OECD publishing

Please cite this paper as:

Warwick, K. and A. Nolan (2014-07-03), "Evaluation of Industrial Policy: Methodological Issues and Policy Lessons", *OECD Science, Technology and Industry Policy Papers*, No. 16, OECD Publishing, Paris.
http://dx.doi.org/10.1787/5jz181jh0j5k-en

OECD Science, Technology and Industry Policy Papers No. 16

Evaluation of Industrial Policy

METHODOLOGICAL ISSUES AND POLICY LESSONS

Ken Warwick,

Alistair Nolan

FOREWORD

Over recent years there has been a revival of interest in industrial policy amongst policy makers around the world. At the same time there have been significant developments in the methodologies for policy evaluation. However, rigorous and systematic evaluation evidence in many areas of industrial policy is scarce, and particular methodological challenges exist.

Initial work overseen by the OECD's Committee on Industry, Innovation and Entrepreneurship (CIIE) involved a review of the literature on industrial policy and its evaluation. This work resulted in a STI Policy Paper, which was published in April 2013 (Warwick, 2013).

Given the relative dearth of peer-reviewed literature in this domain, in order to further reflect on these issues an Expert Group on the Evaluation of Industrial Policy was established by the CIIE at its meeting of April 2012. The objective of the establishment of the Group was to draw upon the first-hand insights of government officials and academics working on the evaluation of industrial policy.

More specifically, the Expert Group had a number of aims, namely: to improve knowledge about evaluation and spread understanding of the related methodological issues; to help improve the consistency of evaluation studies across countries; and to contribute to lesson-learning about policy effectiveness, including the responses to the economic and financial crisis.

The Expert Group met for the first time in September 2012. Subsequent meetings were held in January and September 2013 and January 2014. At each meeting, time was allocated to presentations of national evaluations, with inputs from government officials, academics and independent experts. The Group decided that rather than trying to cover all aspects of industrial policy, it would focus on a limited number of areas, namely:

- Three specific policy instruments – support for R&D; capital market interventions (with a focus on risk capital); and public procurement for innovation; and,
- Three areas where packages of industrial policy measures are generally applied – sector approaches (including public-private partnerships); policies towards clusters and business networks; and national industrial strategies.

The Group also discussed industrial strategy initiatives in a number of countries and approaches to 'meta-evaluation'. Consideration was likewise given to new approaches to evaluating industrial policy, including the use of randomisation and experimental techniques, and the implications of evaluation approaches and findings for the design and implementation of policy.

While the paper draws upon a review of much of the relevant policy evaluation literature, its primary contributions are drawn from the presentations and deliberations of the Expert Group. It does not aspire to be a comprehensive and definitive analysis of the field, but rather a synthesis of the views and insights from "insiders", informed by the existing knowledge base. The paper should be read in this light.

This paper was written by Ken Warwick (Warwick Economics) and Alistair Nolan (OECD Secretariat). The Committee on Industry, Innovation and Entrepreneurship (CIIE) approved this report in March 2014.

Note to Delegations: This paper is also available on OLIS under reference code DSTI/IND(2014)3/FINAL

TABLE OF CONTENTS

Boxes

EXECUTIVE SUMMARY

There has been a revival of interest in industrial policy amongst policy makers around the world in the last decade. At the same time there have been significant developments in policy evaluation methodologies, with a rich vein of work in areas such as health, education, labour markets and poverty alleviation. However, rigorous and systematic evaluation evidence in many areas of industrial policy is scarce, and methodological challenges particular to this field are many.

An OECD "Expert Group on the Evaluation of Industrial Policy" has considered recent evidence from the evaluation of industrial policy, against the background of new approaches to evaluation. Industrial policy, broadly defined, covers a multitude of policy instruments and approaches. This report focuses on:

- R&D support policies;
- Innovation-oriented public procurement;
- Capital market interventions, in particular support for risk capital;
- Sectoral approaches (including public-private partnerships);
- Cluster and regional policies;
- National industrial strategies.

Given the lack of good quality *evaluation*, the findings must be treated with caution. However, the salient policy-relevant conclusions are summarised below.

R&D support policies. A mix of incremental R&D tax incentives and targeted direct support is likely to be most cost-effective, but will increase administrative and compliance costs. Giving direct support to pre-competitive ventures, and to research partnerships, is one means of reducing problems associated with picking winners at the level of individual firms. The use of matching grants might help to identify higher-quality projects and lower total public outlays. Good policy designs need to ensure competitive and transparent selection processes, and avoid favouring incumbents or providing opportunities for lobbying.

If R&D tax credits fail to provide immediate cash refunds and/or carry-over provisions, they may provide less assistance to young firms, even if the scheme is otherwise generous. A refund for expenditure on the wages for R&D personnel is another way to support young firms. New OECD research suggests that R&D tax incentives benefit incumbents at the expense of entrants. The effectiveness of R&D tax incentives also depends on the stability of the policy over time. And overall tax relief for R&D by multinational enterprises (MNEs), when factoring in relief resulting from cross-border tax planning by MNEs, could well be greater than governments foresaw when their R&D tax incentives were designed, with a number of policy consequences.

Innovation-oriented public procurement. Policy-makers need to ensure competition in procurement procedures and adequate capacity in the public sector for developing and implementing innovation-oriented procurement. Sub-optimal scale in procurement bodies should be avoided and linkages established with innovation agencies. Specific risks associated with innovation-oriented procurement must be managed and suitable legislative frameworks developed.

Capital market interventions, in particular support for risk capital. For early stage and seed financing for young innovative companies there is some evidence of successful equity schemes, but many initiatives are relatively new and, on the whole, evidence of the impact of these programmes is not conclusive. Policy seems to have been more successful where there is opportunity for experimentation and learning and when it has leveraged other private funding. Policy-makers need to address not just the supply of finance but complementary policies on the demand side and wider framework conditions.

Sectoral approaches and public-private partnerships. The evaluation evidence on past experience with sectoral approaches is limited. There are examples of sector success stories (for instance from Brazil and Korea) but little systematic evidence that sector performance is related to the degree of support or that the explicit identification of key sectors is a necessary component of modern industrial policy. While a sector focus may help create an alignment of interests around shared objectives, this must be balanced against the risk of capture by sectoral interests and of deadweight losses in public spending. Sector initiatives often involve some form of public-private partnership (PPP). Despite the strong consensus from the new industrial policy literature on the importance of mechanisms for promoting dialogue between industry, government and the research base, evaluation evidence of effectiveness and value for money is limited. 'Soft' industrial policy interventions of this kind are by their nature harder to assess, partly because the intervention itself is qualitative and facilitative. More evaluation is called for, specifically on measures such as sector councils.

Cluster and business networks. Policy should explicitly target market failures and should be based on government working with existing and emerging clusters rather than trying to create them *ab initio*. A policy on clusters should provide a framework for dialogue and co-operation between firms, the public sector (particularly at local and regional levels of government) and non-governmental organisations. Direct financial subsidy of firms' location decisions appears to have only modest effects but could create multiple inefficiencies. Policy-makers should assess the wider determinants of cluster success – such as infrastructure – which may be better targets of policy. Local adaptability should exist in university-industry relationships. On business networks it is important to create informed demand for network services, with networks preferably addressing precise market-driven objectives. A degree of financial support, for instance in feasibility work, is to be expected. Policy must operate with realistic time-frames and experienced network brokers are essential.

National industrial strategies. Many countries have experimented with or are embarking on some form of industrial strategy and debate continues on the conditions for success and the respective merits of selective and horizontal approaches. New industrial policy puts less emphasis on product market support measures and more emphasis on support for technology and skills, public-private partnerships and a government role in facilitation and coordination. These approaches are newer and by nature some are harder to evaluate. Successful implementation of industrial strategy requires well designed plans for monitoring and evaluation.

The evaluation of industrial policy instruments could be improved by making better use of techniques employed in other areas of policy, particularly those involving experimental approaches. Some experience in this respect has already been had, but needs to be built on. To strengthen the evidence base for industrial policy this report makes a number of overarching recommendations. In particular, policy makers should:

- Make an explicit commitment to the evaluation of industrial policy.
- Insist on the development of data and evaluation strategies before programmes can begin.
- Choose evaluation techniques appropriate to the programmes concerned.
- Evaluate industrial strategy using mixed methods.
- Insist on full disclosure in evaluation reports.
- Create robust governance mechanisms to ensure evaluation is objective and free of political influence.
- Develop effective mechanisms for policy learning.

The Annex sets out principles for enhancing the strategic contribution of evaluation to policy development.

1. INTRODUCTION

Recent developments have brought renewed attention to industrial policy and its evaluation. The global economic and financial crisis, competitive pressures in industry from emerging markets, and reductions in industrial capacity in many OECD economies have all increased interest in industrial policy among OECD policy makers. This heightened interest has in turn spurred new thinking on the forms that industrial policy should take. Furthermore, these developments have occurred at a time when, in many fields of public policy, awareness has risen of the need for more sophisticated forms of policy evaluation. This awareness has been driven in part by the constraints on discretionary spending affecting many OECD governments and the consequent drive for greater cost effectiveness in public spending. For a variety of reasons that are considered in this paper, the evaluation record on industrial policy and some of its instruments is weak relative to some other fields of policy. More and better evaluation is essential. A more sophisticated, systematic and strategic use of evaluation is also necessary because the newer forms of industrial policy possess features which complicate the evaluation challenge.

The paper has a significant focus on policy conclusions. However, a recurring theme throughout this paper is that high-quality evaluation of industrial policy is in short supply. So it is worth considering why this is the case. It has been suggested that the limited use of experimental methods and rigorous control groups in the evaluation of industrial policy in comparison to other policy areas may have a number of causes:

- *Identification of controls* – the term "industrial policy" can encompass multiple policy instruments. The feasibility of identifying control groups can vary from one instrument to another. For instance, with R&D tax credits, all firms might be eligible for the policy, which complicates the identification of controls. But other instruments might lend themselves more readily to assessment using controls (as is the case for instance with some forms of enterprise support). However, assessment of the macro-economic impacts of certain industrial policies requires cross-country data covering large numbers of economic, institutional and policy variables that might also explain the outcomes of interest. Compiling such comparative data is problematic, which is one reason why debate continues over the contribution to growth and development of some of the more directive forms of industrial policy.

- *Data availability* – researchers still complain about difficulties in accessing relevant data in the industrial and innovation policy sphere. This is being eased by the greater availability of micro-data and linked business data sets but data problems remain more of a constraint than is the case for evaluators in other fields, such as health, education and labour markets.

- *Unit of analysis* – businesses are large units of analysis. The effects of interventions on firms can be difficult to capture when looking at the impact on the business as a whole. For this reason, measures that rely on the assessment of the recipient often need to be combined with other measures to capture the overall picture.

- *Interdependence* between the outcomes for economic agents. Pecuniary and non-pecuniary externalities – positive or negative – affecting economic agents operating in the same market mean that the true impact of a policy intervention may be hard to establish. In most clinical trials, the health of one patient is completely independent of the health outcome for another. This is unlikely to be the case in a trial of an economic policy instrument where the receipt of an incentive by one firm may give it a competitive advantage over similar firms in the same market. There may also be spillover effects, positive and negative, which are harder to capture, but may be central to the justification for the policy intervention in the first place.

- *Multiple objectives* – in the evaluation of industrial policy, as with other areas of evaluation, policy needs to be evaluated against its declared objectives. An industrial policy measure to stimulate employment in a region of high unemployment is unlikely to score well if evaluated against a national productivity objective. Where policy has multiple objectives measurement challenges will also arise owing to the fact that there will also be multiple influences on economic outcomes, including market, technological and other economic and policy conditions. In the inevitable absence of complete models of the determinants of microeconomic performance, it is hard to identify which changes are due to the "treatment" under consideration and which are due to the many other factors that could be playing a role.

- *Time lags and long-run impact.* Some policies, for example those aimed at improving basic research or creating a culture of entrepreneurship, may bring about gains which are important in the long term but not easily quantified over the short-term. Time lags may limit the usefulness of evaluation as a tool for policy development and also run into the problem that administrations often change before a scheme can be properly evaluated.

- *Context dependence* – it may be hard to translate the results of one policy experiment to a different context. Institutional differences across countries or changes through time may mean that an intervention that has proved successful in one setting may not be so successful in different circumstances.

In undertaking this research, it is recognised that it may not always be possible for cost/capability reasons to do a full economic impact. Nevertheless, in addition to considering economic impact, an evaluation of a programme will also need to cover a range of administrative and general appropriateness issues. These include how much it costs to run, appropriate governance for spending public money, compliance costs to applicants, and continued alignment with government policy.

The plan of the paper is as follows: Chapter 2 reviews the evaluation record for three specific areas of policy, namely: policies to support business investment in R&D (including tax incentives and direct support); innovation-oriented public procurement; and interventions aimed at supporting access to finance for small and young firms, particularly as regards risk capital. The chapter considers the rationales for policy in each area, the specific challenges of performing evaluation for each policy instrument and a synthesis of evaluation findings by instrument.

Chapter 3 addresses the evaluation of what has here been termed "policy packages". While Chapter 2 considers single policy instruments, industrial policy often comprises a mix of policy instruments, which may be targeted at one or more sectors or technologies. At the most general level, a national industrial strategy may be aimed at the economy as a whole with a range of horizontal and selective measures working in combination. When policy packages such as these are considered, the evaluation challenge is magnified. This chapter reviews the evidence from evaluations of sector approaches, public-private partnerships and cluster policies.

Chapter 4 addresses current thinking on new industrial policy and some recent developments in policy evaluation. While the content of industrial policy varies across countries, what has been termed "new industrial policy" exhibits a number of differences relative to the past. Warwick (2013) summarises these as involving:

- A greater emphasis on improving systems, building networks, improving co-ordination, promoting awareness, strategic partnerships with industry and joint consideration of the competitiveness challenges facing business;
- A shift of focus towards tasks/activities and technologies, and away from support for single firms, national champions, tariff protection and other product market-focused interventions;
- Less reliance on direct support in the form of state aids and market-failure correcting subsidies; and
- Some merging between strategic and defensive industrial policy – while policy makers have reacted to the global crisis by adopting defensive measures for certain industries, there are also examples of interventions in support of new areas of economic activity.

New thinking on policy evaluation and its application to the new forms of industrial policy are likewise examined in this chapter. It is shown that there has been a trend in the evaluation community for some time towards greater use of more rigorous techniques for evaluation, particularly using randomisation and experimental methods in creative ways. While long established in medical science and health policy in particular, such approaches have become more widespread in other policy domains, such as education and development economics/co-operation. The application of such approaches to new industrial policy has so far been limited, but positive examples exist, as described in this chapter, and need to be built upon.

Chapter 4 also outlines a possible evidence assessment framework for the evaluation of industrial policy, based on sequential hypothesis-testing. Consideration is likewise given to the role of evaluation in the overall policy cycle. Measures are suggested - and described in greater detail in the Annex – that could enhance the strategic contribution of evaluation to policy development. Ideas are also advanced on how to increase the "clock speed" of the evaluation cycle so as to enable greater responsiveness to the sometimes fast-changing needs of policy makers.

Chapter 5 concludes and draws a number of policy implications. The observations on policy are based on the material reviewed in the paper and are discussed under four headings: a) specific policy areas (those considered in Chapter 2); b) industrial policy packages (examined in Chapter 3); c) new approaches to evaluation, as part of new industrial policy; and d) ensuring that evaluation has policy impact.

2. NATIONAL EXPERIENCES WITH INDUSTRIAL POLICY EVALUATION - EVALUATING SPECIFIC POLICY AREAS

This Chapter reviews the evaluation record for three specific policy areas that often form a part of industrial policy. These are: policies to support business investment in R&D (including tax incentives and direct support); innovation-oriented public procurement; and interventions aimed at supporting access to finance for small and young firms, particularly as regards risk capital. Consideration is given to policy rationales, the challenges of performing evaluation for each instrument, and main evaluation findings.

2.1 Support for business investment in R&D

OECD countries provide a range of support measures for private investment in R&D. These include direct support for R&D – for instance through grants or direct procurement – and indirect support in the form of various sorts of R&D-related tax allowance and credit (the mix of policies used to support private investment in R&D and innovation differs widely across countries). Such support aims to address the inability of firms to fully capture the returns to their investment in R&D, which can lead to a level of private investment below a social optimum (the literature on the socially suboptimal supply of privately funded R&D is vast, and is not reprised here). Government action is also premised on system failures, such as a lack of networks in which to innovate, and incomplete markets, which may pose particular problems for entrepreneurs seeking equity finance for investments in innovation.

2.1.1 Tax incentives for R&D

R&D tax incentives aim to reduce the marginal cost to firms of performing R&D. As of 2011, 27 of the OECD's 34 members provided tax incentives to support business R&D – more than double the number in 1995 – and others are currently considering their introduction. Many non-OECD countries, such as Brazil, China, India, Singapore and South Africa, also offer generous tax incentives for investment in R&D. By 2011, over a third of all public support for business R&D in OECD countries came through tax incentives, a share that rises to more than half when the United States – with its large direct procurement of defence-related R&D – is excluded.

By comparison with other policy instruments considered in this paper, evaluations of R&D tax credits are numerous. R&D tax credits have been evaluated in many countries using diverse methods. Literature overviews consulted in preparing this paper include Hall and van Reenen (2000), Lentile and Mairesse (2009) and Köhler et al. (2012).

For the purpose of policy assessment, once the policy has been fully implemented, firms cannot legally be excluded from a tax incentive to which they are entitled. This removes the possibility of evaluating R&D tax credits using randomisation. Evaluations have therefore been based on the following approaches: surveys; quasi-natural experiments; techniques using statistically constructed control groups; and structural econometric modelling.

The evaluations using surveys suffer from the typical limitations of survey methods: entrepreneurs or managers might be unable to accurately assess the genuine impacts of the scheme, distinguishing these from many other possible determinants of R&D spending; long-run effects might be ignored, especially if the survey is administered shortly after the policy has begun; and respondents could also have strategic

reasons for overstating or understating programme impacts. Surveys have also often been based on small sample sizes.

Another set of evaluations of R&D tax credits uses econometric techniques that exploit thresholds in the administrative process of selecting firms into the scheme. If for instance there is some ceiling on the application criteria for the tax credit, then firms operating immediately above that ceiling could be used as a control group. Haegeland and Moen (2007) evaluate Norway's R&D tax credit following this approach (the evaluation concluded that the scheme, introduced in 2002, increases private spending on R&D).

The third evaluation approach statistically constructs a control group. However, even using a comprehensive set of matching criteria - such as firm turnover, age, sector of operation, geographic location, etc. - perfect comparability between the two groups is not possible. There may always be some unobserved differences between beneficiaries and non-beneficiaries that also affect the policy outcome. For instance, in all observable respects two firms might have the same probability of receiving a tax credit, but the abilities of their respective managements may differ. The managerial abilities in question might also be associated with a higher propensity to innovate. Evaluations that use matching-methods also require that there be a sufficient number of firms that qualify for the tax credit but fail to apply. Lentile and Mairesse (2009) note that such a situation might arise owing to a number of conditions: if firms are unaware of the policy; their R&D spending does not qualify for the tax credit; applying for the tax credit is excessively complex or costly; or they fear that applying for the tax credit could increase the probability of a tax audit.

A fourth approach to evaluating R&D tax credits has been structural econometric modelling, which uses models of R&D investment behaviour and assumes that R&D spending is a function of the cost to the firm of the capital used. This cost includes interest foregone on the money used, economic depreciation, gains or losses from inflation, corporate taxes and the specific fiscal rules governing amortisation of R&D, as well as the tax credit (Hall and Jorgenson, 1971). Modelling first seeks to estimate the sensitivity of the cost of capital to the R&D tax credit. A second step estimates the sensitivity of firms' R&D spending to changes in the user cost of capital. Lokshin and Mohnen (2012) use structural econometric modelling to assess the R&D tax credit in the Netherlands (*WBSO*). Their results indicate that for SMEs, each euro of the tax credit generates .2 of a euro in additional R&D. The figure is considerably lower in large firms, at around .07 of a euro.

Main findings on the effectiveness of R&D tax incentives

Most evaluations of R&D tax credits examine changes in R&D spending brought about by the policy. Most, but not all, evaluations find that responses to R&D tax credits are small initially but increase over time. The greater long-run responsiveness can reflect adjustment costs (Hall and van Reenen, 2000) and, possibly, inelastic supply of scientists and research technicians. On average, a 10% reduction in the user cost of R&D increases the volume of private R&D spending by around 1% in the short run and 10% in the long run (Bloom et al., 2002).

An incremental tax credit (which only applies to R&D expenditures above a baseline) may be more effective than a volume-based tax credit in inducing additional business R&D spending (Parsons and Phillips, 2007; Lokshin and Mohnen, 2012). However, incremental schemes that use a moving average baseline can have a smaller incentive effect owing to inherent instability and reduction of the incentive to invest in R&D as larger scale thresholds are reached (Hall and van Reenen, 2000). A volume-based scheme might also be preferred if the policy goal is to retain R&D-intensive businesses, especially in a context of high overall corporate tax rates.

The size of input additionality effects vary considerably across countries, over time and with the econometric technique used (Köhler et al, 2012). Hall and van Reenen (2000) conclude that "In the current imperfect state of knowledge we conclude that a dollar in tax credit for R&D stimulates a dollar of additional R&D." More recent OECD estimates support this conclusion (Westmore, 2013).

R&D tax credits also produce outcomes other than increased R&D. One of these is the change in the proportion of firms investing in R&D for the first time. This broader outcome is rarely assessed in policy evaluations. However, in Norway, Haegeland and Moen (2007) found that after the introduction of the R&D tax credit, firms that had not previously invested in R&D experienced a 7% increase in their probability of doing so. An evaluation presented to the OECD Expert Group (Donselaar, 2013) showed that a recent scheme in the Netherlands had induced significantly increased spending on R&D-related wages (one target of the policy): each Euro of the tax credit was associated on average with EUR 1.77 of additional spending on R&D-related wages. The evaluation also showed that the policy had led to some substitution between R&D labour costs and other R&D expenditures.

