainly and informative session describing the Danish set-up. Participants are very welcome to ask questions on the set-up etc. Possible issues to address in the discussion: <ul style="list-style-type: none"> Importance of micro data Which data are the right data Data availability vs. resource costs (incl. beneficiaries) Timing of data collection How do we identify individual activities (tech adaptation, core R&D, organisational elements...) |
| 10:45-11:00 | Coffee Break | | |
| 11:00-11:45 | Challenges concerning the size of the impact relatively to the R&D grant based on a RCT analysis | Introductory presentation by Associate Professor, Cedric Schneider, CBS (Confirmed) <ul style="list-style-type: none"> Case: RCT Analysis of the Innovation Voucher Program where results are very positive. Article has just been submitted for peer review. Article can be circulated in advance. Discussion on plausibility of positive results and possible catalyzing effects. What can explain effects? | This session will focus on positive results and possible catalyzing effects. What can explain effects? This could lead to a discussion on: <ul style="list-style-type: none"> R&D behavioural additionality? R&D input additionality? Catalyzing effects? Political justification of very positive results? What type of effects can we measure? Effect of programs on innovation efforts and innovation outputs? |
| 11:45- | Challenges | Introductory presentation by | Based on the introduction the |

| | | | |
|----------------------|--|---|---|
| 12:30 | concerning reference group and outliers (big wins) – Alternative evaluation set-ups. | external expert, Consultant, PhD, Johan Kuhn (Confirmed) <ul style="list-style-type: none"> Learning exercise based on the evaluation of the Impact of Danish Innovation Incubators where some results are fairly challenging | discussion in this first part on session can focus on: <ul style="list-style-type: none"> Treatment of outliers Challenges with reference group Participants are invited to provide examples on how they have treated these issues The Danish evaluation (in English) will be sent to participants prior to the meeting |
| 12:30 – 13:15 | Lunch Break | | |
| 13:15-16:30 | AFTERNOON SESSION | | |
| 13:15-14:15 | Continued: Challenges concerning reference group and outliers (big wins) – Alternative evaluation set-ups. | <ul style="list-style-type: none"> Learning exercise based on the evaluation of the Impact of Danish Innovation Incubators where some results are fairly challenging Active involvement of participants. | In the second part of the session each participant is asked to provide a 5-8 minutes presentation on how they would approach an evaluation of this scheme in their own country? I.e. a presentation of a sketch (quantitative or qualitative) evaluation design of a seed/preseed program based on their own experience/methods/data availability from their own organization The Danish evaluation (in English) will be sent to participants prior to the meeting |
| 14:15-15:00 | Challenges concerning multiple participation | Introductory presentation by Moira Daly, CEBR (TBC) or alternatively Henrik B. Fosse from DASTI <ul style="list-style-type: none"> Case based on not yet published analysis (CEBR 2016) on multiple participation in R&D programs Discussion on alternative approaches – different effects and different methodological approaches Needed information to carry out evaluation of multiple participation in one or more R&D programs Draft analysis can be forwarded to participants | Discussion based on the presentation, which could address: <ul style="list-style-type: none"> How to approach analysis of multiple participation Expected effects of multiple participation Something on attribution if firms combine various R&D schemes? Treatment of timing issues Alternative approaches to evaluation of multiple participation |
| 15:00-15:15 | Coffee Break | | |
| 15:15-16:15 | Use of evaluation – how are they used in policy development and program designs? | <ul style="list-style-type: none"> Invited experts from the department on innovation policy, Nicolai Zarganis (Confirmed) and the Danish Innovation Fund Michael Philip Poulsen (Confirmed) will explain how they use evaluations in their policy development and program designs. Q&A plus discussion on how evaluations are used in other countries | Wrap up discussion based on the seminars in Madrid and Copenhagen. The main discussion question is how evaluation results are used in development of policies and program design. Sub-issues which could be addressed in the discussion are: <ul style="list-style-type: none"> What types of evaluations Are most relevant in policy and program development? What do we learn from econometric type of evaluation for the actual design of schemes? What do we learn from qualitative type of evaluation for the actual design of schemes? What do we learn from evaluations about how the scheme affects innovation |

| | | | |
|----------------------|---------------|---|---|
| | | | <p>decision-making & innovation behaviour in the firm</p> <ul style="list-style-type: none"> • Can results be used in a wider macroeconomic perspective - effects on productivity, exports, employments etc. |
| 16:15 - 16:30 | Final session | <ul style="list-style-type: none"> • Wrap-up by Agneta Bladh | |

Draft Agenda

Final workshop on

Ex-post evaluation of business R&I grant schemes

Brussels, 30.05.2016, 11:00-15:30

| Timing | Sessions | Item |
|----------------------|---------------------------------------|---|
| 10.30 – 11:00 | Arrival | Arrival of participants |
| 11.00 – 11:15 | Opening <i>Chaired by A. Bladh</i> | <ul style="list-style-type: none"> • General Introduction by the Chair and EC • Tour de Table |
| 11:15-13:00 | MORNING SESSION | |
| 11:15-12:15 | The survey results | <ul style="list-style-type: none"> • Overall presentation • Discussion |
| 12:15 - 13:15 | Feedback on MLE as a new instrument | <ul style="list-style-type: none"> • Feedback on the process itself • How does the content/subject matter impact the discussion? • Previous and future expectations of the participants <p>Feedback from EC on how the other groups do and how they report (Eva)</p> |
| 13:15 – 14:00 | Lunch Break | |
| 14:00-15:30 | AFTERNOON Session | |
| 14:00 - 15:00 | A final report of the MLE | <ul style="list-style-type: none"> • Demand of the participants • Discussion of a suggested structure (expert group) |
| 15:00 – 15:30 | Conclusions | <ul style="list-style-type: none"> • Closing remarks by the Chair/EC • Next ERAC meeting (October) |

How to obtain EU publications

Free publications:

- one copy:
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:
from the European Union's representations (http://ec.europa.eu/represent_en.htm);
from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

The 'Policy Support Facility' (PSF) was set up by the Directorate-General for Research & Innovation (RTD) of the European Commission under the European Framework Programme for Research and Innovation 'Horizon 2020', in order to support Member States and associated countries in designing, implementing and evaluating reforms of the national science, technology and innovation systems.

One of the services offered by the PSF are Mutual Learning Exercises (MLE) which are a project-based learning processes whereby over a period of about six months participating countries jointly examine a challenge-driven policy question in detail. The MLEs involve information acquisition and information sharing activities.

The purpose of this particular MLE was to improve the exchange of information and identification of good practices between the participating countries (Denmark, Germany, Norway, Spain and Sweden) and in this way contribute to put in place better systems for the ex-post evaluations of business R&I grant schemes.

This MLE discussed various examples of what some regard as the "golden standard" in evaluation. The three key observations were: evaluations using econometric analyses are far from standardized and complex to perform; econometric analyses are very demanding in terms of data-availability and quality; the obstacles regarding access to data and data confidentiality are far from being removed in only but a few cases.

While the trend towards econometric analysis continues unabated, it was consensus within the group that it would need to be balanced with an understanding of behavioural effects, i.e. the "innovation journey of firms", of the use of R&D and innovation grants. Thus, some of the main challenges and questions discussed by this MLE group included: the use of econometric analyses from other domains, as for example education and labour market studies; how to deal with the interaction of various schemes used in combination and how to measure and understand behavioural effects.

Studies and reports