Fewer studies have focused on output additionality. However, for Norway, Cappelen et al (2012) show that tax credits give rise to new production processes and to some extent the development of products which are new to the firm. However, the tax credit does not appear to contribute to more novel new-to-the market innovations. In Austria, Falk et al (2009) found positive effects on the likelihood of introducing new-to-market products. Examining the effect of R&D tax credits on innovation activity in Canadian manufacturing firms Czarnitzki et al (2011) conclude that tax credits lead to additional innovation output. Recipients of tax credits show significantly increased scores on the frequency of new product development, the introduction of new-to-market products and the share of sales accounted for by new products. However, firm profitability and market share were not affected by the tax credit. In the Netherlands, Donselaar (2013) also found positive indirect effects on innovation and labour productivity.

While R&D is an amply demonstrated determinant of productivity growth, evidence on the impact of R&D tax incentives (and direct support) on firm-level productivity is ambiguous (Westmore, 2013). OECD (2013a) sets out why this seemingly contradictory finding might arise, including that:

- Measurement and identification issues – possibly including time lags before some productivity effects arise - might play a confounding role;
- Fiscal incentives can lead to an increase in the price of R&D (e.g. via higher wages of scientists) rather than the volume of R&D;
- Projects financed by R&D tax incentives might have lower than average marginal productivity and might not have the highest knowledge spillovers. The finding that R&D tax incentives can have a positive effect on incremental innovations but, in some cases, weak effects on new-to-the-market innovations is cited in support of this point;
- Increases in R&D might duplicate other R&D or involve a relabeling of non-R&D activities (although there is some evidence that investment relabeling might not be a sizeable problem); and,
- Firms that benefit the most from tax incentives might be those for which R&D is less likely to generate large spillovers and large increases in aggregate productivity. For instance, many small firms benefitting from a tax credit might cater to the needs of niche markets.

For their literature review, Lentile and Mairesse (2009) found only one full-blown cost benefit analysis of an R&D tax credit: a study of the scheme in Canada by Parsons and Phillips (2007). This

evaluation sought to separately quantify and then combine five effects of policy. The study shows a median increase in social welfare of 11 cents for each dollar of tax credit. After sensitivity analysis the authors conclude that the "tax credit likely generates positive net economic benefits under a reasonable range of assumptions." More recently, Lokshin and Mohnen (2012) undertook a cost–benefit analysis of the R&D tax incentive programme in the Netherlands using a panel of firm data covering the period 1996 to 2004. The authors found some evidence of additionality suggesting that the volume-based programme of R&D tax incentives is effective in stimulating firms' investment in R&D. However, the hypothesis of crowding out is only rejected for small firms. The authors' simulations also indicate that the volume-based nature of the tax incentive leads to the government supporting R&D which firms would have done without the policy. The estimated deadweight loss is as high as 85% of the total revenue loss entailed by the scheme.

While there are a large number of evaluations, comparability across studies is hazardous. In part this reflects variations in method. Problems of comparability also arise because of variation in the design of tax incentive schemes: some are based on R&D volumes and others on incremental R&D spending. The generosity of schemes also varies, and can vary within the same country for different parts of the enterprise population. The scheme in the Netherlands targets the wages of R&D workers as well as social insurance liabilities. A variety of additional eligibility criteria can also apply. These might favor co-operative research, or the hiring of young researchers. Programme objectives also vary, from increasing the overall volume of R&D spending, to encouraging firms to begin investing in R&D, to stimulating the recruitment of R&D workers and attracting/retaining direct investment by foreign firms engaged in R&D.

2.1.2 Direct support for business R&D

Direct support for business R&D through grants or loans and loan guarantees has a long history, dating in many countries to at least the immediate post World War II period. Compared to indirect support for business R&D, direct support for R&D has the advantage that it can be focused on activities and actors that are of greatest interest in meeting public policy goals and that may have the highest social returns. A case in point is environmental amelioration and pollution abatement. Many governments have made international commitments on reducing carbon emissions. The search to develop, diffuse and apply low-emission technologies is taking place within defined time-horizons. Direct support can help to focus innovation efforts aimed at meeting such timeframes in ways that generic instruments might not.

Furthermore, direct public funding might be more beneficial than R&D tax incentives for small and young firms facing constraints in access to debt or equity. The receipt of a grant (or procurement contract) might also provide a signal of project quality to third-party sources of finance. Such signalling could help reduce the information asymmetries which are at the source of some of the difficulties of access to finance (particularly debt) among small and young firms. Such information asymmetries may be more severe for radical than for incremental innovation (Czarnitzki and Hottenroot, 2011). Some evaluation evidence confirms this positive effect on access to finance (for instance Czarnitzki and Licht [2006] examining data on firms in East and West Germany, and Feldman and Kelley [2003] for the United States).

Main findings on the effectiveness of direct support for R&D

As with tax credits, the evaluation evidence on direct support focuses largely on input additionality (i.e. the extent to which the support has induced new business spending on R&D). Less attention has been given to the effects on outputs, including innovation, employment and productivity. Furthermore, in a review of the evaluation literature Cunningham et al (2013) find only two studies that have assessed impacts over significant periods of time.

Recent OECD research shows that direct support (subsidies) can induce additional business R&D (Westmore, 2013). However, as discussed in Chapter 1 of OECD (2013), this result is only found for data

after 2000. This result accords with earlier OECD research that also failed to find a significant relationship between R&D subsidies and additional business R&D between 1982 and 2001 (Jaumotte and Pain, 2005). The estimated increase in the effectiveness of R&D direct support may reflect changes in how public support is delivered: the support is now more focused on subsidies for commercial R&D activities and involves a greater use of matching grants.

Few evaluations specify the distribution of impacts [Cunningham et al (2013)]. In some cases a skewed distribution of impacts may be a function of programme design when, for example, governments support high-risk R&D. Indeed, the studies that do report the distribution of impacts generally show considerable skewness. For instance, examining a sample of Spanish manufacturers, González et al (2005) estimate that "almost half of large non performing firms could be induced to perform innovative activities by financing less than 10% of their R&D, and one out of three small nonperforming firms by financing up to 40% of their expenses". [1]

Evidence varies on the extent to which public support substitutes for private R&D. Part of this variation reflects study methodology, with substitution effects found more often in firm-level rather than industry or macro-level studies. Garcia-Quevado (2004) reviewed 74 studies and found that 38 identified complementarity between public support and private R&D (17 revealed substitution and 19 had insignificant results). The results of econometric modelling partly explain these findings, as many beneficiary firms receive public support precisely because they intend to undertake R&D. Guellec and van Pottelsberghe de la Potterie (2003) find that public and private R&D are complementary, but become substitutes after a subsidisation rate of 20%. While the evidence is inconclusive, various studies find that the additionality of direct support is greatest for small firms, firms in relatively low-technology sectors and firms in lagging regions (Cunningham et al, 2013).

Evaluations of output additionality – in terms of productivity, innovation, employment and the introduction of new products – are fewer than the evaluations of input additionality. Cunningham et al (2013) conclude that the results of the available evaluations are not clear-cut. However, output additionality is made more likely by the combination of direct support with other factors such as firm-level capabilities and the availability of other forms of support. The influence on output additionality of firm-size, location, industry and other firm characteristics is less evident than with input additionality.

Evidence on the relative effectiveness of direct support as compared with R&D tax credits is limited. For Norway, Hægeland and Moen (2007) find that an additional dollar of tax credits had a slightly larger effect on R&D than an additional dollar of direct support. But, in a cross-country setting, Westmore (2013) shows that direct support has a larger impact than volume-based tax incentives on R&D (a result that might not hold if the comparison were with incremental forms of R&D tax credit, which, as noted above, appear to have greater additionality).

The earlier observation that direct support may relieve financing constraints for young and small firms suggests that direct support and R&D tax credits could have complementary effects. Further research is needed to better examine this relationship. However, Guerzoni and Raiteri (2012) find significant positive synergies among R&D tax credits and R&D subsidies (as well as innovation-oriented public procurement). And, in Canada, Bérubé and Mohnen (2009) show that firms receiving both tax credits and direct support are more innovative than firms that only receive tax credits.[2]

2.2 Innovation-oriented public procurement

The concept of fostering industrial development and innovation through procurement is not new and some countries have pursued active technology procurement policies for many years, particularly in defence, energy and transport. A number of OECD governments have recently given renewed impetus to

using procurement to foster innovation. For instance, the United Kingdom has actively sought to integrate procurement for innovation across government departments since 2003. Finland, Germany, the Netherlands, Spain and the United States all operate programmes for innovation procurement. Legislation to support innovation procurement is in place in France.

The potential for use of innovation procurement has also been highlighted in a number of European Union reports, including the Innovation Union Communication (European Commission, 2010). Indeed, this latter communication calls on European Union Member States to create dedicated budgets for pre-commercial procurement and public procurement of innovative products and services. The proposed EU Framework Programme for Research and Innovation ("Horizon 2020") likewise points to the importance of public procurement in the uptake of innovation (European Commission, 2011). The Industrial Policy Flagship proposes the use of public procurement, among other measures, to boost innovation performance in specific sectors, such as space and bio-based products (European Commission, 2010a). And among the twelve instruments in the 2011 Single Market Act, the European Commission proposed that legislative frameworks on public procurement be made more simple and flexible, in part to foster innovation (European Commission, 2011a).

OECD (2012a) suggests that while most OECD countries are trying to use procurement for innovation in some way, few have set aside a separate budget for this purpose. However, many countries have used performance-based tender specifications to encourage innovation, provided guidance to procurement officers, or involved suppliers at an early stage in the tender process to help foster innovation.

At least four types of procurement can be distinguished. The distinctions are important because the interpretation of evaluation work can be complicated by the breadth of approaches to procurement. Indeed, there does not seem to have arisen yet a body of (comparative) evaluation work that tailors outcome metrics to – and seeks to assess the impacts of - the characteristic features of each type of procurement. The four types of procurement are:

- *General public procurement*. A number of studies using different methods try to measure the impacts on firms of procurement contracts (see below).

- *General public procurement that employs design features aimed at increasing innovation*. This might involve, for instance, a greater focus on performance metrics, rather than on specific features of the procured *items* (defined *a priori*), thereby creating more scope for innovation. New procurement criteria, beyond price, can emphasise innovative outcomes, and can be added in the tender specifications and in the assessment of tender documents. OECD survey evidence suggests that about 65% of OECD countries used such approaches in 2011 (OECD, 2012a). A variant has been used in the United Kingdom, whereby the market is given advance information on future procurement needs. An agreement is given to purchase a product or service that currently may not exist, at a specified future date, providing it can be delivered to agreed performance levels and costs.

- *Catalytic procurement.* Under this approach the government plays a role in strengthening demand for products and services still at an early stage of development or diffusion and for which there are network economies. Government purchases may provide firms with early market support and may act as a signalling device. For example, several governments have recently purchased electric cars to promote the market for such alternative fuel vehicles.

- *Pre-commercial strategic (or innovative) procurement (PCP)*. This type of procurement is aimed at purchasing research and development, design, prototyping and testing services for products or services that do not yet exist on the market. This pre-commercial procurement requires novel

technological *development* work on the part of the companies or institutions responding to the call for tender. PCP has taken on different forms in different countries. A critical feature can be the introduction of step-wise competitive phases. Such a step-wise process maintains competition throughout. It likewise allows the procurer to guide the outcome such that this most closely matches the public sector need.[3]

There are various rationales for using public procurement as a policy tool for fostering innovation:

- Because of their purchasing power, governments can shape innovation directly (and indirectly). Firms benefit because procurement can help them recuperate the sunk costs of large and sometimes risky investments. Indeed, a number of major technological innovations have their origin in public procurement, including Internet Protocol technology and the Global Positioning System. And public *procurement* has facilitated the emergence of high-tech sectors in a number of countries, including the United States, Japan and France (where public procurement has been used, for instance, to develop high-speed rail technology). Where the scale of public sector demand is significant, profitability among successful suppliers might also be increased on account of economies of scale. This in turn can prompt innovation and learning, which can lower the cost of technologies and processes, making them even more attractive for adoption by users (Romani et al, 2011).

- Furthermore, by providing a signalling or demonstration effect as an early and high-profile user, governments can also influence the diffusion of innovation and catalyse private demand. Businesses that win procurement contracts can also achieve an enhanced reputation within their sector (Binks, 2006).

- If targeted innovation is achieved, the delivery of essential public services can become more cost-effective. New products and services might enable governments to innovate to improve process efficiency and enhance the quality and availability of public service delivery. More efficient procurement might also liberate public resources that can be deployed to satisfy private needs through other channels.

- The possibility of inducing innovation through procurement that would have occurred anyway – but suitably modified - is particularly attractive in the context of fiscal constraint.

- Public procurement – irrespective of whether it is innovation-oriented – might also help to counter problems of access to finance, particularly for small firms (the provision of a market entailed in the awarding of a contract, and the fact that a public agency has evaluated information on the firm that is awarded the tender, might also serve to attract additional finance from private sources for innovative activities). Depending on their design, procurement processes might also help offset problems of bias against small firms in the public tendering market.

The evaluation challenge for innovation-oriented public procurement

Innovation-oriented procurement has been under-evaluated compared to other categories of innovation support. This reflects the technical challenges of such evaluation and the relative novelty and underdevelopment of demand-side policy. Evaluation is further complicated by the fact that procurement takes many forms, and is not usually innovation-oriented. Consequently, data are often inadequate for the assessment of innovation impacts. Few public authorities have sought to classify data on procurement so as to distinguish regular from innovation-oriented procurement.[4] There have been almost no systematic assessments of innovation-oriented public procurement, apart from some evaluation in the context of the US Defense Advanced Research Projects Agency (DARPA). Furthermore, review of the documentation on

public procurement produced by many governments suggests little overt consideration to date of how this instrument might best be assessed.

One of the conceptual problems of evaluating innovation procurement is that a number of possible outcomes are of interest. These exist among both procurers and suppliers:

- For the procurer, there are costs and (potential) benefits. The costs include any added transactions costs entailed in implementing non-standard procurement (for instance in writing tender documents in new ways, in liaising across governmental departments, in including new criteria and processes when assessing tender submissions and in engaging stakeholders in new ways). Additional financial costs might also arise, for instance in paying for or incentivising research conducted during pre-commercial procurement. Financial costs might also arise if the procured item is more expensive than the item it replaces (which might be the case if other characteristics still make the item preferable). New administrative costs may also be incurred, for instance in establishing an entity with responsibility for implementing or overseeing innovation procurement.

- Public-sector benefits stemming from innovation-oriented procurement might also take a number of forms: radically or incrementally new services might be created; existing services might be delivered at lower cost; or existing services might be delivered at the same cost but bring some other benefit (such as a lower carbon footprint). The precise form taken by any public sector benefit could also vary across technologies, implying that the metrics of success will also vary. Indirect and hard-to-measure benefits might have to be assessed in the case of catalytic procurement – whereby government procures a good or service with the goal of encouraging others in society to follow suit - notably relating to increased public or industry use of a targeted technology.

- As with other policies, it can be difficult to determine the counter-factual. In attempting to evaluate public sector benefits a problem is raised by the fact that an innovation might emerge independently of any change in procurement policy. Even in the absence of changes in procurement practice, this innovation might have been incorporated into standard procurement as the new technology eventually became widely accepted. This particular attribution problem could perhaps be addressed when evaluating innovation procurement in lower levels of government. That is, an evaluation of public-sector benefits experienced by a given subnational authority could examine whether, over a comparable timeframe, other subnational authorities in the same country had procured differently as a result of independent technical change. This information would help attribute public sector benefits to the change in procurement practice.

- Innovation-oriented procurement might also induce supply from SMEs that would not otherwise have had access to public-sector contracts. SMEs can encounter specific obstacles as suppliers to the public-sector. A perceived lack of legal and administrative resources, as well as limitations in electronic systems operated by SMEs – for instance in order processing and invoicing – have been linked to low SME participation in public procurement (Karjalainen and Kemppainen [2008]). The fact that innovation procurement might engage otherwise unengaged SMEs has beneficial implications: an increased engagement with SMEs could allow the procurer to draw on a wider range of potentially innovative ideas, especially given the important role that new and small firms play in radical innovation. Furthermore, ideas might also transfer in unexpected ways from sector to sector, a phenomenon made more likely when larger numbers of firms are brought into the procurement process (for instance, evidence from the SBIR programme in the United Kingdom has shown that a small firm working with perishable foodstuffs was able to apply aspects of its know-how in this field to the problem of reducing hospital infections).

For supplier(s), there are also costs and benefits:

- The costs are the additional outlays undertaken to bring about the innovation. These costs might not be recouped in the case of failure to secure the procurement contract. But some or all of these costs might be recovered in some other way, for instance by making use of the know-how and capacities developed in another market.

- The benefits to suppliers include possible increased access to third-party funding (see below for evidence of this outcome), access to a public sector market in the case of success (with the advantageous predictabilities that such markets can afford), and possible access to wider markets as a consequence of the know-how and capacities developed. Benefits might also accrue to firms that did not participate in or win the procurement, for instance through information spillovers, imitation and learning effects.

In assessing the benefits accruing to suppliers, a number of points could be kept in mind:

- The evaluator is asking "has this procurement brought about innovation that would not have occurred otherwise?" In some cases an answer to this question might appear straightforward, such as when the procurer buys a good or service for which it is the sole, or main, source of demand. A new solution is provided, and so it could seem evident that the solution has arisen because of the specific procurement. A case in point might be a medical device requiring some particular new functionality procured by a health service. However, suppliers may already have been working to develop similar or even identical innovations. Indeed, there is some evidence that suppliers, rather than the procuring organisation, often identify and initiate innovations (Department of Enterprise, Trade and Employment, 2009). An innovation has occurred in the successful firm, but this may represent a marginal step for the firm – something brought about with little extra internal capacity development (indeed, the procurement process is not designed to transfer research or engineering skills to suppliers). Such marginal innovation might often be the outcome when procurement processes seek incremental improvement to some standard item, as contrasted with pre-commercial procurement that requires knowledge creation.

- Accordingly, well-structured questionnaire and interview techniques appear the best way of assessing whether the procurement is associated with some additional innovative behaviour among suppliers. Indeed, there is no control group. This is because all firms in a relevant group of suppliers will have had access to the procurement process (indeed, ensuring such access is a tenet of competitive tender). So there is no way to isolate the effects of the policy on suppliers through random assignment. And because only one firm wins the tender – or a small number of firms, at most - there is little foundation for statistically inferring change in behaviour to the procurement. It cannot be excluded that the firms more likely to innovate will also be more likely to take part in the innovation-oriented procurement. This is an acute form of the selection bias problem in evaluation.

- Besides innovation outcomes, there may be other questions that evaluators wish to assess. One is whether the procurement has facilitated participation of SMEs.

Main evaluation findings for innovation-oriented public procurement

Overall, the evidence base on the efficacy of innovation-oriented procurement is weak. Most evidence is based on self-reports, rather than attempts to quantitatively assess a counterfactual. Methodologically, the most systematic assessment is that of Guerzoni and Raiteri (2012), reported below. This study, however, suffers from significant data limitations (but is illustrative of the sort of the evaluation work

needed to better gauge the policy merits of innovation-oriented procurement). No evaluation work seen during the preparation of this report has sought to identify the additional costs and benefits of innovation-oriented procurement – relative to general procurement – and to relate these to specific design features of the procurement process. More evidence is available on the effects of general public procurement, but much of this is still based on company self-reports.

A number of studies assess the impacts on firms of their success in becoming contractors for general public procurement. Aschhoff and Sofka (2009) sought to quantify the effects of general public procurement on innovation, and to compare these effects with other determinants of innovation. A survey of 1100 innovative firms in Germany was used, with effects differentiated by firm size, industry and geographic location. The survey data were self-reported and subjective, raising problems of possible response bias and accuracy. However, the methodology used was the same as that employed in the Community Innovation Survey, which has been widely pre-tested and piloted. Response characteristics were therefore relatively well understood. A comprehensive non-response analysis was also undertaken of over 4 000 firms. This showed no systematic differences between responding and non-responding firms with respect to innovation activities. This work however was not able to control for selection bias (*i.e*. that more innovative firms may also be those more likely to engage in public procurement). Public procurement was found to have a particularly marked effect on innovation in smaller firms in regions under economic stress. The effects on innovation of public procurement and knowledge spillovers from universities are similar, and the benefits of university knowledge occur almost equally in firms of all sizes.

Starzyńska and Borowicz (2012) examined public procurement in Poland using a quantitative survey of 100 awarding entities and 685 enterprises. The survey was combined with a qualitative analysis based on three case studies on a local government unit, a healthcare unit and a tertiary education institution. Seven case studies were also undertaken on enterprises experienced in public procurement and engaged in innovative activities. Participants' perceptions of the effectiveness and efficiency of the programme were examined along different dimensions, such as:

- The extent to which awarding entities use innovativeness as a criterion in the tender, and the weight that this criterion carries in the final decision.
- The ability of the public administrator to implement pro-innovation procurement.
- The degree of understanding of the innovation requirements in the tender among participating firms.
- The impact of public procurement on innovation outcomes in participating firms, also in relation to the type of innovation.

The survey findings were in some senses disappointing. Only 7% of the rewarding entities confirmed that during the previous three years their organisations had been involved in public procurement for innovative projects. Almost 45% of the surveyed firms reported that between 2007 and the end of 2010 they had not invested in innovation. However, only 5.3% of the enterprises, regardless of their size, had fulfilled or were carrying out innovative contracts. And more than 45% of responding firms could not tell whether the public procurement in which they had participated required innovations.

Edler et al. (2012) report the results of a survey of 800 companies in the United Kingdom that had been suppliers of central government, local authorities and the English National Health System in 2010. 25% of the surveyed firms attributed all of their innovations to procurement and 67% reported that public procurement had had some impact on innovation. Close to 80% of respondents indicated that innovations brought about through public procurement had helped them win other public sector contracts. 55% held

that such innovations had helped increase sales to the private sector and almost 30% indicated that procurement-enabled innovation had helped them increase sales overseas. Positive innovation effects appeared most frequent among larger firms, central government suppliers and suppliers of professional services. Half of the surveyed firms that had invested in R&D in the last 3 years reported that procurement had led to additional investment in R&D. Firms considered a number of procurement practices to be particularly innovation friendly. These included: having innovation requirements in tenders; interacting early with the procuring organisation; having tenders specified in terms of outcomes; and receiving advanced communication of future procurement needs. These practices, considered conducive by firms, were also among those that occurred least frequently. It should be underlined, however, that these findings are based purely on companies' self-reports.

With respect to the evaluation of innovation-oriented public procurement, Guerzoni and Raiteri (2012) seek to disentangle and compare the effects of innovation-oriented procurement, R&D tax credits and R&D subsidies. The effects of these instruments are assessed in terms of the size of R&D investment, using a propensity score matching approach. The data source is a survey of 5 238 companies with more than 20 employees across many sectors in the 27 member states of the European Union, plus Norway and Switzerland. One goal of the study is to examine the extent to which other studies of the efficacy of R&D subsidies and R&D tax credits may have yielded results which were driven by hidden treatments (i.e. by the fact that firms in receipt of such support were simultaneously benefitting from innovation-oriented public procurement).

The study finds a reinforcing effect between public R&D subsidies and private R&D, and firms that benefit from R&D tax credits are 10 percentage points more likely to report an increase in R&D spending than those in the control group. A positive and significant effect is also found for innovative public procurement on private R&D: 11.2 percentage points more firms report increasing their R&D in the treated group.[5] However, when assessed in isolation from the other instruments, the positive impact of R&D grants on R&D private investment ceases to be significant. And, again, when evaluated in isolation, the effect of R&D tax credits is almost halved. But, when examined in isolation from R&D tax credits and subsidies, the effect of innovative public procurement is unchanged. Positive synergies among innovative public procurement, R&D tax credits and R&D subsidies are found to be significant. More firms than in the control group – 22 percentage points more - increase their R&D investments when receiving both R&D subsidies and innovative public procurement. But only 8.4 percentage points more firms in the treated than in the non-treated subsample increase R&D spending when engaged only in innovative public procurement.

This study is an example of the sort of work which is needed to better understand the impact of innovation-oriented procurement. More research using better data is clearly needed. The Guerzoni and Raiteri (2012) study is not able to characterise the procurement systems themselves. Rather, the study infers participation in innovation-oriented procurement from firms' "yes/no" responses to the following survey question: "Did at least one of the public procurement contracts that you have won since 2006 include the possibility to sell an innovation (i.e. new or significantly improved products or services)?". Future research is needed to relate actual procurement practices – which are likely to vary from programme to programme and across countries - to company outcomes. The authors of this study were also forced to rely on firm-level self-reports of whether R&D grants or tax credits had contributed to innovation, rather than actual measures of the scale of support received.

When considering pre-commercial procurement, the most significant evaluation effort to date has addressed SBIR-type programmes. The Small Business Innovation Research (SBIR) programme was introduced in the United States in 1982 and has given rise to similar programmes in a number of OECD countries. In reviewing the available evaluation evidence Rigby (2013) notes that three key issues have arisen:

i. Problems of study comparability across programmes that have different designs. Programmes in the United States can vary significantly, Rigby notes, by government departments, in terms of the technologies addressed, the management teams responsible and the rules that apply.

ii. There is little evaluation material that uses control-group based methodologies. Reviews of the SBIR by the United States' National Research Council in 2007 and 2009 - which suggested a range of positive outcomes - employed case studies and self-reported information on programme benefits. And while some evaluation work does concern effects on firms, very little material thoroughly assesses the merits of pre-commercial procurement against other forms of procurement or against other innovation policy instruments.

iii. There has been inadequate commitment to the sorts of long-term data collection needed to assess an instrument, of this sort, which can give rise to complex effects over the long-term, with some of the most important outcomes (i.e. significant research findings) being seen in a minority of beneficiaries.

Some evaluation work has shown that that SBIR funding has led to increased growth and employment creation and a greater likelihood of attracting venture financing (Lerner, 1999). However, the assessment of this programme itself points to the critical importance of evaluation method. Considering additionality – the extent to which outcomes are achieved beyond what would have occurred anyway – an assessment by Wallsten (2000) suggested that SBIR resources had almost entirely crowded out privately-funded R&D.

Audretsch, Link and Scott (2002) examine the Department of Defense's (DoD) SBIR programme using a three-pronged evaluation approach: a broad-based statistical analysis of SBIR recipients (but without control groups); a case-based investigation of recipients regarding the impacts associated with SBIR awards; and, a case-based investigation of the social rate of return from SBIR-funded research. The authors conclude that this SBIR programme increases technological innovation and private sector commercialisation of innovations derived from federal R&D in ways that are additional. Furthermore, the spillover effects of the created commercial activity generate substantial positive net social benefits. The authors caution however that the conclusions are specific to the DoD's SBIR programme.

Bound and Puttick (2010) examine whether the United Kingdom's Small Business Research Initiative (SBRI) – initially modelled on the US SBIR - has helped to stimulate innovation. The SBRI is a model process that involves identification of a public policy problem, an open competition awarding R&D contracts to promising solutions offered by small firms, applications for further prototype development for proposals that passed the feasibility stage, and final public procurement, market commercialisation (or both). Bound and Puttick's research sought to provide qualitative insights on the SBRI's performance. The study entailed 30 interviews. The evaluation found that government departments had been able to widen the search for solutions to procurement needs.

Government and the independent evaluations of the SBIR programme in the Netherlands have been survey and interview based. This allowed the gathering of descriptive statistics on the population of firms participating in the programme (*e.g.* size), their innovation performance and their networks of collaboration as well as their perceptions of the effectiveness of the programme.

2.3 Capital market interventions (in particular for risk capital)

Since the financial crisis, there has been increasing concern from policy makers around the world about the growing financing gap for young, innovative firms. Banks are less willing to lend to start-ups, who often have little or no collateral. Venture capital firms have become more risk averse and have mostly left the seed and early stage to focus on later stage investments. As a result, governments in many OECD

countries have sought to address financing gaps and correct perceived market failures by supporting risk capital, and in particular seed and early-stage finance.[6]

Wilson and Silva (2013) discuss the significant role that financing plays in firm creation and growth and the financial barriers that entrepreneurs face as a result of market failures. For instance, information asymmetry can lead to high monitoring costs and principal-agent problems and is particularly pronounced in the seed and early stage and for young technology-based firms which have limited internal funds and no track record with which to signal their "ability" to investors. Externalities associated with the creation of a well-functioning entrepreneurial and financial ecosystem also mean there is a rationale for government intervention to play a catalytic role in the development of the financial system.

The relative newness of some programmes to support risk capital is one of the main aspects of the evaluation challenge in this area of policy. Recent initiatives have been in place for only a few years, whereas the effects of equity programmes may take many years to come through, longer than typical grant or loan guarantee programmes. Data limitations and the establishment of appropriate control groups are inherently much more difficult in the case of the new and emergent industries in which these schemes tend to operate, and there may also be constraints due to commercial sensitivity. Other problems are the need to disentangle the impact of public and private funding and the contribution of the various actors involved in the capital market; the difficulty of establishing a counterfactual; and factors such as the impact of the institutional setting, the policy mix and interaction of policies. Finally, if the intervention is targeted at improving the ecosystem for entrepreneurial finance and innovation, it will be important to take account of externalities and systemic impacts, which by their nature are hard to measure.

In Sweden**,** the Agency for Growth Policy Analysis is currently addressing some of these challenges in its work on an evaluation of the Swedish Regional Venture Capital Fund. The Fund is the Swedish response to the EU/JEREMIE initiative and the evaluation follows the Fund from its start in 2009 until 2015, using both process and impact evaluation methods. Background research has included a study of international experience with similar measures in Finland, Norway and Scotland (Swedish Agency for Growth Policy Analysis, 2011). A process evaluation and overall summary of the portfolio is to be published shortly and an impact evaluation is planned for 2015. Based on the international case studies and early work, the emerging conclusions on the evaluation of regional venture capital funds from this stage of the project are:

- Evaluation evidence in this area is often lacking – advance planning and preparation is essential. Both quantitative and qualitative approaches will be required to evaluate the impact of the programme – econometric approaches alone will not suffice.
- A long-term strategy and institutional approach is required but is rarely used. There is a need to shift from short-term policy measures focused on the supply side to more long-term, indirect and systemic venture capital strategies, such as changing incentive structures and improving regulation. The challenge is to build up and maintain an institutional structure that is stable and long term yet still permits and encourages learning and innovation.
- Context dependency is not sufficiently considered, in particular the need in designing policy to take account of factors such as history, the nature of financial markets and differences in business structure.
- The rationale for policy actions is not clear – is the intention, for example, to be an agent of regional development or a "pure" venture capital fund; is it to support specific companies or to stimulate the development of the venture capital market?
- The geographical dimension is not properly addressed – there may be a trade-off between understanding the needs of local investors and achieving critical mass.

A study from the German Federal Ministry of Economics and Technology (Ziegler, 2012) focuses on an evaluation of Germany's Central Innovation Programme for SMEs (*Zentrales Innovationsprogramm Mittelstand*, or ZIM).[7] ZIM is a technology-neutral promotion programme of the Federal Ministry of Economics and Technology to support research and innovation activities in small and medium-sized businesses in Germany. It dates from 2008 when four previous programmes to support innovation in SMEs were merged and streamlined. Funding for the programme was increased by EUR 900 million at the beginning of 2009 under Germany's Second Economic Stimulus Package following the financial crisis.

Two evaluations of the programme have been conducted. The first evaluation, which focused on the streamlining of earlier programmes and the administration of the new programme, produced broadly favourable conclusions. It found that the programme mostly met SME needs and that there had been a strong response from firms. No change in the operation of the programme was recommended. The second evaluation, which looked at the macroeconomic impact, concluded that the widening and expansion of the ZIM in early 2009 contributed to stabilising the economy and to improving the participating companies' growth potential. The initial funding was estimated to have given rise to R&D projects totalling EUR 3.7 billion, 2.8 times the initial cost of the initiative, stimulating production and employment directly in the participating enterprises, but also leading to other benefits via supply chain linkages and multiplier effects. Moreover, the R&D activities stimulated through ZIM contributed to medium and long-term growth through product and process innovations, although these longer term supply effects were not measured.

While these evaluations suggest that the streamlining of the ZIM programme and its subsequent expansion in response to the crisis were a success, questions may be raised about the wider applicability of the findings. The additionality ascribed to the programme and the supply chain and multiplier effects were undoubtedly influenced by the aftermath of the crisis and may not hold in more normal economic times. Once again, this is a problem of determining the appropriate baseline. The evaluations did not attempt to measure the longer term supply-side effects of R&D on productivity and competitiveness, which could be the more enduring impact of the scheme. Nor did the evaluation consider whether the move from a set of hitherto technology-specific and region-specific programmes to one that was more technologically neutral and open to all SMEs had a positive or negative long-run effect.

Government-supported equity investment programmes in the United Kingdom are the subject of a risk capital meta-evaluation in Cowling et al (2014). In a comparative economic assessment for the Department of Business, Innovation and Skills, the authors examine business benefits arising from the schemes, including general indicators of capacity building (for example, investment in fixed assets and new plant and equipment, and employment changes) and final outcome measures (sales growth and value added). Using data for businesses assisted through equity schemes, loan guarantees and venture capital trusts, and a comparison group of non-assisted businesses, the researchers' main findings are that:

- Even with modest assumptions, all the schemes examined, with the exception of Regional Venture Capital Funds (RVCFs), were found to have a net benefit to the economy within a few years of investment.

- It is likely that additional benefits lasting beyond this time period will occur, although there is evidence of some erosion of benefits over time.

- There has been substantial policy learning over the last decade, through experimentation with different delivery models and institutional innovation such as the creation of Capital for Enterprise Limited (CfEL) as a fund management company to embed learning.

- Non-economic constraints such as geographic selection have important and negative effects on final fund investment performance, in particular for RVCFs.

- Escalators are important – i.e. sufficient money per individual portfolio company should be allocated in order to assist the financing of early growth stages beyond start-up.

- The full effects of equity finance schemes tend to occur, and play out, over longer time horizons than grant programmes.

- Programme architects gave insufficient attention to *ex ante* consideration of methodology and data gathering issues, making *ex post* evaluation more difficult.

In Estonia, the Government's venture capital fund, the Estonian Development Fund (EDF), invested EUR 9 million in a portfolio of 18 companies between 2008 and 2012. An evaluation by the Estonian Ministry of Economic Affairs and Communications (2013) adopted a multi-method approach using descriptive business statistics, semi-structured interviews and online network analysis.

EDF assess the value of their portfolio in 2013 as EUR 11 million. Since the investments are made in early-stage companies whose value depends on future financing and on the potential for exit, the evaluation concluded that it is impossible to confirm the accuracy of this assessment. One EDF investment in a telecom equipment maker has already proved successful as the firm was sold to a Nasdaq-listed company. However, three investments have failed, resulting in losses totalling EUR 1 million and four are judged to have a high probability of failure. The other ten investments are at an early stage where it is not yet possible to judge their true value.

The evaluation considered other measures of performance, such as sales revenue, export revenue, profit, number of employees and wages. Statistics from these sources were complemented by network analysis and semi-structured interviews which highlighted the importance of the EDF role in the network of venture capital ecosystem of Estonia. The evaluation concluded that the EDF has contributed significantly to the development of the ecosystem for venture capital and innovation in Estonia but noted that:

- There is still some way to go, as the interviews revealed that several companies are eager to leave Estonia for a more conducive environment elsewhere;

- There may be a trade-off between securing and demonstrating a return on the portfolio and the creation of positive externalities for the local economy; and

- The lack of openness about losses and exits, as well as other data difficulties for the EDF portfolio companies, though understandable for reasons of commercial sensitivity, works against principles of accountability and transparency in government funding.

The OECD secretariat recently concluded a cross-country research project to investigate the role of public support to promote seed and early-stage financing (Wilson and Silva, 2013). This subject is of particular interest because of the dramatic growth in the number of such programmes in the aftermath of the 2008-09 economic and financial crisis. Figures for European private equity and venture capital investment suggest that the share of government agencies jumped from 9.9% in 2007 to 56.9% in 2011. Within the total, the importance of hybrid public-private models, such as co-investment and participation in Funds of Funds, has shown a significant increase.

Despite the growth of this form of financing, there is a lack of systematic evaluation of government programmes in this area. According to responses provided by member countries to a questionnaire for the project, only 13 out of 32 OECD countries have evaluated their seed and early stage tax incentives and/or equity instruments. By contrast, grants, loans and guarantee schemes, which in many cases have been in place longer, have been evaluated in 21 OECD countries. Moreover, many of the evaluations of public support for seed and early-stage financing are qualitative process evaluations rather than quantitative impact evaluations, still less full economic assessments. The majority of evaluations have been conducted by outside experts and academics. Tax and equity instruments are also under-evaluated relative to debt instruments.

The main findings from the OECD project, in terms of policy indications and lessons for future evaluation, may be summarised as follows:

- Given the data and methodological challenges of evaluations of seed and early-stage financing interventions, it is critical to have a well-defined evaluation strategy *ex ante*, including a well-defined policy objective and an idea of the policy questions and evaluation design at the beginning of the process.

- Evaluations need to consider both supply-side and demand-side factors, as well as framework conditions. Policies often focus on the supply side when some of the key barriers are on the demand side. Further attention to the demand side may be warranted, including a focus on the different models of public and private programmes in these areas.

- Given the increasing reliance on public sector funding in the seed and early stage market, more emphasis should be put on initiatives to attract institutional investors as well as on various equity risk-sharing arrangements between public and private investors.

- There is also interest in exploring firm involvement in the financing system and which types of support they use. Often the same companies receive support at multiple stages so it could turn out that not as many companies are really supported as policy makers might think. Analysing the implications of firms benefiting from multiple support schemes could be useful.

- Data are a key factor in successful evaluations but also a major challenge and cost. The approach should be proportionate to programme size, and data needs to be tracked and harmonised and a combination of different statistical methods used. Ideally, indicators should allow cross-country comparisons not only of the supply-side, but of the full policy mix, including demand and framework conditions.

Notwithstanding the evidence from the United Kingdom suggesting there have been some successful equity schemes, the OECD survey suggests that the evaluation evidence to date for these programmes is not conclusive. Policy seems to have been more successful where there is an opportunity for experimentation and learning, as was seen for the United Kingdom. Policy was also found to be more effective where it leveraged other private funding and this is evidenced in the trend towards hybrid models. Another policy lesson is that policy makers need to address not just the supply of finance, which is the focus of most of the interventions, but also the demand side and wider framework conditions, suggesting a mix of complementary policies is required. Government must strive to avoid crowding out effects and structure instruments effectively to address the specific policy goals. Finally, although there is scope for learning from other OECD countries, policy makers should be wary of attempting to transplant schemes from one country to another without understanding the differences in institutional context and other framework conditions.

3. NATIONAL EXPERIENCE WITH INDUSTRIAL POLICY EVALUATION – EVALUATING POLICY PACKAGES

The previous chapter focused on distinct policy areas and generally covered evaluations of discrete single instrument interventions. In practice, however, industrial policy generally comprises a mix of policy instruments, which may be targeted at one or more sectors or technologies. At the most general level, a national industrial strategy may be aimed at the economy as a whole with a range of horizontal and selective measures working in combination. When policy packages such as these are considered, the evaluation challenges for industrial policy take on a new dimension. Accordingly, this chapter reviews evidence from evaluations of sector approaches, public-private partnerships, cluster policies and national industrial strategy.

3.1 Sectoral approaches and public-private partnerships

A "sectoral approach" to industrial policy can be considered, in its most general interpretation, as any industrial policy intervention or set of interventions that is focused on one or more sectors, technologies, tasks or activities.[8] In practice, sectoral approaches can take many forms, ranging from a single sector-specific instrument, through multiple instruments aimed at improving the performance of a single sector, to multiple instruments aimed at multiple sectors. Sectoral approaches may involve fiscal instruments, access to finance measures, business advice, coordination mechanisms and the like. They may also involve different institutional mechanisms, including for example public-private partnerships, competence centres, sector councils, industry task forces and other forms of government-industry collaboration. At their most general level, sector-focused initiatives may be virtually indistinguishable from regional or national competitiveness or growth strategies.

With sectoral approaches, the evaluation challenge is made more difficult by the fact that the whole sector is the target of the intervention. This makes it harder for the evaluator to identify comparable treated and untreated firms, since the option of comparing treated firms with similar untreated firms in the same sector may not be available.[9] There may also be effects on other sectors that need to be taken into account, particularly where the existence of knowledge spillovers or other externalities is part of the motivation for the intervention. When a package of measures is involved, the possibility of systems effects and dynamic interactions also need to be taken into account. It may also be hard to identify the impact of any one policy instrument when multiple instruments are being used. Evaluators need to think about the overall framework for evaluating a package of sectoral measures, which will be easier if prior consideration has been given to the logic model, or theory of change, underlying the intervention. However, this degree of pre-planning and coordination may not always take place.

A simple example of a sectoral intervention from the recent past would be ***car scrappage schemes*** (for example the Car Allowance Rebate System (CARS) in the US, commonly known as "Cash for clunkers"). Under these schemes, fiscal incentives were offered to car-owners to trade in older cars with high fuel emissions for newer models that would be more fuel-efficient. The motivation was partly environmental and partly to provide a boost to demand at a time when the global economy was in recession with the automotive industry particularly severely affected.

A number of attempts have been made to evaluate the impact of car scrappage schemes, particularly in the United States (Mian and Sufi, 2012; Copeland and Kahn, 2013; Gayer and Parker, 2013). The

findings of these evaluations, some of which were able to exploit differences between US states to establish a counterfactual, were generally that any impact on aggregate demand and GDP was modest and mainly temporary, as car purchases were brought forward, and the environmental impact, though positive, was secured at a cost that was much higher per tonne of CO_2 saved than other carbon abatement schemes. From an industrial policy point of view, however, these studies usually contain little evaluation of the supply-side impact on the car industry.

Evaluation of the overall impact on the *automotive industry* is further complicated by the fact that there were other interventions to support the industry, in particular – in the case of the United States – the bail-out and Government-managed restructuring of Chrysler and General Motors (GM) in 2009. Klier and Rubinstein (2013) consider the numerous evaluations spawned by the US policy actions, pointing out that, while the White House evaluation was positive, Congressional evaluations have been mixed. On the one hand, there is evidence that both GM and Chrysler were more financially viable than in December 2008 and the Detroit Three all saw a return to profitability in 2011. On the other hand, it seems likely that, at least in narrow fiscal accounting terms, taxpayers will not see a return on their investment and there are some concerns about the sustainability of improvements in the Big Three's market share.

All of these evaluations are partial and descriptive, with none attempting a rigorous evaluation of the full package of support given to the automotive industry from an industrial policy point of view. Such an evaluation would doubtless be difficult, because of the problems in establishing a counterfactual, and any lessons to be drawn may be difficult to generalise beyond the particular circumstances of the 2008-09 economic and financial crisis. It is nevertheless disappointing that economists and evaluators have not done more to assess the overall impact of the response to the crisis on the automotive industry.

An earlier example of a sectoral evaluation is the study undertaken by Lechevalier et al (2010) of the public programmes aimed at supporting the emergence of *next generation robots* in Japan from the early 1990s[10]. The focus was on support for public sector-led R&D consortia. Using indicators of the quality of patents, which enabled the estimation of a quality-adjusted research production function, the study found that participation in government programmes had a positive impact on the research productivity of participating firms. Moreover, participation in government-sponsored consortia had a greater impact on research productivity than participation in collaborative R&D among firms. The researchers suggest there may be a particular role for government involvement in R&D consortia as a coordinator of R&D collaboration, perhaps because governments bear the coordination costs, helping to realise larger scale R&D collaboration that cannot be achieved by markets.

Lechevalier also noted the importance of institutional design to promote trust among participants, particularly in a situation where some of the programmes have the characteristics of a strategic bet and *ex ante* priorities can only be established when the process has advanced and uncertainty is reduced. In these circumstances, it may be necessary to look beyond specific indicators such as patents to more systemic effects. A question also arises as to whether the policy achieved its ultimate objective, namely the emergence of a new industry. Although a success from a technological viewpoint, the policy has so far failed to stimulate the emergence of a new industry, other than a few niche applications. One lesson, therefore, from the evaluation is that in order to achieve the policy objective, it might be necessary to look beyond a single instrument such as support for R&D consortia to a broader package of demand and supply side measures designed to influence the system as a whole. A reason sometimes advanced for the lack of success of industrial policy in the United Kingdom in the 1970s is that it was too focused on engineering and technology, without consideration for the market. Understanding the objective of the policy intervention, the nature of the system, and what specific interventions can and cannot be expected to influence, is clearly important.

As the robot case study illustrates, sectoral interventions are sometimes channelled through *Public Private Partnerships* (PPPs). PPPs can take many forms ranging from their use in public procurement of goods and services as a way of securing value for money for government (usually in the financial sense rather than as a tool of industrial policy)[11], through collaborations in R&D consortia[12], industry councils and other forms of sectoral PPPs, to national-level strategic public-private alliances.[13] PPPs for science, technology and industry are taking on greater importance in the policy mix for research and innovation in many countries.

Some work under way at the OECD is focusing on strategic PPPs in science, technology and innovation. For the purposes of this work a strategic public/private partnership is defined (OECD, 2013b) as:

> "a strategic cooperative and contractual agreement between three or more private and public stakeholders, involving substantial long-term public and private investment commitments in high-risk projects that revolve around broad, emerging scientific and technological fields which are primarily initiated by the government and aligned with industrial and innovation strategies"

On this definition, initiatives such as the Top Sectors approach in the Netherlands and the industrial strategy in the United Kingdom would involve strategic PPPs. Thus there is clearly a close link between this work and the evaluation of industrial policy.

OECD (2012b) found that evaluations of PPPs to date point to contradictory results regarding their effectiveness and value-for-money and cautioned that, despite continuing political popularity, greater care is needed to strengthen future evaluations. Existing literature shows that ex-post evaluations of PPP are mostly lacking or forthcoming. In addition, there are few *ex ante* assessments to provide a basis for monitoring and later evaluation. Selected PPPs are monitored and evaluated at either the programme conclusion or during the programme's life-cycle. However, this type of evaluation concentrates on clarifying whether the identified goals/objectives have been achieved rather than the economic impact of the PPP as such.

In a later scoping paper (OECD, 2013b), drawing on a number of case studies in the EU, China, Israel, Japan and Spain, and an earlier review (OECD, 2005), the difficulties of evaluation in this area were further stressed:

- Costs and benefits of partnerships being inherently hard to measure as they imply both direct and indirect benefits as well as opportunity costs;
- Difficulty in establishing the counterfactual in the absence of benchmark or control groups;
- The existence of multiple stakeholders and multiple levels of evaluation;
- Varying timelines for outputs, some short (3-5 years) and others longer term (10- 20 years); and
- The interaction with other policy instruments such as grants and R&D tax credits.

The evaluation challenge for PPPs is therefore similar to that for other forms of sectoral intervention – in particular the existence of multiple policy impacts, the absence of a counterfactual and the problem of selection bias (policy makers choose the sectors they expect to perform best). With PPPs, there is also the need to be clear whether it is the institutional arrangement that is being evaluated or the associated policy measures. PPPs focusing on "soft" forms of sector coordination may be especially difficult to evaluate.

One example of an evaluation of a sector approach in Sweden involving PPPs is a study (Stern et al, 2013) of the impacts of the Competence Centres programme, devised as a funding instrument for collaborative R&D and innovation. The centres are typically located on a university campus and involve a consortium of companies working together with researchers from more than one academic department. Compared against other government-funded R&D support, Competence Centres (CCs) are longer term and have higher rates of subsidy, to encourage more fundamental research and to promote behavioural additionality. In Sweden, the programme ran from 1995-2007 and from more than 300 applications, 28 centre consortia were selected to receive 10 years of funding. All programmes included government support, conditional on at least equal contributions from industry, including in-kind contributions, which were the predominant form.

The evaluation was undertaken on the basis of evidence from five sources: desk research (including a review of international experience with competence centres and past evaluations); interviews with centre managers and other university representatives; interviews with company representatives; statistical databases on companies; and a survey sent to PhD holders who had graduated in the programme.

One general finding of the study was to confirm the long period of time that can be needed for the results of research to be felt at large scale. Also, in line with modern thinking about industrial policy, the evaluation concluded there was a need for careful programming of research funding in a way that is neither wholly bottom-up nor top-down but a mixture of the two, informed by stakeholders' knowledge and interests. In terms of the industrial impact, the study identified the following main effects:

- Direct impacts on industry, through generating directly usable outputs, including new and improved products, services and processes;
- Direct impacts through behavioural additionality, including more open innovation and the creation of knowledge networks;
- Economic impacts on participating firms, through cost savings and increased revenue and profits;
- Improvements in the economic performance of individual small and medium-sized enterprises (SMEs) participating in CCs;
- Indirect effects through adding to the firms' stock of human capital and research capabilities;
- Spillovers from participants to other knowledge users; and
- Indirect effects on companies via the university system, for example access to a pool of graduates with relevant skills.

Although the evaluation found that the Competence Centre initiative was successful and should continue, the findings could not be considered entirely robust, as there was no adjustment made for selection bias and no comparison with a control group or reference scenario. Following the change of government in 2006, Sweden has pursued a policy more focused on framework conditions and generic support for research and development. In recent years, the Swedish Government has strengthened and renewed the Swedish industrial research and technology organisations (RTOs), whose role is to act as a public/private partnership providing a bridge between academic and business-oriented research. State support for the RTOs, amounting to almost SEK 500 million in 2013, is devolved to a state holding-company, Research Institutes of Sweden AB (RISE AB), comprising a network of 16 Swedish industrial RTOs and their subsidiaries.

The Swedish Agency for Growth Policy Analysis is charged with tracking and measuring the impacts of RISE projects and has proposed an *ex post* assessment system (Hjalmarsson, 2014) based on:

- Citations, patents and IP rights generated.
- Follow-up survey of the RTOs' customers (customer satisfaction survey).
- Survey approaches.
- Benefit-cost analysis, including the wider economic benefits sought by the government in terms of growth and sustainable development.

In implementing this multi-method system of assessment, the Agency will use triangulation across the different approaches and will look in particular to ensure that a balance is struck between the research-focused outcomes sought by universities and the more commercially focused interests of business. Again, this focus on what can be achieved in terms of higher growth and economic welfare by working at the interface between government and the market is the essence of the new industrial policy.

The Netherlands is another country that has made frequent use of strategic public/private partnerships, although it has done so with a greater degree of sectoral selectivity than has Sweden and the evolution of policy over time has been somewhat different. A precursor to the current Top Sectors approach, the Programmatic Approach (PA) to Dutch Innovation Policy 2006-10 was the subject of a (qualitative) cost-benefit analysis in 2012 by Dialogic (Den Hertog, 2014). Dialogic conducted the final overall evaluation of the operations and effectiveness of the approach. The evaluation concluded that the programme had been effective in promoting PPPs in the ten preferred sectors, boosted investment in R&D by "at least several hundred million euros" and encouraged skills development. But, in the absence of experimental or quasi-experimental evidence and an appropriate control group, the evaluators found it impossible to judge whether the gains in terms of productivity and competitiveness were sufficient to justify the EUR 900 million of subsidy and EUR 100 million of administrative cost expended on the programme.

Although not the main focus of the study, the evaluation also raised questions about the selection of preferred sectors, noting that the number of sectors chosen was relatively large for a small country, and that there was a bias towards established sectors and those that drew heavily on the science base rather than those closer to the market. The risk of capture by lobby groups and the exclusion of dynamic small firms and new entrants were also noted as a danger.

Since 2010, the focus has changed in the Netherlands with the emphasis now on the Top Sectors approach, which is generally seen as a leading example of modern industrial policy. The initiative focuses on nine sectors in which the Netherlands occupies a leading position worldwide. As with the previous Programmatic Approach, a key feature is the establishment of a public/private partnership providing the basis for close collaboration between businesses, knowledge institutes and the government (known as the "golden triangle" or "triple helix"). A team has been put together for each sector, consisting of a sector expert, an innovative SME entrepreneur, a high-ranking civil servant, and a representative from a research institution. To strengthen cooperation in R&D between the private sector and public sector knowledge institutions, Innovation Contracts have been drawn up for the nine top sectors. In the course of 2012, the parties collaborating in the top sectors established 19 Top consortia for Knowledge and Innovation (TKIs) and the TKIs have started to implement the research agendas in the Innovation Contracts.

The Dutch are planning to evaluate the Top Sectors initiative and a possible framework for an evaluation, scheduled for 2015, has been drawn up in the report of the Impact Evaluation Expert Working Group (2012). Table 1 shows the hypotheses that the report suggests should be tested and the indicators and data sources that it suggests could be used. The report suggests that the evaluation might proceed by comparing TKI companies with companies in the same economic sector that do not participate in TKIs, or by comparing TKI companies participating and not participating in projects. The same could be done for

knowledge institutes, using patents or publications as the key measure, though these tend to be available only with a lag. Survey approaches could also be used to test support for the hypotheses. The proposal is therefore for a structured hypothesis testing approach using a mix of evaluation methods.

It should be noted that the framework set out in Table 1 focuses very much on R&D and technology. However, a key difference between the Top Sector approach and the earlier Innovation Dialogue initiative is that the scope of the Top Sector approach runs wider than innovation to include, for example, export promotion, regulatory reform and human capital development. A complete evaluation of the Top Sectors approach should ideally cover these aspects as well, although it is recognised that doing so would add considerably to the complexity of the task.

Why is evidence on the impact of sector targeting more elusive? One reason is that sector targeting is usually part of a wider industrial policy which might include horizontal measures or other forms of selective policy such as technology targeting. It is rare for evaluations of industrial policy to attempt to identify the value added by sectoral targeting *per se*. Moreover, even if there were clear conclusions about the value added by sectoral targeting, the findings would be more than usually context-dependent. A successful case of sector targeting in the presence of technology targeting and sound framework conditions might not read across to a situation where the institutional setting was different and the authorities were relying on sector targeting alone.

Table 1. List of indicators for TKI evaluation design

Hypotheses	Indicators		Data sources
	Dependent	**Independent**	
1) A TKI increases public-private-partnership in the knowledge chain	• More PPP publications • More PPP patents	• Indicator TKI Yes/No	• *TKI* • *PATSTAT* • *CWTS* • *NL Agency/Netherlands Patent Office*
2) A TKI improves the match between public R&D investments and the innovation efforts in the top sectors	• No quantitative indicators available	• No quantitative indicators available	• *Survey*
3) A TKI leads to more private R&D	• Private R&D investment	• Indicator TKI Yes/No	• *TKI* • *NL Agency/WBSO* • *PATSTAT*
4) The channelling of public research funds through a TKI leads to more commercial and social application of research output	• More patents for knowledge institutes	• Indicator TKI Yes/No	• *TKI* • *NL Agency/Netherlands Patent Office* • *PATSTAT*
5) The TKI supplement encourages private R&D investment in the TKIs	• Private R&D investment	• Indicator TKI Yes/No	• *TKI* • *NL Agency/WBSO* • *Statistics Netherlands*
Control variables *[not exhaustive]*	• Company characteristics • Characteristics of knowledge institutes		• *PS/Company Finance Statistics* • *CIS*

Source: Impact Evaluation Expert Working Group (2012).

Another set of reasons why the evidence on sector targeting is patchy is that sector targeting on its own is not a well-defined policy. If a sector is identified for support, the instruments used need to be specified and the choice can range across the whole spectrum from traditional state aids, through factor subsidies to support for public-private partnerships or a light touch coordination and facilitation role.

Given the patchy evidence and the methodological limits to gauging the efficacy of sectoral targeting, the conclusions that can be drawn about sectoral targeting are not very different from the general conclusions that can be drawn about industrial policy. They may be summarised as:

- There are undoubtedly many successful examples of explicit sectoral targeting. The rise to prominence of Korean electronics firms such as Samsung and LG in the 1990s can be traced to support provided in earlier decades. Rodrik (2004) pointed out that there are examples of successful targeting not just in East Asia but Latin America as well and commented that "it is not true that there is a shortage of evidence on the benefits of industrial policy. To the contrary, it is difficult to come up with real winners in the developing world that are not a product of industrial policies of some sort".

- However, it is more difficult to find systematic evidence of a relationship between sectoral support and sectoral performance. Studies by Beason and Weinstein (1996) and Lee (1997) failed to find a clear association between sectoral support and total factor productivity growth in Japan and Korea respectively during the years of rapid catch-up and growth.

- New industrial policy puts less emphasis on product market support measures and more emphasis on support for technology and skills, public-private partnerships and a government role in facilitation and coordination. These approaches are newer and by nature harder to evaluate.

- The targeting of key sectors will help policy makers prioritise their efforts and also help create clarity of vision and an alignment of interests around shared objectives.

- With this sort of approach, however, there are dangers of capture by sectoral interests or of backing losers. These risks can be mitigated by having mechanisms in place to permit the withdrawal of support at the appropriate stage and to ensure accountability and openness.

Public/private partnerships are very much in line with the consensus view from the new industrial policy literature about the importance of dialogue between government, industry, and the research base. Apart from some indirect evidence from the business networks literature, it has so far proved difficult to find hard evaluation evidence on PPPs. The absence of econometric evidence for the effectiveness of PPPs does not necessarily invalidate conclusions drawn on the basis of more qualitative and subjective assessments. But this is an area where more evidence is needed, for example on the experience of some countries with sector councils.

3.2 Cluster policies and business networks

"Clustering" is the tendency of firms in related lines of business to concentrate geographically. In recent years, policy initiatives to foster enterprise clusters have been common in both OECD and developing economies, in wealthy and lagging sub-national regions, and in jurisdictions with *laissez-faire* and *dirigiste* approaches to economic development. Inspired by the idea of replicating well-known clusters such as Silicon Valley or the Italian industrial districts, policy makers have devoted significant public resources to cluster policies. For example, in France, between 2006 and 2008, EUR 1.5 billion was spent on "competitiveness clusters" (Martin, Mayer and Mayneris [2010]).

Despite the popularity of cluster initiatives there is a shortage of robust policy evaluation. Much of the available material is based on self-reports among target firms, or examines process and management issues that have little bearing on the economic merits of policy. Furthermore, as described below in greater detail, there sometimes appears to be a confusion between the extensive literature which demonstrates the benefits to firms of cluster membership and the separate question of the rationale for, and effects of, government

policy. A literature review by Uyarra and Ramlogan (2012) also highlights the lack of methodologically robust evaluation

The shortage of useful evidence may have a number of causes. One may be that policy towards enterprise clusters often encompasses many different types of intervention, some with multiple and interrelated objectives. In addition, the theoretical basis of cluster policy is in fact much weaker than it is often thought to be (as discussed in Duranton, 2011) and is imprecisely articulated in some of the literature which has inspired policy. In such circumstances it is inevitable that evaluation will be problematic. As examined in a multi-disciplinary literature dating back to Alfred Marshall's *Principles of Economics*, the agglomeration of firms and their suppliers can confer competitive advantage to the enterprises involved.[14] Agglomeration, or clustering, can permit locally concentrated labour markets, specialisation in production and the attraction of specialised buyers and sellers. Such concentration and specialisation can bring a range of (self-reinforcing) benefits, including: more efficient sharing of infrastructures and facilities; a greater division of labour between firms (offering greater scale economies for individual firms); and more efficient matching between all economic agents (employees and employers, firms looking for partners, buyers and sellers. For example, by operating in close proximity firms can also more easily subcontract to competitors those orders that exceed their own capacities - because proximity allows greater knowledge of the capabilities of potential contractors – which may allow firms to retain valued customers. The clustering of firms can likewise facilitate the flow of ideas and information. Such flows occur formally and informally, for example when employees change employer, through contacts with common suppliers, and through social exchanges.[15] Some observers also hold that locally overlapping commercial and social institutions can create a social tissue that facilitates the reduction of transaction and other business costs. In the United Kingdom and in Italy the location of firms in clusters of innovating enterprises has been found to positively affect the probability of innovating for cluster newcomers (Beaudry and Breschi, 2000).

Intimately linked to the subject of clusters is the theme of business networks. Business networks operate with varied forms and objectives. Some aim at general sharing of information, while others tackle more specific goals. Networks can allow rapid learning, and small companies often favour the peer-based learning that networks permit. Networks can help firms to reconfigure relationships with suppliers. In some instances networks have led to a new division of labour in a group of firms, allowing individual companies to reap economies of scale and scope. The benefits of collective action in many areas of business – from quality assurance to training to joint marketing initiatives to joint purchases of inputs to production - have all been enabled by networks. Distinctions between clusters and networks are important.[16] Both types of initiative can entail different resource requirements, objectives and - consequently – different evaluation metrics. Business network programmes might be easier to design and implement among co-located – or clustered – firms, but such programmes need not exactly coincide with the geographic boundaries of a cluster.

What are the characteristics of commonly pursued cluster policies? Such policies often have marked similarities. Frequently, networking or promotion of public-private partnerships is at the centre of a cluster programme. Many programmes concentrate on small and medium sized enterprises. Others provide generic information on business and economic trends as well as cluster-specific information on parameters such as markets, technologies and competitors. The specific infrastructure and training requirements of a cluster are a common focus. Governments sometimes also provide business services ranging in sophistication from basic research to advice on bookkeeping. Schemes sometimes seek to attract direct investment to a cluster. Initiatives often also seek to assess and upgrade an array of general public policies and programmes that affect a cluster.

While there are recurrent features to many cluster programmes, there are also important differences. For example, differences exist in the level of government involved. The public sponsors of cluster strategies have included local, regional, national and even supranational governments. Programmes differ

as to whether they focus on developing the existing economic base, attracting firms into the cluster or a combination of the two. Cluster programmes also differ in the process of cluster selection, with some programmes using detailed criteria – such as industry growth rate, multiplier effects, job creation and income potential, the match with local resources, relationships with local suppliers, etc. However, many other programmes use little rigour in the choice of target cluster. In fact, programmes frequently exhibit marked similarities in the selection of the sectors assisted, often being those in so-called "sunrise" industries such as biotechnology, new materials, information technology and others.

A number of observations are relevant regarding the economic logic of clusters policy and therefore, also, the evaluation of their impacts:

- It is not the case that economic benefit for firms always arises from clustering. For example, successful clustering might increase land prices or lead to congestion effects, especially for firms located in clusters belonging to industries other than their own.[17] This raises the possibility that under some circumstances a policy to increase the size of clusters could become self-defeating. Furthermore, because of limited micro-level data, policy makers typically will not know when policy has started to create more costs than benefits.

- The available evidence suggests that while clustering can bring benefits to firms, the magnitude of these benefits is modest, especially when viewed in relation to changes in cluster size. From their survey of the literature on agglomeration Rosenthal and Strange (2004) conclude that doubling the size of a cluster yields productivity gains of between 3% and 8%.[18]

- Micro-level evidence from France shows that firms are able to appropriate some of the productivity gains from belonging to a cluster (Martin et al, 2011). To the extent that firms capture these gains, the need for policy to alter firms' location behaviour is reduced. Indeed, Martin et al (2011) show that in France the size of existing clusters is not very different from the size that would maximise short-run productivity benefits.

- Some of the productivity or other benefits of cluster membership might not be caused by the cluster, but might reflect a process whereby higher-productivity firms choose to locate in clusters (in part because their higher productivity allows them to locate in attractive places where factor costs, such as land values, are also higher).

- Related to the preceding point, few cluster initiatives are explicit with respect to the market failures they intend to address. The mere fact that firms in a cluster might be more productive than firms elsewhere is not an economic justification for policy support (as it is a state of affairs that can be consistent with efficient markets). Policy needs to start from the identification of market failure(s) that could merit correction.

- When stated rationales for cluster policy do refer to market failures, they usually concern market failures which equally affect firms not located in clusters. For instance, cluster initiatives commonly aim to enhance access to finance for SMEs. But the market failures which purportedly limit access to finance for SMEs are not more present in clusters than elsewhere. Rather, they are symptoms of specific market characteristics (such as minimum efficient deal sizes in the case of formal equity finance) which are unrelated to clusters. It may be that the cost-effectiveness of some measures to address market failure could be greater when addressed to a clustered rather than a dispersed group of firms (although there seems to be no evidence in this connection). However, there are perhaps no frequently offered forms of support for clustered firms from which there is a basis for excluding firms outside clusters.

- The goals of cluster policy can appear sound from a local standpoint, but be economically unsound from a national perspective. For instance, wasteful competition is evident when – as surveys have shown to occur – different subnational bodies pursue policies to develop clusters in the same sector or industry (typically, as noted above, these are "sunrise" industries such as biotechnology, new materials and information technology).[19] Similarly, a high degree of labour mobility may be good for the national economy, but a hindrance to local efforts to maintain or develop a cluster.

- Underlying programmes of cluster development is the idea that firms and industries are part of larger inter-linked economic and social systems involving market and non-market exchanges. In fact, many well-known clusters have long historical roots: the origins of some of the Italian districts for instance can be traced back to the medieval guilds. It is difficult to imagine that governments possess the competencies with which to create and manage such complex systems through policy. Indeed, there is almost no evidence of a significant cluster that has been created *ab initio* by intentional policy (this observation alone also creates doubts about the efficacy of policies that support existing clusters but which have not been evaluated in robust ways).

Main evaluation findings on cluster policies and business networks

The paucity of high-quality evaluation evidence on cluster policies has been noted above. There also appears to be little evidence on how specific design and implementation features of cluster policy have contributed to policy outcomes. Various evaluations using statistical controls are cited below. A salient finding is that policy effects for the clusters programmes are often modest, and that assessments of long-term impacts are almost entirely lacking. Network-oriented schemes appear to give more positive outcomes.

Martin et al (2011) provide a quantitative evaluation of a cluster policy that makes use of firm-level data (this appears to be the first evaluation of a cluster policy using such data). Aimed at "Local Productive Systems" (LPS) in France, the policy sought to support inter-firm co-operation and firm competitiveness. Firm-level data for the period 1996-2004 were used to assess impacts in terms of Total Factor Productivity (TFP), employment and exports. Difference-in-difference, triple differences and propensity-score matching techniques were employed. A key finding was that the policy's main beneficiaries were firms in declining industries in low-income regions. Indeed, firms benefitting from the cluster policy were already receiving more public subsidy than other firms. The authors comment that this finding has at least two possible explanations: the pattern of support for declining industries could reveal government political preferences, or declining firms could have greater incentive than other firms to lobby for public subsidies (and are successful in doing so). The policy was found to have no effect on employment or exports. The policy did not reverse the relative decline in TFP in the declining industries. No significant effect was found on enterprise survival.

Subsequent cluster-related policy in France - aimed at "poles de competitivité" - has taken somewhat different forms. The *poles de competitivité* policy was launched in 2005 and based on calls for tender leading to financial subsidies for innovative projects managed collectively by firms, research departments and universities. Bellego and Dortet-Bernadet (2013) assessed this policy using a matching technique. The evaluation found that firms targeted by the policy increased R&D spending more than similar firms elsewhere. But increases in sales, patents and exports were not seen. However, some aspects of resilience appear to have been strengthened: export shares among the target firms were less likely to fall (although this was not the case during 2008-2009).

The Danish Agency for Science, Technology and Innovation (2011) has prepared an evaluation titled "The impacts of cluster policy in Denmark". The evaluated policy appears to focus on the development of

business networks and institutional linkages (located in defined areas or clusters). This quantitative impact assessment used 2002-2008 data covering 1 225 firms, and employed propensity-score matching to create a counterfactual. The policy had a number of goals: to provide access to professional competencies from scientists, specialised companies and others; to foster knowledge sharing and ideas generation; to provide opportunities to identify collaboration partners and launch joint projects; to develop linkages with research environments; and to assist internationalisation of companies.

The Danish evaluation showed strikingly positive programme impacts. After just one year of participation, the probability of being innovative was 4.5 times higher for firms in the networks compared to a control group of non-participants. Again after one year of participation, firms in the innovation networks increased their probability of engaging in R&D collaborations by a factor of 4. Assessment of the economic impacts associated with the behavioural changes in firms (i.e. doing more innovation and engaging in more R&D collaboration) will require a longer time series (which is planned).

Also examining a network-oriented scheme, Nishimura and Okamuro (2011) evaluated the impact of Japan's "Industrial Cluster Project" (ICP) on the R&D productivity of participants. A key goal of the ICP was to develop collaborative networks between universities and industries. The authors estimate that participation in the initiative does not affect R&D productivity. In fact, research collaboration with a partner in the same region decreases the quantity and quality of patents. To improve the R&D efficiency of local firms, the authors found, collaborative networks should reach beyond the clusters. However, most initiatives focus on networks that only have local scope.

Using interviews and self-reports, various other evaluations suggest that, as with the Danish evaluation described above, initiatives can effectively foster inter-firm collaboration and/or increased business contacts. As reported in Uyarra and Ramlogan (2012) a survey of more than 1 000 companies and research institutions in the biotech sector in Germany indicated that collaboration was achieved that would not have occurred without the BioRegio and BioProfile initiatives. And Engel et al (2012) showed that BioRegio and BioProfile gave rise to increases in biotech patent applications and public R&D projects.

Viladecans-Marsal and Arauzo-Carod (2012) use a difference-in-difference approach to assess a policy in the city of Barcelona aimed at forming a cluster of knowledge-based firms. In 2000, after decades of economic and industrial stagnation, knowledge-based firms were encouraged to establish in a district called Poblenou. The authors found that the cluster initiative did increase the share of knowledge-based firms in the locality, but only modestly (between 1.3 and 2.1%). Moreover, the authors find the effect to have stagnated over time. In addition, at least some of the positive effect of the initiative may have come at the expense of neighbouring areas.

Falck et al (2010) evaluate a cluster programme introduced in Germany in 1999, the Bavarian High Technology cluster initiative. This programme aimed to increase innovation and competitiveness in the region of Bavaria by stimulating co-operation between science, business, and finance in five target industries. The main activity was to improve the supply of joint research facilities. The authors used a difference-in-difference-in-differences estimator (comparing the innovation performance of target and control firms in Bavaria as well as firms in other German States, both before and after the policy). The authors found that the initiative increased the probability that firms in a target industry would innovate by 4.5% to 5.7% (depending on the indicator of innovation used). Interestingly, at the same time, R&D spending fell by 19.4% on average for firms in the target industries, although the use of external know-how, cooperation with public scientific institutes, and the availability of suitably qualified R&D personnel increased. The authors suggest that the increase in innovation in the presence of falling R&D can be explained by improvements in R&D efficiency consequent on greater access to complementary inputs such as qualified R&D personnel.

Uyarra and Ramlogan (2012) and others note that many policies that are likely to have a major impact on clusters are almost never assessed. Such policies include transport, land-use planning and labour market policies. Instead, evaluations focus on activities implemented under discrete enterprise-oriented and time-bound initiatives.

3.3 National industrial strategies

With the revival of interest in industrial policy there is increased interest in past evaluation evidence and evaluation frameworks for packages of industrial policy at the national level. However, as previously noted, there is a dearth of good examples of *ex post* evaluation and few examples of good practice in establishing evaluation frameworks for new industrial strategy initiatives.

The evaluation challenge is at its most stark at the level of a national industrial strategy. Most of the evaluation problems identified for single instruments and for sectoral and cluster policies apply to national industrial strategy as well. Sector and regional initiatives at least allow the possibility of a reference group in the same country, but that may no longer be an option for a national industrial strategy, even when the industrial strategy has a selective aspect. Industrial strategy initiatives often involve a package of inter-related industrial policy measures, as well as horizontal economic policy measures, complicating the identification and attribution of policy impacts. The challenge is accentuated when the industrial policy interventions include "softer" forms of collaboration and coordination, where even the inputs are harder to measure than for programmes that involve transfers of financial resources. The impact of an integrated and strategic industrial policy on business performance and on the wider economy may also take many years to materialise, making it hard to assess progress within the time horizons typical of most policy makers.

Looking at countries that are attempting to evaluate industrial policy initiatives, the Netherlands is further ahead than most in putting in place structures to evaluate the Top Sectors initiative. The work by the Impact Evaluation Expert Working Group (2012) on impact measurement of business support policies, as well as setting out a framework for evaluating the TKIs as a strategic PPP, also made specific recommendations on methodological issues. It found that natural experimental research design can be applied to most policy instruments and should be used in combination with complementary surveys. It stressed the need for setting out the logic model for the approach and the need for improved data collection. It also found that it was important to clearly recognise complexity, and the preconditions and limitations of evaluations.

The Group also stressed the point that new industrial policy is often based on a dialogue with many stakeholders on a broad range of topics. Econometric approaches work best while analysing an instrument that targets one problem of a clearly defined user group. It was recognised that evaluating broad strategies and new demand driven/bottom-up policy initiatives is therefore difficult and likely to require a mix of methods.

Work on the evaluation of support for business under way in Ireland also highlighted some broader issues in the evaluation of industrial policy, notably the challenge for Irish policy makers in coming to terms with a shift in the role of the state from one providing support for capacity building to one performing a more facilitative and developmental role. Forfás, the national policy advisory board for enterprise and science, is undertaking a programme of evaluations of enterprise agency supports. The evaluations are undertaken in line with the Forfás evaluation framework, which has been informed by the international experience with industrial policy evaluation. The evaluations are being undertaken across three main thematic areas: start-ups and entrepreneurship; R&D and innovation support; and business development programmes (which include internationalisation, productivity, scaling and capability and management development). The evaluations assess the appropriateness, effectiveness and efficiency of each programme in its own right, but also in relation to other programmes and the overall policy context.

The European Union's new approach to industrial policy, launched in 2010 in response to the economic and financial crisis and updated in 2012. The approach includes horizontal measures to facilitate trade, access to capital and skills development, but also more targeted measures to foster investment in innovation in specific sectors, specifically: advanced manufacturing technologies; bio-based products and technologies; clean vehicles; key enabling technologies; Smart Grids; and sustainable construction. In addition, the EU Commission is developing new appraisal and evaluation tools, including *ex ante* competitiveness proofing, and sector-specific fitness checks and cumulative cost assessments. The use of these tools may reveal something about the success of past policies and/or the need for new policy measures. The Commission is also considering new approaches to regional policy evaluation, and the possibility of making a commitment to evaluation a requirement for state aid approval.

In Sweden, the Agency for Growth Policy Analysis has been tasked with monitoring and evaluating the Swedish Innovation Strategy, launched in October 2012. The strategy takes a broad view of innovation and sets a vision of having a world-class innovation climate in 2020. The strategy sets out goals for human resources, research and higher education, framework conditions and infrastructure, businesses and organisations, public services and regions and environment. Sub-targets, numbering 17 in total, are also set for each area. But it lacks a specific budget and operational goals, making conventional monitoring and evaluation difficult. Faced with the remit of producing an annual report on progress towards the overarching goal in order to support policymaking, the Agency's approach is based on the following principles:

- Use existing indicators aligned with the goals and sub-targets of the strategy.
- Avoid composite indicators which are good as signals but less good in communicating the areas that need attention and action.
- Use "real" indicators linked as closely as possible with policy action.
- Allow for international comparisons as far as possible.

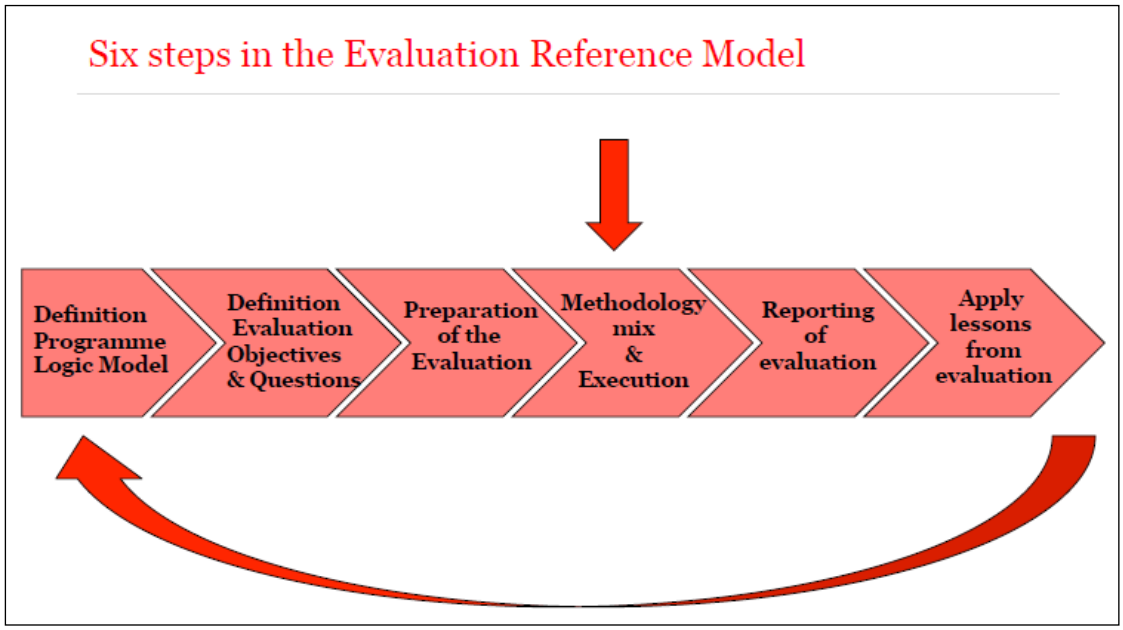
Benchmarking approaches like this are commonly used by a number of countries in monitoring the progress of industrial or innovation strategy. Such approaches have their value, and there are strong arguments for using a variety of indicators, rather than a composite, to capture the full breadth and complexity of the policy area. There is, however, a risk that the exercise loses steam if annual monitoring shows little change in the indicators or they move in conflicting directions. Monitoring systems and reporting systems therefore need to be designed to be robust to political and economic change. It is also important to find ways to present annual reports that strike the right balance between rigour and consistency on the one hand and impact and influence on the other.

An interesting example of meta-evaluation in the industrial and innovation policy sphere is the exercise currently under way under the auspices of the TAFTIE Task Force on Benchmarking the Impact, Efficiency and Effectiveness of Innovation Agencies.[20] Eleven innovation agencies are participating in the task force, which is chaired by NL Agency and supported by research from Technopolis (Boekholt, 2014). The Task Force is currently developing three "benchmarks", on evaluation techniques, impact/effectiveness and efficiency. The findings on the latter two in terms of the performance of the innovation agency will provide interesting examples of meta-evaluation.

Perhaps of even greater relevance is the work on benchmarking evaluation techniques, for which a "reference model" is being developed. The idea is to draw from an overview of state-of-the-art methodologies and evaluation tools and existing generic evaluation guidance to produce an idealised standard, against which the practice of the participating innovation agencies can be assessed. The

instruments included in the comparative exercise are: grant schemes aimed at business R&D, grant schemes aimed at R&D cooperation, innovation vouchers and cluster programmes/competence centres. The reference model (Figure 1) will cover guidance on questions such as: specification of the logic model; definition of research questions; management of the evaluation process; choice of methodologies; and analysis, drafting and publication. For each step in the reference model, the TAFTIE project, drawing on evaluation guidance from six countries[21] and the European Commission, identifies key quality criteria that are considered to constitute best practice for the four innovation instruments under consideration. Evaluations submitted by participating innovation agencies are then assessed against the benchmark.

Figure 1. TAFTIE Evaluation Reference Model



Source: Boekholt (2014).

The TAFTIE project, which has not yet been finalised, is designed to generate insights on the impact, effectiveness and efficiency of policy instruments used by innovation agencies. However, one of the main emerging findings is that benchmarking impact and effectiveness in an internationally comparable way is not really possible given the difference in evaluation approaches and methods used. In addition, many of the evaluations focus on inputs, activities and programme outputs, with little evidence of the impact on innovation capabilities, competitiveness or productivity. Even for the better evaluations, differences in national context, programme design and timing of the evaluation hamper comparison. Against this background, the emerging conclusions from the TAFTIE review (Boekholt, 2014) are mainly focused on evaluation methodology:

- Good evaluation requires a mix of methods and triangulation between them. There is no "golden rule" for the best methodology mix – it depends on the instrument, the context and the timing of the evaluation.
- A different set of methods will be appropriate for "simple" programmes with a clear rationale (for example, business grants for R&D) and "complex" programmes, where attribution is a major issue (for example Competence Centre programmes).
- Programme objectives must be clearly defined from the outset and an evaluation framework developed at the start of the programme.

- Evaluation questions should not be unrealistically ambitious or too narrowly focused – they need to be adapted to the timing of the evaluation and take account of the time lags before impacts at company level and spillovers to the rest of the economy can be observed.
- Data requirements for an evaluation should be specified at the start of the programme and data should also be collected for counterfactual analysis – this may require work on access to appropriate linked micro-data sets.
- Conclusions should match the evidence – commissioners of evaluations should insist that the conclusions can be easily derived from the evidence gathered and that all the methods used are transparent.

Drawing on the TAFTIE findings, and the experience of other countries grappling with the problems of meta-evaluation in the context of industrial strategy, a number of conclusions can be drawn on the evaluation of industrial strategy:

- There is a dearth of good examples of monitoring and evaluation of industrial strategy, though there are some encouraging signs of good practice being put in place.

- Being clear about the objectives and the policy rationale or programme theory is an important first step in any programme evaluation. It is easier if this is established from the outset but it may need to be reconstructed *ex post.*

- Econometrics and other quantitative methods should be used as far as possible. Experimental research design methods could be used in a wider variety of circumstances than so far applied. Reference models are being developed for some commonly used instruments.

- However, evaluation of broad strategies will require complementary perspectives and methodologies to be used. Evaluation of the main components of industrial policy package can be carried out but care needs to be taken in attributing impacts and further work will be needed to assess systemic effects.

- Even with improved evaluations, thinking about systemic effects will have to be done with limited evidence and in conditions of uncertainty – but it should be attempted nonetheless.

- External peer review and benchmarking are valid evaluation techniques when it is otherwise difficult to establish robust control groups but the results need to be interpreted with care.

- Evaluation is itself a learning process for policy makers.

4. NEW INDUSTRIAL POLICY AND THE ROLE OF EVALUATION

This chapter brings together thinking on new industrial policy and some recent developments in approaches to evaluation. There are ideas in the new industrial policy literature which can be implemented without any change in the approach to evaluation, and there are important strands of the new thinking on evaluation which apply to traditional approaches to business support and indeed to other areas of government policy. But it is at the intersection of these two trends that the most interesting conclusions and policy lessons can be drawn. The two trends are first briefly assessed independently and then the potential for the use of new approaches to evaluation as a tool of new industrial policy is explored.

4.1 New industrial policy

The renewed interest in industrial policy around the world has a number of causes and manifests in different ways (O'Sullivan et al, 2013; Warwick, 2013; Stiglitz et al 2013). Many countries are being more explicit about industrial policy than in the past and there is renewed interest and debate about the appropriate role of Government in seeking to influence the evolution of the structure of the economy. Some countries are adopting clear sectoral priorities for the first time, others are building on earlier sectoral approaches, while some are adopting more horizontal approaches to industrial policy. Countries are responding to changing circumstances in different ways. However, at the risk of over-generalising, by comparison with earlier historical experience of industrial policy, new industrial policy might be characterised as exhibiting some or all of the following characteristics (Warwick, 2013):

- A greater emphasis on improving systems and in particular interventions aimed at building networks, improving co-ordination, promoting awareness and securing strategic alignment, including more use of strategic partnerships with industry and joint consideration of the competitiveness challenges facing business;

- Notwithstanding the continued use of sectoral targeting in some countries, a shift of focus towards tasks/activities and technologies, and away from support for single firms, national champions, tariff protection and other product market-focused interventions;

- At a time of fiscal austerity, less reliance on direct support in the form of state aids and market-failure correcting subsidies; and

- Some merging of the boundaries between strategic and defensive industrial policy – while policy makers have reacted to the global crisis by adopting defensive measures where industry has faced acute finance constraints and/or demand pressures, there are also many examples of more strategic interventions in support of new areas of activity.

Consistent with this picture, O'Sullivan et al (2013) discern a trend towards greater alignment of longer-term investment and policy planning with the needs of industry. The same authors also see evidence of growing understanding amongst policy makers of the systems nature of manufacturing and the challenges posed by the increasingly complex interdependencies in the modern economy.

Advocates of new industrial policy see government as a facilitator in the face of complexity and uncertainty, enabling closer co-ordination between individual economic agents and greater experimentation

in the economy. In a globalised world, the policy space open to governments is increasingly constrained. In these circumstances, some countries are shifting their approach away from one-time attempts to "pick winners" to the design of better processes for search and "self-discovery". In the process, it is accepted that mistakes and errors are inevitable; the policy challenge is to design governance procedures to detect and correct them and to manage the associated vested interests (Rodrik, 2008).

Mazzucato (2013) has also stressed the importance of public-private partnerships in modern capitalist economies and argued that the role of the state should not be seen purely as de-risking or market-failure-correcting, but that it can and should play an entrepreneurial role, taking the lead in the development of new technologies, such as green technology, just as it has done in the past with the Internet, agricultural science, bio-technology and nano-technology. As with the notion of the experimental state – discussed for instance by Bakhshi, Freeman and Potts (2011) - collaboration between government and industry in the face of complexity and uncertainty is central.

Box 1. Smart specialisation – new industrial policy in action

Specialisation is a driving force that enables the exploitation of economies of scale and differentiation for the creation of economic value. Smart specialisation is a policy framework combining industrial, innovation as well as educational policies (including their design, implementation, and evaluation) in order to promote new growth opportunities based on innovation and knowledge. The smart specialisation approach thus aims to support innovation-based economic growth strategies at different levels of government and policymaking based on:

- A more effective spending of public resources, concentrating on certain domains of knowledge or expertise.
- The creation of synergies between public support mechanisms for R&D and innovation, industrial promotion and training institutions.
- The elimination of fragmentation and duplication of policy interventions that may result in a waste of public resources.
- The identification of the strongest or promising domains for entrepreneurship and growth through a careful analysis of the existing capabilities, assets, competences, competitive advantages in a city, region or country.
- Mechanisms to enable strategic development based on multi-faceted and multi-governance interactions.
- Mapping and benchmarking of clusters including analyses of the role and influence of key players.
- Evidence-based monitoring and evaluation systems to select the knowledge domains and innovation projects

Source: http://www.oecd.org/sti/inno/smartspecialisation.htm.

The notion of the state as match-maker and collaborator is also emphasised by Kuznestov and Sabel (2011). Rather than interventions to select firm- or sector-level winners, or policies designed to improve framework conditions, Kuznetsov and Sabel identify a set of "mezzo level" interventions designed to make connections between economic agents ("matching winners", rather than picking winners or backing winners). There are similarities here too with the concept of "Smart Specialisation" [Box 1 and Foray (2012)]. The smart specialisation approach also places emphasis on interactions between government and business and on evidence-based monitoring and evaluation in selecting the innovation projects on which to focus public resources.

4.2 Recent developments in evaluation

While the evolution of thinking about new industrial policy in the policymaking and academic communities is relatively recent, there has been a trend in the evaluation community for some time towards greater use of more rigorous techniques for evaluation, using randomisation and experimental methods. In the words of Glennerster (2012):

> "Advances in the theory and practice of running randomised evaluations mean that a wider range of questions can be answered than ever before. Elsewhere in the world, fundamental questions are being answered about how humans behave, which in turn are being used to design new policies which themselves are rigorously tested. By learning from these results, and by conducting more randomised evaluations.... it will be possible to design more effective policies and do more with less."

Rigorous evaluation requires the establishment of a counterfactual (what would have happened if the policy had not been in place). Researchers must also deal with the problem of selection bias, whereby those receiving support exhibit characteristics that differ from non-recipients in a way that may bias the estimated impact of the policy. In many social policy spheres (for example, health and education), techniques and practices have been in use for some time allowing researchers to carefully select control groups and run randomised trials in order to establish the difference between outcomes in treated and untreated groups that can be attributed to the treatment. Piloting and evaluating are also common in other areas such as labour market economics where the collection and use of evaluation evidence is more developed. Banerjee and Duflo (2011) describe the "recent surge in experimental work" in development economics (Box 2).

In general, however, much less has been done in the industrial policy sphere. The establishment of more rigorous counterfactuals using propensity score matching techniques (for example, Oh et al, 2006) and instrumental variables (for example, Criscuolo et al, 2012) and the use of experimental methods in developing industrial policy is becoming more common. And NESTA and the Kauffman Foundation have recently launched an initiative to promote the use of randomised control trials (RCTs) in the industrial and innovation policy sphere and to help build capacity and a network of researchers in this area (Box 2). However, Bakhshi et al (2011) find that experimental policy evaluation approaches remain marginal, with non-experimental policy evaluations the norm. Potter and Storey (2007), for example, provide an extensive review of OECD best practice in the evaluation of SME and entrepreneurship policies without any mention of either the application or the potential of experimental methods.

In addition to the long-established trend in the evaluation literature towards greater use of experimental methods and econometric techniques, another more recent trend has been the growing interest in the developmental role of evaluation. To some extent, developmental evaluation draws on the insights gained through the use of experimental techniques but, as emphasised by Patton (2006, 2010), there are more fundamental differences between the traditional approach to evaluation and developmental evaluation. In Patton's view, developmental evaluation can complement the two conventional categories of evaluation: formative (process) evaluation and summative (impact) evaluation. Process evaluation is used to help improve a programme or policy and the way it is delivered. Impact evaluation is typically used to assess the merit or worth of a programme or policy in order to determine whether it should be sustained, discontinued or scaled up. However, both forms are at their most effective when an intervention or programme of work is finite or where there is the option to pause the programme while an evaluation is conducted. Policy-makers often seek an approach that delivers results during the planning and execution of a programme and helps them react to new information and emerging results.

Box 2. A J-PAL for Innovation and Growth Policy?

The Abdul Latif Jameel Poverty Action Lab (J-PAL) (http://www.povertyactionlab.org/about-j-pal) was established in 2003 as a research centre in the Economics Department at the Massachusetts Institute of Technology. Over the last 10 years, it has developed into a global network of researchers who use randomised evaluations to address critical policy questions in the fight against global poverty. Founded by MIT Professors, Bannerjee, Duflo and Mullainathan, it has grown to have independent regional offices in six continents, 91 affiliated professors, 1 620 people trained in the use of RCTs and 448 ongoing or completed evaluations in 55 countries. J-PAL activities include:

- **conducting rigorous impact evaluations**: randomised evaluations to test and improve the effectiveness of programmes and policies aimed at reducing poverty;
- **policy outreach**: analysing and disseminating research results and building partnerships with policy makers to ensure that policy is driven by evidence, and effective programmes are scaled up;
- **capacity building**: equipping practitioners with the expertise to carry out their own rigorous evaluations through training courses and joint research projects.

Inspired by the J-PAL model, NESTA is launching, together with the Kauffman Foundation and innovation agencies and government bodies in several countries, a new global innovation, entrepreneurship and growth lab (http://www.nesta.org.uk/project/experiments-innovation-entrepreneurship-and-growth). This new lab will aim to experiment with new approaches to increase innovation, support high-growth entrepreneurship and accelerate business growth, improving the evidence base on their effectiveness. Like J-PAL, the lab will undertake RCTs, contribute to capacity-building and play a matching and brokerage role, connecting academic researchers with each other and with organisations, both public and private, open to running RCTs.

As a first step, NESTA and the Kauffman Foundation have launched a grants programme to fund researchers and organisations proposing to undertake RCTs related to innovation, entrepreneurship and business growth. The aim of the programme is to build a critical mass of researchers and organisations engaged in RCTs in these areas and showcase the potential of experiments to lead to insights and actions that make innovation and industrial policy making more effective.

Source: J-PAL.

The developmental evaluation approach is designed to be helpful in such a context. Some of the main differences are summarised in Table 2. Central to developmental evaluation is the notion that evaluator and policymaker are part of a team seeking to understand the impact of policy in real time and adapt it in a complex and changing environment. It is thus particularly well matched to the modern conception of industrial policy where policy makers engage in an iterative process of dialogue with business and others, and there is a combination of top-down and bottom-up approaches. If done well, developmental evaluation should also be more embedded into the policy making process and not seen as an additional burden placed on policy makers.

Table 2. Traditional evaluation and developmental evaluation compared

	Traditional evaluation	Developmental evaluation
Purpose	Renders definitive judgements of success or failure.	Provides feedback, generates learning, supports direction or affirms new direction.
Success measure	Measures success against predetermined goals.	Develops new measures and monitoring mechanisms as goals emerge and evolve.
Independence	Positions the evaluator outside to assure independence and objectivity.	Positions evaluation as an internal, team function integrated into policy development.
Design	Design the evaluation based on linear cause-and-effect logic models	Design the evaluation to capture system dynamics, interdependencies and emergent interconnections.
Learning	Aims to produce findings generalisable across time and space	Aim to produce context-specific understanding that informs further policy development.

Source: Adapted from Patton (2006).

4.3 New approaches to industrial policy evaluation

New and complementary approaches to the design and evaluation of industrial policy are being explored against the background of the new trends in thinking about industrial policy and growing interest in experimental and developmental evaluation. Partly, this is about the spread of more rigorous techniques into areas of industrial policy where they can be most directly applied. But new evaluation approaches are also being used as a tool of modern industrial policy. Typically, new approaches to industrial policy require processes of information discovery, policy experimentation and networking. These are partly built on the premise suggested by Rodrik (2004) that "...the task of industrial policy is as much about eliciting information from the private sector about significant externalities and their remedies as it is about implementing appropriate policies." These new approaches seek to tackle one of the main challenges in policy, namely the information asymmetry between government and business. In eliciting more information from the private sector, more informed policy making is made possible (Bakhshi, Freeman and Potts, 2011).

Examples can be found of the use of techniques such as random assignment in the industrial and innovation policy sphere. One such example is found in the innovation voucher scheme in the Netherlands. The goal of the policy was to increase exchanges of knowledge between SMEs and (semi-) public research institutions. With the vouchers, SMEs could profit from the “ready to use” knowledge at such institutions to develop new products, processes and services. The policy instrument was developed through a series of pilots in which vouchers were randomly allocated to scheme applicants. In the first evaluation, based on interviews with 71 of the 100 “winners” in the lottery and 242 of the 944 “losers”, it was found that 87% of “winners” commissioned a research project, compared with 8% in the control group, suggesting 79% of vouchers led to additional projects. A follow-up study after 18 months found evidence of a small effect on process improvements but no significant effects on product innovation or the introduction of new processes amongst the winners.[22] The evaluation of subsequent pilots produced similar results, albeit with slightly lower rates of additionality. In the light of the positive findings on additionality, the popularity of the vouchers with SMEs and the fact that the voucher scheme initiated a more positive approach of SMEs by

the knowledge institutions, the policy was rolled out across the economy in 2007-10 and the budget increased. The voucher scheme was, however, suspended in 2011 due to budget cuts.

The Dutch scheme showed that random allocation can be used in an industrial policy setting. Concerns about the potential unfairness of a random allocation process proved to be unfounded (or at least there was no greater concern than for other allocation mechanisms) and it was possible to compare treated and untreated firms on a comparable basis. But changes in the scheme as it was rolled out and as the budget increased reduced the relevance of the pilots and there were some difficulties in tracking firms over time, particularly the "losers" who were less inclined to cooperate.

Bakhshi, Edwards et al (2011) discuss the results of a randomisation experiment with innovation vouchers in the United Kingdom. Constructed as a test of the hypothesis that SMEs that spend more on creative goods and services tend to be more innovative, firms in the Manchester area were offered GBP 4 000 vouchers, issued by lottery, to be spent with Manchester-based creative businesses (for example, in design, advertising, publishing, software and film). The study found that SMEs receiving the voucher were 78% more likely to undertake their project than those who did not and there was strong evidence of short-term additionality in terms of increased innovations after six months. However, no significant changes in terms of innovation additionality or other behavioural benefits were found after 12 months.[23] Qualitative research suggested that the benefits were more likely to dissipate when SMEs made the wrong choice of partner, where the relationship was seen as purely transactional or where there was no shared understanding of the brief. Further measures, therefore, might be needed to ensure that the improvements were sustained, and this would affect the cost effectiveness of such initiatives.

Also in the United Kingdom, a "growth voucher" scheme is being developed which will use randomised control trials in the delivery of a programme of business advice for small and micro businesses. The aim will be to establish the effectiveness of different channels for advertising the scheme and for delivering business advice. The programme will be rolled out from the beginning of 2014, outcomes will be tracked using self-reported survey data six months and two years on, and the longer term impact will be assessed using administrative data.

These studies illustrate that experimental methods are increasingly being used in the evaluation of some facets of industry and innovation policies, but there is potential to do more. The use of experimentation and the iterative approaches of developmental evaluation fit well with the notion of a smarter state, which seeks to learn from the market and the discovery process of entrepreneurs in selecting appropriate targets for public policy.

White (2013) seeks to apply the ideas and techniques of developmental evaluation to industrial and innovation policy. White takes as his starting point that new approaches to industrial policy are rooted in information and coordination failure – the state and market are both ignorant and informed in different ways. The new approaches often involve collaborative and experimental engagement with the private sector. They tend to be "place-based" to tackle information gaps and information asymmetry through local interactions and collaboration and they require local knowledge to be elicited and aggregated for learning and for future policy development.

White goes on to argue that, as the problems to be addressed move from the simple to the complicated and complex, the policy response also becomes more complex (Table 3) and the design of evaluation strategies needs to adapt accordingly. Hence, in White's categorisation, for "*simple"* problems, the design of industrial policy would typically be fairly straightforward, based on a single instrument and probably a single delivery agent, with cause and effect relatively well understood and best practice identifiable. A variety of evaluation methods might still be used, ranging from basic monitoring through analysis of secondary data and case study analysis to experimental or quasi-experimental methods.

In more *"complicated"* situations, there would typically be a variety of different industrial policy interventions simultaneously, with several delivery partners and a range of other influences on outcomes. Policy responses could still be predictable in principle but the combination of different instruments and influences makes it difficult to be precise. Good but not best practice can be identified. In these circumstances, previous evaluation evidence needs to be used with care, and it will be more difficult to establish reliable control groups. Policy responses could still be predictable in principle but the combination of different instruments and influences makes it difficult to be precise.

For "*complex"* problems, the uncertainty is compounded still further. Responses may be non-standard and highly context-dependent, and there is uncertainty about how the context will change, the role of other actors, and the impact of other influences. Cause and effect are not well understood or even well-defined at the outset and the potential to generalise from past experience is limited. In this world, previous evaluations can give a general guide but not precise insights, quantitative methods may be difficult to apply and process evaluation will be as important as impact evaluation. Qualitative approaches to evaluation and in particular case studies will be important not just for *ex post* "audit-style" evaluations, but for active learning and policy adaptation as knowledge of the impact of policy evolves.

White's analysis can be extended to make more explicit the two underlying difficulties that confront industrial policy makers and evaluators in the real world – that which is *complicated* (lots of parts) and that which is *complex* (uncertain and emergent outcomes).[24] The first arises in situations where, even though cause and effect are reasonably well understood, there may be multiple interventions and influences at work. An example might be support to improve access to finance where a variety of different instruments and agencies are used to provide support to SMEs – in such circumstances, the interaction of multiple sites, levels of governance and channels of impact will complicate the evaluation of the impact of any one intervention or indeed the whole package of such measures. The second dimension is complexity, which can arise when there are important systems effects, including non-linearities and externalities, or where cause and effect are not well understood and policy makers need to react to changing circumstances. This may arise both in the case of a single measure (for example, innovation vouchers, as already described) or in the case of a package of measures. In the latter case – which may be typical – those designing policy and strategy for innovation and industry face a challenge which is both complicated and complex (Table 4).

Table 3. Simple, complicated and complex problems and the developmental role of evaluation

Problem	*Industrial policy characteristics*	*Evaluation method*
Simple	• Standardised and discrete change in previous intervention practice • Usually a single delivery agent • Cause and effect reasonably well understood • Works similarly in different contexts • Best practices are identifiable	• Monitoring data supplemented with results from previous evaluations • Experimental or quasi-experimental methods to assess what works • Analysis of secondary data to identify interruption in time series • Qualitative methods – case studies of theory testing kind
Complicated	• Multiple components and/or requiring supportive context or interventions • Often involving multiple partners with clearly defined roles • Cause and effect dependent on context and other factors • Works differently in different contexts but in predictable ways • Good rather than best practice identifiable	• Previous evaluation evidence only of use if clearly “like-for-like” • More difficult to establish control groups and allow for other variables • Quasi-experimental methods may be feasible • More sophisticated analysis of secondary data required • Qualitative methods – case studies of theory building kind
Complex	• Non-standardised and responsive to changing circumstances • Multiple partners and stakeholders with changing values/roles • Cause and effect not well defined at outset • Uncertainty about future context, relationships and values • Rapid decay in potential for generalisations	• Previous evaluations likely to be helpful as a guide to what methods work • Integration of formative and summative evaluations • Quantitative methods difficult to apply - limited internal/external validity • Use of systemic approaches such as network analysis • Qualitative methods - action learning and problem solving case studies – to understand ‘what’s working’ to help build and sustain more steady states

Source: White (2013), based on Rogers (2011).

Table 4. Two-way classification of the industrial policy evaluation challenge

	Single measure	Package of measures
Standard/ well understood	Simple – just do it better, use rigorous counterfactuals, control groups, state-of-the-art estimation techniques.	Complicated – apply single measure techniques to components where possible, take account of interactions and multiple treatments and influences.
Uncertain/complex	Complex – use experimental methods, test/learn/adapt.	Complex and complicated. Counterfactuals may not be possible. Apply single measure techniques to components, take account of interactions and systemic effects, use qualitative measures and more informal methods of learning by doing.

Source: White (2013).

Viewed within this framework, the challenges for industrial policy evaluation can be summarised as follows:

- *Simple:* how to get better evaluation of single instruments or discrete areas of policy in situations where cause and effect are relatively well understood. This is mainly about getting more rigorous evaluation using state-of-the-art techniques in the analysis of, for example, R&D support policies, loan guarantee schemes, equity market interventions and other discrete instruments of business support. Where possible, industrial policy evaluators need to rely less on subjective self-assessment of impact and do more to establish rigorous counterfactuals, using control groups, propensity score matching, instrumental variables and other econometric techniques.

- *Complicated:* how to get better evaluation of industrial strategies and packages of measures (which might include cluster development, or sector initiatives as well as national strategies). This is likely to require the use of a programme evaluation approach and a more eclectic range of evaluation methods. It will also require that frameworks be put in place clearly specifying desired outcomes and channels of impact. Where there are multiple levels of governance, an overarching evaluation framework will also need to be put in place. Where there are multiple interventions, meta evaluation based on auditing the performance of industrial interventions may be possible, but interaction effects need to be taken into account.[25]

- *Complex:* how new evaluation techniques may be used in the context of discrete industrial policy initiatives where outcomes are more uncertain and the environment more complex. Experimentation, iteration and learning are at the heart of this approach, so the use of randomisation and experimental/quasi-experimental methods is potentially just as important here as in the "simple" category. However, in complex environments, evaluation will be far from straightforward and the audit approach to evaluation may not be appropriate or possible. More iterative approaches will need to be used and it may also be necessary to supplement econometric methods with the use of case studies and other more qualitative methods.

- *Complex and complicated:* how to apply developmental evaluation to new industrial policy. As new industrial policy typically involves a package of measures so the problem encountered in the "complicated" category are present here as well – being explicit about the logic model, or theory of change, may be even more important. But, in a complex environment, the model may need to evolve as new information emerges. In addition, the instruments used in this tend to be "softer" and more focused on institutions than in more traditional sector targeting industrial strategies, posing new challenges again.

This approach does not itself provide a framework for undertaking industrial policy evaluation. Indeed, given the multitude of different approaches to industrial policy and the difficulty of drawing up a generic framework that would cover all eventualities, it would be as well to avoid being too prescriptive. However, the approach set out above prompts a number of observations that can be made about good practice in industrial policy evaluation:

First, it is important to avoid a kind of "evaluation bias" towards "simple" problems and solutions. If there is too much emphasis on hard evaluation evidence or too much insistence on building in rigorous quantitative evaluation from the outset of a programme then policy makers may be steered towards the subset of industrial policy measures for which hard evaluation evidence is available (for example R&D support policies, some capital market interventions) and away from softer interventions (for example institution building, public-private partnerships) which might be just as important for innovation and growth. Since most real world problems tend to be complicated and/or complex rather than simple, it is likely that a mix of policy measures and an eclectic approach to evaluation and the use of evaluation evidence will be required.

Second, interaction effects and the systems nature of policy impacts need to be recognised and taken into account in industrial policy evaluation. Evaluations which focus on the difference between treated firms and a non-treated control group will miss the linkages to other sectors, spillovers, cluster effects and other externalities and system impacts which are often the main rationale for industrial and innovation policy. Conventional measures may also fail to take account of dynamic effects and the time trajectory of responses to industrial policy interventions. Experimental and quasi-experimental methods may still have a role to play but they must be supplemented by other techniques to take account of as full a range of impacts as possible.

Third, industrial strategy – whether for the economy as a whole or for one or more sectors, regions or clusters – generally involves a package of measures, whose influence will be difficult to disentangle using even quite sophisticated econometric techniques. By their very nature, many industrial strategy initiatives belong in the "complicated" category and can only be assessed using a mix of different techniques, including case studies and qualitative assessment. However, in most situations, evaluation will be easier if the industrial strategy is accompanied by a theory of change or logic model against which progress can be assessed at the appropriate points in time.

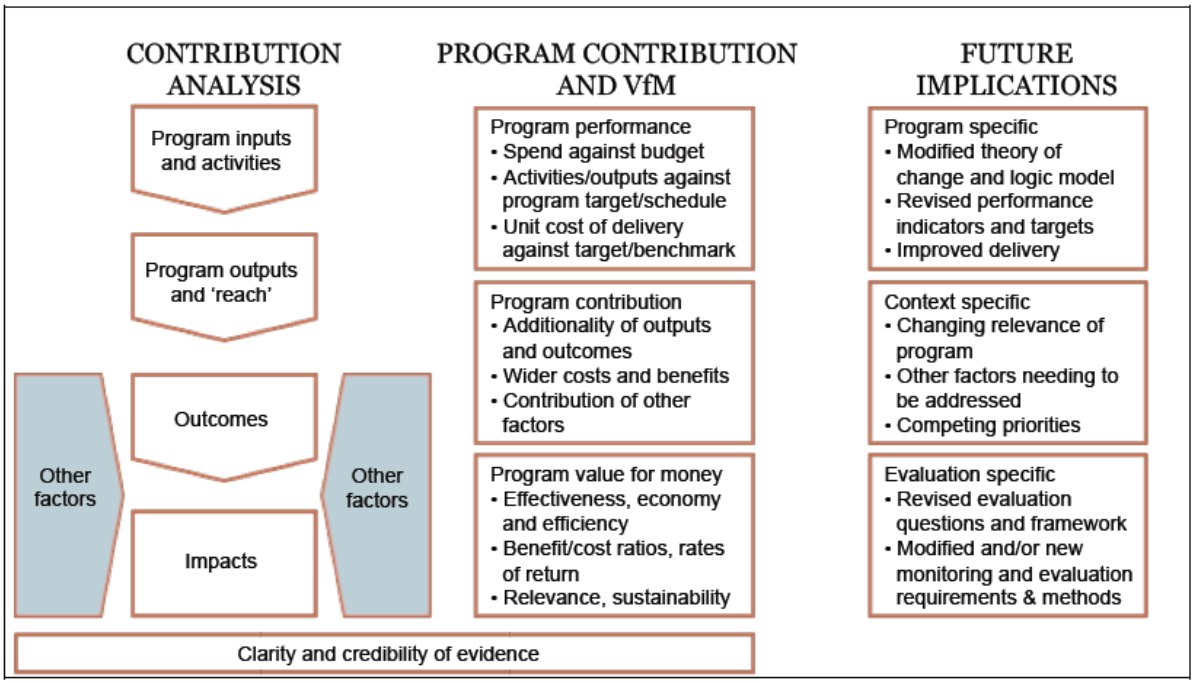
Fourth, in complex problems, where there is uncertainty about programme impacts and even about the channels of impact, evaluation has an important role not just as a form of *ex post* audit, but as a policy instrument in itself, one that helps deal with uncertainty and complexity. This is relevant to single interventions where cause and effect are not well understood but it is also very much the world of new industrial policy, in which it is accepted that government cannot operate industrial policy levers in predictable ways, that instead policy makers need to construct policy interventions as a system of discovery through collaborative and experimental engagement with the private sector. In the "experimental state", evaluation should not be an overhead that is resented by policy makers; rather it is a tool for eliciting information about what works and a way of refining and improving policy in a world where ignorance and complexity is the norm (Bakhshi, Freeman and Potts, 2011; White, 2013).

Fifth, in a world which is both complicated and complex, where uncertainty, multiple influences and interactions are the norm, there may well be a need for a more developmental approach to evaluation. One possible approach is discussed in the next section.

4.4 An illustrative evaluation evidence assessment framework

Matthews and White (2013) propose an approach to developmental evaluation of industrial policy based on sequential hypothesis-testing. Their framework is being piloted at state level in Australia, following a decision to bring evaluation work in-house. It is also being explored by the Australian Productivity Commission as the basis for an evaluation of the impact of regulation on the business sector. The approach, based on techniques used in the intelligence community, is described as a "faster and smarter" alternative to traditional audit approaches to evaluation, which are often resource-intensive and slow to deliver. The proposal is to adopt a framework for formulating and testing succinct propositions about policy in conditions of complexity and uncertainty in a structured and sequential manner. A standardised reporting template is used to collate and analyse the available evidence from different interventions in a single framework. The emerging results from hypothesis testing and evaluation are then used to inform policy and programme design. The approach is illustrated in Figures 2 and 3.

Figure 2. Possible components in a standardised reporting template



Source: Matthews and White (2013).

In Matthews and White's proposed Evaluation Evidence Assessment Framework, the process for a structured approach to hypothesis testing (Figure 2) and the template they suggest should be completed at each stage (Figure 3) can be divided into five sections:

> *Section A* – Corresponding to the leftmost panel of Figure 2 and the upper panel of Figure 3, this would give basic data such as the name of the evaluated programme, period for evaluation and state objectives, as well as setting out the logic model - specified in terms of motivation, inputs, activities, outputs, outcomes, beneficiaries and impacts - and a narrative description of the rationale or theory of change.
>
> *Section B* – This section would develop the key propositions on programme outcomes and impacts, accompanied by a brief summary of the available evidence, conclusions drawn and commentary. It would draw on the indicators suggested in the central panel of Figure 2.

Section C – This section would develop the key propositions on factors contributing to programme outcomes and impacts, accompanied by a brief summary of the available evidence, conclusions drawn and commentary. Like Section B, it would draw on the indicators suggested in the central panel of Figure 2.

Section D – This would provide an overall assessment of the effectiveness and value for money of the programme, a summary of the available evidence, conclusions drawn and commentary. Together with Sections B and C, this section would comprise the Programme Contribution and Value for Money analysis indicated in the central panel of Figure 2.

Section E – Corresponding to the rightmost panel of Figure 2 and the last component panel of Figure 3, this section would contain a short narrative giving an overall summary, conclusions and implications for future programme formulation and delivery. It could form the basis for formulation of a new set of hypotheses and the next iteration of the framework.

A template approach such as this is intended to encourage policy makers and evaluators to be explicit in setting out the rationale, objectives and hypotheses to be tested from an early stage in the policy intervention. Monitoring of key indicators and outcomes would be used as the basis to revise the hypotheses to be tested and to fine-tune the policy as evidence emerges of its impact.

The above approach should be seen as a complement to existing methods rather than a replacement. It could, for example, serve as a less resource-intensive way of providing a real-time monitoring mechanism that would provide an early warning of whether any problems are emerging. Full evaluation and review would be needed to diagnose the root causes of any problems identified by this monitoring function. There would still be potential to make cost savings because a full evaluation and review would be triggered only in certain circumstances. Such an approach also has the attraction of being more helpful to policy makers and programme administrators, making them more likely to engage positively with the evaluation.

Figure 3. Evaluation Evidence Assessment Framework – an illustration`

Evaluation Evidence Assessment Framework (EEAF)						
Policy, programme or project						
Policy, Programme or project objectives						
Period evaluated						
Section A: policy or programme logic model and theory of change ➔	Section B: Policy or programme outcomes and impacts ➔	Section C: Factors contributing to outcomes and impacts ➔	Section D: Overall assessment of effectiveness and value for money ➔	Section E: Summary conclusions and policy implications		
Section A: Policy or programme logic model and rationale						
Motivation	**Inputs**	**Activities**	**Outputs**	**Outcomes**	**Beneficiaries**	**Impacts**
Rationale or Theory of Change						
Propositions-hypotheses	**Evidence**	**Conclusion**	**Commentary**			
Section B: Key propositions on programme outcomes and impacts – achieved and in prospect						
Section C: Key propositions on factors contributing to outcomes and impacts, especially role of the evaluated intervention						
Section D: Overall assessment of the effectiveness and value for money of the public sector investment – achieved and in prospect						
Section E: Summary conclusions and implications for future policy or programme formulation & delivery						

Source: Matthews and White (2013).

4.5 Evaluation in the policy cycle

The approach described in the previous section may be particularly helpful to policy makers facing the need to monitor the impact of their intervention in real time. Full impact evaluations play an important role but they can take many years to deliver. The 2008-09 economic and financial crisis is a good example of how tensions may arise between the time required for a full impact evaluation and the needs of policy makers and programme managers seeking to design policy and adjust programmes to respond to emerging knowledge about the nature and extent of the crisis. A developmental approach of the kind described here

may be a useful complement to existing methods. Taking a pragmatic approach to indicators that are likely to be available at an early stage rather than requiring evidence of final impact could enable the "clock speed" of the evaluation cycle to increase, bringing it more into phase with the policy cycle.

The use of evaluation in the policy cycle could also be improved if good processes are put in place for feedback from evaluators to policy makers, both during the programme and at its conclusion (Matthews and White, 2013). During the programme, evaluation can be part of a process of iterative and experimental policy development. Process evaluation can also offer real time identification of barriers that could pose a risk to successful delivery. Evaluation at the conclusion of the programme can be used to learn lessons for new policies and projects. But, no matter how good the quality of the underlying evaluation, its value will only be realised if there are effective channels for communication and influencing to increase the likelihood that the results are used. Ensuring evaluation is influential is of course every bit as important as ensuring that it is soundly based. This applies to both traditional and developmental approaches.

The role of evaluation in the programme and project management cycle is discussed more fully in the Annex, with particular reference to the system and practice in the United Kingdom, but also drawing on experience elsewhere. Starting from the role of evaluation in the policy cycle, the Annex discusses the various uses of evaluation evidence and suggests some principles for the effective use of evaluation findings and ways of organising the evaluation framework. The principles for more effective practice in the use of evaluation in the policy cycle include:

- *Securing early stakeholder involvement* - involving key programme managers, analysts and decision makers from the outset.
- *Maintaining links with decision makers* – putting in place mechanisms to prevent evaluation becoming a box-ticking exercise and to promote close links between managers and evaluators.
- *Good communications planning* – clear and well communicated messages, effective dissemination, aimed both at key decision makers and more widely, in order to enhance credibility and stimulate public debate – this includes publishing the underlying data.
- *Requiring a management response* – putting the onus on senior managers to state publicly what they will do in response to evaluation findings.
- *Building good knowledge management systems* – making evaluation reports available in a structured way – there may be a role for the OECD in coordinating a common template for relevant meta-data.
- *Managing evaluation activities strategically* – taking a proportionate approach and focusing on key priorities, ideally established as part of an evaluation strategy.
- *Choosing the right evaluator* – striking the right balance between ensuring independence and lesson-learning.
- *Enhancing credibility* – maintaining professional and ethical standards and the methodological quality of the evaluation while recognising the need for an eclectic approach.
- *Fostering an evaluation culture* – working on the demand for evaluation and providing better training for commissioners, evaluators and users.
- *Creating institutions for embedding better use of evaluation* – a number of countries have created independent institutions with the formal role of collating and synthesising evaluation and other evidence for use in policymaking.

5. CONCLUSIONS AND SELECTED POLICY IMPLICATIONS

Based on the evaluation material reviewed here, a number of general lessons can be drawn about industrial policies, evaluation methods and evaluation as it relates to industrial policy. This concluding chapter draws out selected policy-relevant conclusions under four headings: a) specific policy areas; b) policy packages; c) new approaches to industrial policy evaluation; and d) achieving policy impact through evaluation.

5.1 Specific policy areas

5.1.1 Policies to support business R&D

Direct support for R&D

A policy mix of incremental R&D tax incentives and targeted direct support is likely to be most cost-effective. However, the associated administrative and compliance costs are likely to be higher than for volume-based tax credits and generic subsidies.

Focusing direct support on activities that either have high positive externalities or are prone to market failures can increase the additionality of public support. However, because direct support can be more targeted, the danger exists of distorting competitive market processes. The increasingly common practice of giving direct support to pre-competitive ventures, and to research partnerships, is one means of reducing problems associated with picking winners at the level of individual firms. By eliciting firm-level information and resources, the use of matching grants might help to identify higher-quality projects and lower total public outlays.

Good policy designs need to ensure competitive and transparent selection processes, and avoid favouring incumbents or providing opportunities for lobbying.

R&D tax credits

Chapters 1 and 2 of OECD (2013a) and a recent OECD Policy Note titled *Maximising the Benefits of R&D Tax Incentives for Innovation*[26]give detailed consideration to the design of R&D tax credits. Indeed, OECD (2013a) suggests that the design features of R&D tax credits are at least as important as their generosity. The key messages from those documents are summarised here.

- Incremental tax incentives may be more effective in inducing additional business R&D spending than volume-based tax credits. They are less costly to taxpayers because they are less likely to subsidise R&D that would have taken place without the tax credit. However, incremental schemes typically have higher compliance costs than volume-based schemes. Higher compliance costs are more likely to deter young and small firms.

- Incremental schemes are more complicated and costly to administer than volume-based schemes.

- Young firms typically experience financial losses in the early stages of an R&D investment. If R&D tax credits fail to provide immediate cash refunds and/or carry-over provisions, they may

provide less assistance to young firms, even if the scheme is otherwise generous. The use of payroll withholding tax credits for R&D wages, whereby firms receive an immediate refund for expenditure on the wages for R&D personnel, is another way to support (young) firms in a loss position.

- New OECD research suggests that R&D tax incentives benefit incumbents at the expense of entrants, thereby slowing the process of reallocating market share to more productive firms. But differences in the extent of direct support appear to have a more neutral impact as between incumbents and entrants.

- The effectiveness of R&D tax incentives also depends on the stability of the policy over time. When R&D tax policy changes often, the impact of the R&D tax incentives appears to be much reduced (Westmore, 2013).

- New OECD analysis – outlined in Chapter 2 of OECD (2013a) - has examined the effective tax rates that apply to investment in R&D by multinational enterprises (MNEs). A central insight in this work is that, when making R&D investment decisions, the calculus facing MNEs includes not just incentives that affect the marginal cost of R&D, but also the treatment of corporate income tax on the returns to the investment in R&D. The new research finds that overall tax relief for R&D by MNEs, when factoring in relief resulting from cross-border tax planning by MNEs, could well be greater than governments foresaw when their R&D tax incentives were designed. Countries may be losing tax revenue from the commercialisation of subsidised R&D and also foregoing some potential domestic knowledge spillovers associated with production. Furthermore, "stand-alone" firms that are not part of a multinational group of companies, and thus unable to adopt cross-border tax-planning strategies, may be placed at a competitive disadvantage, relative to MNEs, in undertaking and exploiting R&D. Careful attention to the design of R&D tax credits is essential to reduce these risks.

5.1.2 Innovation-oriented public procurement.

A number of challenges can be identified in the design and implementation of innovation-oriented procurement. These are summarised here:

- *The need to ensure competition*: as with traditional procurement, innovation-oriented procurement must avoid the risk of capture by vendors and/or other anti-competitive effects. For example, in some countries procurement procedures may end up giving preferential treatment to state-owned enterprises, which hinders competitive neutrality. At the same time, when special measures for SMEs or other groups are considered, these must fall within the framework of national competition policies as well as international standards and obligations. The challenge of avoiding anti-competitive effects will likely be more acute for pre-commercial procurement, as some interaction with suppliers of not-yet-existing products may be needed in formulating tenders that are technically feasible.

- *The need to ensure adequate capacity in the public sector for developing and implementing innovation-oriented procurement.* Procurement officials are increasingly requested to integrate innovation (and other) considerations in their purchasing decisions. When award criteria include considerations other than economic value, subjectivity in the decisions of procurement officials may arise. Particular expertise may be needed in the procurement to make such judgements. But recent work in the United Kingdom, at least, found that only 14% of surveyed firms strongly agreed with the statement that "public procurers are knowledgeable about the market in which our product and/or service operates." Just 18% of firms strongly agreed with the statement "public

procurers are knowledgeable about the technical aspects of our product and/or service" (Edler et al., 2012).

- *The need to avoid sub-optimal scale in procurement bodies:* General public procurement is often highly fragmented across local, regional and national government agencies. Sub-national governments account on average for around 64% of public investment in OECD countries. This can entail challenges in having requisite expertise across many small procurement bodies. In addition, a decentralised procurement system may also lack scale-efficiency and risk-mitigation possibilities open to more centralised systems.

- *The need to establish linkages to innovation agencies:* Many public procurement agencies operate separately from government agencies tasked with fostering innovation. Specialised procurement agencies are mainly responsible for the efficiency of purchasing, and expertise in the respective fields of innovation may be lacking. A number of studies suggest that innovation is often low on the list of priorities among procurement agencies (Technopolis, 2011). In addition to the need for legal expertise on procurement regulations at the various levels of decision-making, procurers will also need expertise on technologies and markets.

- *The need to manage specific risks associated with innovation-oriented procurement:* Procurement of innovation entails risks beyond those that arise in traditional procurement. These risks include:

  - Technological risk – that is, non-completion risk stemming from technical features of the procured good or service.
  - Risks related to the uptake by users of the good or service. These might stem, for instance, from incompatibilities with existing technologies or routines.
  - Market risks – these risks exist on the side of both supply and demand. On the demand side, risks are greatest for wholly novel items. Public bodies might mitigate such risk by implementing additional demand-side measures, such as user training schemes, or by bundling public demand. On the supply side, the main risk is that suppliers do not respond to the tender.

- *The need for suitable legislative frameworks:* Aspects of legislation might unnecessarily hinder innovation-oriented procurement. In some countries legislation has prohibited innovators – i.e. the suppliers of the original prototypes – from bidding for later procurement, as this might provide them with an unfair competitive advantage over their rivals (Technopolis, 2011). While no doubt aimed at ensuring competition, such a restriction could conceivably act as a disincentive to investment in innovation.

- *The need for additional evidence of the impact of procurement for innovation.* Few countries analyse public procurement to support innovation in rigorous ways, even if most countries collect basic data on a regular basis on the number of bids, contract awards and the use of open versus non-competitive procedures. Evidence from an OECD survey on public procurement also suggests that most countries do not explicitly consider the opportunity costs and potential risks when using procurement to support socio-economic objectives (OECD, 2012a). The expense of achieving these goals should be considered, and the trade-offs, if they exist, need to be made explicit – *e.g.* finding out whether procurement is a more cost-effective way to achieve innovation objectives than other innovation policies.

5.1.3 Capital market interventions, in particular support for risk capital.

Notwithstanding evidence from the United Kingdom of some successful equity schemes, the survey of evaluations undertaken suggests that the evidence of impact of these programmes is not conclusive. Policy seems to have been more successful where there is opportunity for experimentation and learning. Policy was also found to be more effective where it leveraged other private funding, hence the trend towards hybrid models. Another policy lesson is that policy makers need to address not just the supply of finance, which is the focus of most of the interventions, but also the demand side and wider framework conditions, suggesting that a mix of complementary policies is required. Finally, although there is scope for learning from other OECD countries, policy makers should be wary of imitation and in particular should not attempt to transplant schemes from one country to another without understanding the differences in institutional context and other framework conditions.

5.2 Policy packages

5.2.1 Sectoral approaches and public-private partnerships

This is an area where the evaluation evidence has proved elusive. Industrial policy by definition involves some attempt to influence the structure of the economy but the sectoral targeting may be either explicit or implicit and the range and mix of instruments used vary enormously. Sectoral approaches are therefore hard to distinguish from industrial policy in general. There are many examples of sector success stories (for example from Brazil, Chile, Japan, Korea) but little systematic evidence that sector performance is related to the degree of support or that the explicit identification of key sectors is a necessary component of modern industrial policy. Sectors, however, offer a natural conduit for pursuing new industrial policy and a sector focus may help create clarity of vision and an alignment of interests around shared objectives, provided the risks of capture by sectoral interests, of deadweight losses in public spending and market distortion can be contained.

Public-private partnerships are often the vehicle through which sectoral approaches are implemented. Despite the strong consensus to emerge from the new industrial policy literature on the importance of public-private partnerships and other institutional mechanisms for promoting dialogue between industry, government and the research base, it has proved difficult so far to find hard evaluation evidence on the impacts of key partnership approaches. The impacts of "soft" industrial policy interventions of this kind are by their nature harder to capture through evaluation, partly because the intervention itself is facilitative. The absence of econometric evidence does not necessarily invalidate conclusions drawn on the basis of more qualitative and subjective assessments. But this is an area where more evidence is needed, for example on the experience of some countries with sector councils. Recent initiatives in the Netherlands and the United Kingdom will be followed with interest and it is noted that there are plans in place to formally evaluate the Dutch experience with Top consortia for Knowledge and Innovation (TKIs).

5.2.2 Cluster policies and business networks

The discussion in Chapter 2 suggests that policy towards clusters and inter-firm networks should be distinguished. A programme of network development might be developed within and be supportive of a geographic cluster, but it will have different resource requirements and goals from many other measures that could strengthen a cluster (such as infrastructure development). Moreover, business networks can operate over a wide geographic scope, not being limited to the physical scope of a cluster. The earlier discussion suggested that it remains to be proven which, if any, cluster policy measures are effective, and to what extent they might increase innovation or productivity. Caution in policy development is therefore prudent. In this connection, the policy recommendations set out here also reflect the conclusions of policy

discussion on clusters and networks which took place at the first OECD Ministerial-level conference on SMEs and globalisation, in June 2000 in Bologna, Italy.[27]

Policy should explicitly target market failures. The fact that clusters can afford competitive advantages for firms does not in itself constitute a justification for public action. Furthermore, not all market failures merit correction through policy. Policy should be implemented only if there are good grounds for believing that it will produce outcomes superior to the "do nothing" situation.

Several forms of market failure are relevant. These include under-supply of public goods, and co-ordination failures.[28] Also important can be failures affecting small and medium sized firms more generally, such as in the supply of industrial real estate and, possibly, the provision of certain technical and financial services. Nevertheless, the fact that markets can fail in some of the above fields does not imply that they will be failing everywhere: indeed, many clusters have thrived in the absence of policy. A proper assessment of how well the relevant markets are functioning should be an integral part of policy formulation. In addition, by not explicitly identifying market failures a cluster development programme might simply become a source of interest-group support. Indeed, assisting firms to better act in concert could have the consequence of helping those same firms press for support that is economically unjustified.

A policy towards clusters should be based on government working with existing and emerging clusters rather than trying to create them *ab initio*. A policy aimed at developing entirely new groups of firms in selected sectors can entail high costs, high risks, serve as a screen for outmoded forms of industrial targeting, and give rise to destructive competition should many regions follow the same policies in pursuit of the same industries. Furthermore, there is almost no evidence of a significant cluster of firms having emerged anywhere as an intentional outcome of policy.

A policy on clusters should provide a framework for dialogue and co-operation between firms, the public sector (particularly at local and regional levels of government) and non-governmental organisations. This dialogue could identify and lead to the development of inter-firm networks (see below). Such a dialogue could also lead to an improved quality of policy and government action (such as in the provision of information, infrastructure supply and better decisions in co-locating complementary public investments with related concentrations of private investment).[29] Direct financial subsidy of firms' location decisions appears to have only modest effects while giving rise to multiple possible inefficiencies.

The content of the policy should reflect the type of cluster in question. For example, a cluster comprised of innovative small-scale firms that engage in collaborative ventures and access international markets will require very different policies from one for a cluster containing many large companies in traditional industries with limited inter-firm collaboration and weak competitive positions.

Policy-makers should assess the wider determinants of cluster success and judge whether these should be the preferred targets of policy. Uyarra and Ramlogan (2012) and others note that many policies that are likely to have a major impact on clusters are almost never analysed in depth in the context of cluster development. Such policies include, transport, land-use planning, housing, the quality of public amenities and labour market policies. Rather, the evaluations that have been performed focus on activities implemented under discrete time-bound initiatives.

Policy-makers should invest in better data. Better micro-level data are needed to identify the drivers of clustering, the sources of increasing returns to cluster development and specific market failures (indeed beyond a given scale, public subsidy of agglomeration might reduce overall welfare, although policy makers do not generally have the data with which to know when this occurs). Better micro-level data are also essential for policy evaluation.

Initiatives should be matched to the most suitable level of government. The ideal level of government will correspond to the physical scope of the cluster while having substantial influence over relevant programmes and expenditures. Initiatives should likewise be framed so that they do not invite deadweight losses from competition among subnational authorities.

Local adaptability should exist in university-industry relationships. Incentive structures should encourage local linkages to industry. Many institutional permutations are possible as regards the interaction of local firms, universities and training institutions. For example, university-industry partnership mechanisms can range from grants and fellowships to targeted research contracts, collaborative research and consortia agreements, training, mobility and networking programmes. In terms of functional goals such partnerships often seek to enhance the commercialisation and diffusion of technology, create enterprise spin-offs and support strategic research and technology objectives. A key consideration is that local flexibility should exist in the collaboration that educational and training bodies enter into with adjacent firms, whether in joint development of specialised courses and curricula – national curricula may be too slow to change and be unsuited to the technical specifics of particular enterprise agglomerations – the distribution of financial benefits from collaborative undertakings, or the precise forms of partnership in research and development.

If firms benefit from networks then it is natural to expect that they will invest in networks of their own accord. However, governments might justify a facilitating and/or co-ordinating role in network development owing to the fact that in some places and industries there may have been no, or limited, prior familiarity with the opportunities that networks afford (if "demonstration" is the policy rationale, then this implies that the policy should have short duration).

It is important to create informed demand for network services, with networks preferably addressing precise market-driven objectives. Successful business networks appear to organise around specific goals. Those goals should be set by, or in conjunction with, private actors. But lasting private sector engagement requires awareness of the benefits and opportunities of networks in order to increase informed demand for network services.

A degree of financial support, in feasibility work, start-up activities, and the costs of network brokerage, is to be expected. Aversion to and unfamiliarity with inter-firm co-operation, as well as problems of co-ordination, could create barriers to the spontaneous emergence of networks. Co-ordination problems among geographically dispersed firms may also create inertia in the establishment of business networks. Public action, at least in a catalytic role, may be needed. However, funding should be modest, and should be phased out as participants start to engage more formally and obtain benefits.

Policy must operate with realistic time-frames: a commitment of 3-4 years is usually required for a significant business network programme.

Experienced network brokers are essential. As with many schemes to support enterprise, the quality of management is critical. Persons with direct experience of SME development should be employed as network brokers, providing advice and a neutral corner for firms hesitant at the prospect of co-operation. Network brokers can also help allay concerns over loss of control and appropriation of benefits. Establishing broker teams and facilitating exchanges among them could help maintain effectiveness and motivation.

5.2.3 Industrial strategy

Many countries have experimented with, or are embarking on, some form of national industrial strategy and debate continues on the conditions for success and the respective merits of selective and

horizontal approaches. Increasingly industrial policy around the world may be focused on technologies, activities or tasks, or cross-cutting issues.

It is apparent, however, that there is a dearth of good practice and a desire for practical guidance on the problem of evaluating industrial strategies, particularly those involving a package of inter-related industrial policy measures. Part of the difficulty in drawing general lessons in this area is that industrial policy will inevitably be context-dependent and the policy approach will vary according to the country's stage of development, institutional features of the business environment and the specific characteristics of the sector being targeted.

Examples can be found in the literature of success stories of more general industrial strategy programmes in both OECD and developing countries (Rodrik, 2004; Bianchi and Labory, 2011; O'Sullivan et al, 2013). But systematic evidence is harder to come by and many economists remain to be convinced. New industrial policy puts less emphasis on product market support measures and more emphasis on support for technology and skills, public-private partnerships and a government role in facilitation and coordination. These approaches are newer and by nature harder to evaluate. The jury is still out.

Rodrik (2008) has attempted to distil some of the key lessons to be learned from the experience to date:

- *Embeddedness.* To design effective industrial policy, governments need to discover and elicit more information about the constraints that markets face, which typically requires close, strategic co-operation between the government and the private sector. Advisory councils and public-private partnerships are among the tools that governments can use.

- *Carrots and sticks.* Well-designed public support should help to amplify and build on market dynamics. Firms need to be able to earn rents from investment in innovation and entrepreneurship. But these rents should be temporary and capable of being eroded over time by competition.

- *Accountability* is essential. Accountability can take various forms, including: regular reporting on targets and achievements, together with explanations for deviations from planned outcomes; a high degree of openness in discussions between government and business; transparent public accounting of government support; and openness of such programmes to new entrants, as opposed to capture by vested interests.

The emerging consensus is that the risks associated with selective-strategic industrial policy can be minimised through a "soft" form of industrial policy, based on a more facilitative, coordinating role for government, with the emphasis on systems, networks, institutions and capabilities. Successful implementation of such an approach will require, even more than usual, well designed plans for regular monitoring and evaluation, especially of strategy and policy programmes. Ideally such plans should be put in place before the programme begins, so that data gathering and real-time tracking of progress against deliverables and outcomes can take place from the outset.

5.3 New approaches to industrial policy evaluation

Chapter 4 reviewed recent developments in thinking about industrial policy and evaluation and suggested ways in which they could be brought together. A distinction was drawn between four categories of problems and the associated evaluation challenges:

- *Simple* – typically, the use of industrial policy as a single instrument in conditions where the causal relationships are relatively well understood and best practice identifiable. There is still a need to establish 'what works' and the size of the expected policy response. Evaluators need to rely less on subjective self-assessment by beneficiaries of impact and do more to establish rigorous counterfactuals.

- *Complicated* – where a package of related measures is involved, with multiple instruments or multiple levels of governance. Policy responses could still be predictable in principle but the combination of different instruments and influences makes it difficult to be precise. There is a need for consistency of approach, a mix of methods and more careful use of the data and where frameworks and logic models are important. Policy responses could still be predictable in principle but the combination of different instruments and influences makes it difficult to be precise about attribution or overall impact.

- *Complex* – where uncertainty is high and the effect of behavioural interventions may be non-standard and highly context-dependent. In such circumstances, even in cases where single instruments are being used, the potential to generalise from past experience is limited. Iterative approaches to evaluation will need to be used and the design and implementation of policy adapted as experimentation and iteration reveal new knowledge about outcomes.

- *Complicated/complex* – typical of new industrial policy where a package of measures is being applied in an environment of uncertainty so the problems encountered in both the "complicated" and "complex" categories are present. Understanding the behavioural responses, system effects and interactions that determine policy outcomes may be even more important, but, in the face of uncertainty and complexity, the theory of change itself may need to evolve as new information emerges. In addition, the instruments used in new industrial policy tend to be "softer" and more focused on institutions than in more traditional sector-targeting industrial strategies. The evaluation challenge is at its most extreme here. Iterative approaches based on a structured sequence of evaluations may be a way forward.

In summary, the three main messages to emerge from thinking about the evaluation of industrial policy in this framework are:

- Evaluation of industrial policy has some catching up to do with other policy areas in the application of rigorous techniques. There are some legitimate reasons why rigorous evaluation is more difficult but for simple evaluation of single instruments, better understanding and better data now provide an opportunity for strengthening the evidence base.

- Central to new industrial policy and smart specialisation approaches is the notion of an "experimental state" which uses evaluation as a policy tool in developing modern industrial and innovation policy. Developmental evaluation approaches offer a way forward.

- Industrial strategy is inevitably more difficult to evaluate particularly where it involves "softer" forms of industrial policy, often involving dialogue and public-private partnerships. It should, however, be possible to devise a structured approach to testing hypotheses about industrial strategy and tracking progress, perhaps along the lines proposed by Matthews and White (2013).

The approach proposed by Matthews and White is of course highly generic. It does not pretend to be a 'reference model' for industrial policy evaluation in the sense being developed by TAFTIE for four innovation policy instruments, or by the European Commission for State Aid. The OECD's Expert Group

on the Evaluation of Industrial Policy discussed a number of frameworks (for example, Andreoni, 2013) but the group's conclusion was that it is too difficult to draw up a single framework for industrial policy evaluation, given the multitude of approaches, and it would be unhelpful to be too prescriptive.

5.4 Achieving policy impact through evaluation

The work of the Expert Group has confirmed a relative lack of rigorous evaluation of industrial policy. 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. From the work of the Expert Group, as documented in this report and in the Annex, the following principles might be suggested as the basis of good practice in industrial policy evaluation:[30]

- **Make explicit, at the highest possible level, the commitment to evaluation of industrial policy.** There should be an explicit commitment to undertake ex post evaluation of significant industrial and innovation policy interventions and industrial strategy initiatives, adopting a proportionate approach and using appropriate methodologies. An overt recognition of the importance of evaluation, by senior policy makers and agency heads, is vital. Human and financial resources for evaluation are more likely to become available once such recognition is forthcoming.

- **Consider mandating evaluations when public funding is provided.** A key reason why various state and federally sponsored programmes in the United States have been thoroughly evaluated – such as the Manufacturing Extension Partnerships programme – is because mandatory evaluation requirements were attached to the use of federal government funds.

- **Insist on the development of data and evaluation strategies as a pre-requisite for the commencement of programmes.** A clear evaluation strategy for a programme should be established from the outset, with an ex ante evaluation plan which, to the extent possible, articulates the theory of change and shows the main channels of impact from inputs and activities through to impacts and outcomes. This should be accompanied by a data strategy to ensure that the data necessary for evaluation are collected from the outset. Governments also have a duty to make more data available in order to assist researchers, "armchair evaluators", and future historians to assess policy effectiveness. There could be a role for the OECD in coordinating the release of data in an internationally comparable format or in collating or promoting the exchange of data.

- **Choose the evaluation technique in the light of the size and nature of the programme concerned.** Studies of major programmes – especially pilot schemes that could be rolled out later at larger scale – should use a variety of methods: random assignment, quasi-experimental assessments, interviews with beneficiaries, participatory approaches involving all stakeholders. There should be a move to more use of randomised experiments and, where appropriate pilots, as the basis of *ex post* impact assessment. However, assessments of outcomes using experimental

evaluation approaches are expensive – for which reason they should best be applied to large programmes or to pilot or small initiatives which are likely to be rolled out at larger scale over time. They are also data intensive and require particular statistical expertise. Studies of smaller programmes might employ other methods.

- **Evaluating industrial policy requires an eclectic approach mix of methods.** At the level of industrial strategy, a mix of evaluation methods is likely to be needed – 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.

- **In the face of complexity, or when outcomes are uncertain, consider the approach of developmental evaluation.** Evaluation can play a role in policy development which is very much in line with modern industrial policy thinking. It needs to be recognised that time lags may make it difficult to track the variables of ultimate interest in the short term and there will be a need to find proxies which can give useful information. An iterative approach can be used, with periodic evaluations to improve knowledge of policy impacts and fine-tune programmes. Evaluation is not just about policy impact – process evaluation and the learning and development role of evaluation are important as well.

- **Insist on full disclosure in evaluation reports. There should be a commitment to public diffusion of evaluation findings of publicly funded programmes.** The choice of methods and evaluation parameters used, methodological drawbacks and areas of subjective judgement should be described in full. There should be a commitment to transparency and early publication of evaluation findings and the data on which they are based. Published evaluation findings should be accompanied by relevant meta-data that will facilitate online searches. The OECD could play a role in establishing a common format for such meta-data.

- **Robust governance mechanisms are needed to ensure evaluation is objective and free of political influence**. If in-house staff administers surveys of programme beneficiaries, especially of clients they have worked with directly, there is a heightened risk of response bias. External consultants may not give programme managers objective advice if they fear it may affect funding for future projects. Programmes should be evaluated by, or in collaboration with, genuinely independent experts, possibly from an Audit Office. Ideally, the body that implements the evaluation would work with programme managers but would not be dependent on continued contracts from the sponsor of the programme.

- **Good mechanisms for policy learning are needed to ensure that the findings of evaluation feed back into future policy making.** While accountability and audit are important, mechanisms for lesson-learning must also be strong. This applies to both traditional evaluation and developmental evaluation. The Annex sets out some principles for establishing good lines of feedback and communication and some desirable features of the institutional setting that would help foster a more positive culture for the evaluation of industrial policy.

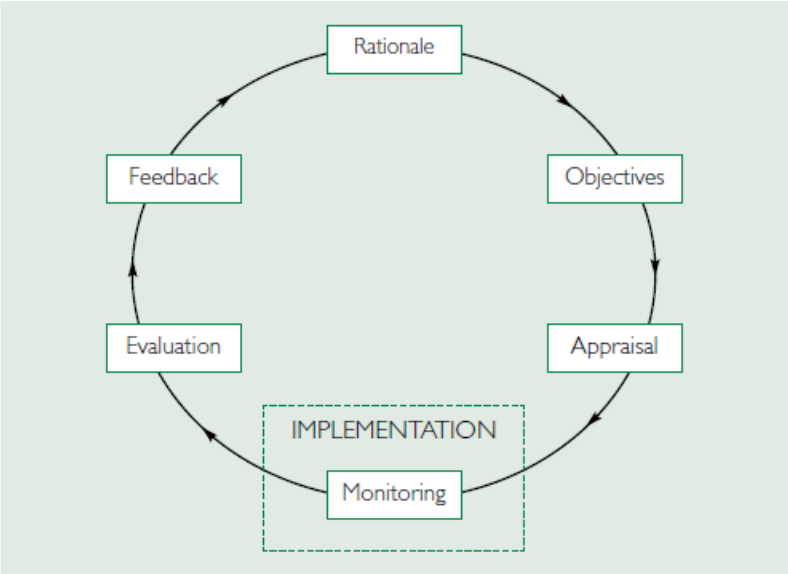
ANNEX – EVALUATION IN THE POLICY CYCLE

This Annex discusses the role of evaluation in the programme and project management cycle, with particular reference to the system and practice in the United Kingdom, but also drawing on experience elsewhere. Starting from the role of evaluation in the policy cycle, it discusses the various uses of evaluation evidence and suggests some principles for the effective use of evaluation findings and ways of organising the evaluation framework. It concludes with a discussion of how an evaluation culture can be fostered and the scope for creating institutions for embedding better use of evaluation. Ensuring evaluation is influential is of course every bit of important as ensuring that it is soundly based. This applies to both traditional approaches and to developmental evaluation.

a) Role of evaluation in the policy cycle

Figure A1, drawn from the guide to evaluation used by the United Kingdom Government (HM Treasury, 2011), sets evaluation in the context of the policy making cycle. As the diagram illustrates, appraisal, monitoring and evaluation are integral parts of the broad policy making and delivery cycle, sometimes known as the ROAMEF (Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback).

Figure A1. Role of monitoring and evaluation in the policy cycle



Source: HM Treasury, 2011

Frameworks similar to this are used in a number of countries (for example, GAO, 2012) and will be familiar to users in other social science and science disciplines. Engineers and managers will recognise the similarity to the Plan- Do-Check-Adjust method, also known as the Deming cycle (Deming, 1986), an iterative four-step management tool used in business for the control and continuous improvement of processes and products. Another variant is the Test-Learn-Adapt methodology, which focuses on

understanding better what works and continually improving policy interventions to reflect learning from evaluation (Haynes et al, 2012). In a sense, these are all just applications of the "scientific method" whose main components may be characterised as part of a cycle of hypothesis, experimentation and evaluation.

Key to the effective operation of this cycle is that evidence generated from monitoring and evaluation activity can be used to build on the analysis undertaken at the appraisal stage by testing and refining the assumptions made. This applies not just at the conclusion of the programme in order to learn lessons for new policies and projects. In developmental evaluation, it can also be as part of a process of iterative and experimental policy development. Or, more generally, it can be to feed evidence back throughout the whole policy and delivery cycle to help track progress and offer real time identification of any barriers that could risk inhibiting successful delivery. The cycle can operate at different "clock-speeds", but in all cases, effective management systems are needed to ensure that the processes for feedback from monitoring and evaluation are robust.

b) Uses of evaluation evidence

The Magenta Book used in United Kingdom Government circles identifies four ways in which evaluation evidence can be used:

- Immediate decisions about policy options; for example whether to roll-out a pilot as a national or local programme;
- Longer term decisions about the policy/programme; for example informing Government Spending Reviews and the future scale of investment;
- How the programme/policy could, or should, be improved; for example if the evaluation identifies major flaws; and
- How future policies should be designed and implemented.

Examples can be found around the world of good practice in each of these applications. For instance, in the Netherlands, the design of the innovation voucher scheme evolved through various pilot programmes whose results were evaluated. In the latest United Kingdom spending round, capital spending allocations to departments were in part based on the evaluation evidence for past programmes presented to the Treasury. As an example of how policies have been improved, the early monitoring and evaluation of one of the United Kingdom's access to finance interventions in response to the economic and financial crisis (an export insurance guarantee product) led to the correction of a design flaw in order to ensure that the market failure was properly addressed.

Sometimes the link to future policy development may not be immediately obvious. In the case of industrial policy, formal and informal evaluation contributes to the evidence base and to the body of corporate knowledge. Policy-makers draw on this in making decisions about the amount of resources to devote to industrial policy, how much to devote to each programme, priority sectors, technologies or activities and individual projects. In these processes, and notwithstanding the simple representation shown in Figure A1, the use of evaluation evidence should not be seen as a linear or closed-loop process, with implementation of a policy being followed by monitoring, evaluation and feedback into the next stage of policy making. This has been characterised (Lavis et al, 2003; Davies, 2012) as the *instrumental use* of evidence, as distinct from the *conceptual* use, described as "using research results in less specific, more indirect ways, than in instrumental use". The distinction is akin to the debate in innovation policy between the linear model and the innovation systems approach. Both the instrumental and conceptual uses of

evidence are to be distinguished from the *symbolic* use which is more about using research results to legitimise and sustain pre-determined positions.

Sadly, there would appear to be a dearth of good examples of influential industrial policy evaluation. Partly this reflects some genuine difficulties in undertaking industrial policy evaluation, but there is also more that could be done to ensure that evaluation in this area is more successful.

c) Some principles for more effective use of evaluation

Although there are a number of obstacles to undertaking industrial policy evaluation and applying the latest techniques, it is nevertheless important to ensure that such evaluation that is carried out is as influential as possible. No matter how good the quality of the underlying evaluation, its value will only be realised if there are effective channels of communication and influencing in place to ensure that the results are used. Based on experience around the world, from industrial policy and other policy spheres, it is recommended that that the following principles should be followed in order to promote the more effective use of evaluation evidence[31]

Securing early stakeholder involvement

If evaluations are to have influence, stakeholders, including senior officials and analysts responsible for policy and programme development, should be involved from the outset. Various approaches can be deployed in order to secure staff ownership and buy-in. Staff involved in policy decisions on industrial policy and resource allocation can be appointed to evaluation committees or involved through steering or advisory groups. Staff involved in project design can be brought into the planning, selection, design and guidance of decentralised evaluations. The determination of evaluation outcomes can be designed as a learning process through the use of participatory evaluation methods and dialogue with users and staff. Such mechanisms can not only improve the evaluation but also create consensus and ownership for a change process. However, care should be taken to avoid the risk of capture and to ensure the effort involved is not disproportionate.

It helps build stakeholder confidence and secure buy-in if evaluations are timely and focused on priority issues for key stakeholders. There is some value in "quick-wins" – evaluations that can be conducted quickly and economically and that provide information on issues of immediate concern. However, there may be a trade-off between speed and quality. Stakeholders must also be persuaded of the need for broader and more complex evaluations that take longer to complete but may be of higher quality and result in longer term impacts on the design of future projects and policies.

Maintaining links with the decision-making process

Evaluation should not be a box-ticking exercise or an end in itself. It should be relevant and address issues that are significant for political, budgetary, management and other strategic reasons. The proposed use of evaluation should be clearly defined at the outset and the evaluation approach should be tailored to the purpose of the policy intervention and the objectives of the evaluation. Building requirements for evaluation into policies from the start, and defining the objectives, will improve the usefulness of evaluation and facilitate planning and data collection.

It can also help institutionalise an impact evaluation system with strong buy-in from key stakeholders if there is a powerful central government champion who can help make the case for undertaking evaluation, check that it is well executed, and make sure that the results get used. Finance Ministries and Planning Departments have an important role to play in this, for example by insisting that evaluation evidence is brought forward before decisions are made on future resource allocation.

Where relevant, the evaluation plan should be agreed with key policy makers to maximise the likelihood that evaluation results will be applied to policy development. It will include an agreement on how and when interim and final results will be disseminated and fed back into the policy cycle. Particular attention should be placed on project and policy cycles to identify entry points for use of evaluation results. It should include the explicit identification of the mechanisms and processes for feeding the results back into policy and programme design.

Good communications planning

Clear and well communicated messages increase the likelihood that evaluation studies will have impact. It helps when the evaluation results point to clear policy implications. A good evaluation will have a limited number of well-focused conclusions and recommendations that flow logically from the analysis and findings. The recommendations should be realistic, relevant and preferably prioritised, distinguishing the most important and best-evidenced recommendations from those that are more difficult and/or controversial.

Reporting of the evaluation outcomes may need to be targeted to the intended users and to specific audiences. In more complex cases, this may require both a technical annex to establish credibility with an expert audience and a non-technical summary to bring out the main conclusions for policy makers and the wider public.

Effective dissemination of evaluation findings is also essential. Even the most clearly written evaluation will be of no value if it lies on the shelf and is never seen by future decision makers. This can be avoided if there is continuous and targeted communication with the intended users to build interest and confidence and to ensure there are "no surprises" at the time of the final report. There is also lesson-learning value in regular communication and dialogue through, for example, one-to-one briefings, workshops and short update and briefing papers.

Publication of evaluation reports increases credibility, stimulates wider dialogue and debate and creates pressure to act upon findings. It also helps build trust and confidence in the process if evaluation findings are presented in positive language that focuses on overcoming problems rather than assigning blame.

Placing evaluation data in the public realm, possibly in anonymised form, could facilitate (cost-free) academic analysis and serve as a form of evaluation quality control. Furthermore, in the pluralistic policy systems that characterise OECD countries, evaluation should be expected to serve the knowledge needs of a wider set of actors beyond programme managers and public policy officials.

Requiring a management response

The requirement for senior managers to publish a formal response to evaluation findings is one way to ensure that evaluation studies are not simply shelved or ignored. In the United Kingdom, the Department for International Development (DfID) already requires a formal management response for all of its evaluations. Managers are required to set out agreement or disagreement with each key finding, conclusion and recommendation, give reasons and state what action will be taken. DfID is committed to ensuring that management responses are completed, published and followed as part of an annual Directors' Statement of Assurance. As part of a refresh of its evaluation strategy, the Department for Business, Innovation & Skills (BIS) is also undertaking to publish a departmental response alongside evaluation reports in order to help complete the feedback loop to policy (Figure A1).

Going beyond this, senior management should also be given the opportunity to reflect more generally on the lessons from evaluation evidence – this can be encouraged through regular opportunities for board level discussion, whether as part of a formal evaluation strategy or in the context of wider strategic planning.

Building good knowledge management systems

Knowledge management systems are no substitute for good communication but can be an important adjunct to it. Access to a structured database of evaluation findings in a computer-searchable format will make it easier for analysts and policy makers to make more use of evaluation findings in policy and programme development.

In order to make it easier to locate relevant evaluations, good systems for meta-data are essential. This is important within a country and across a policy area. In the United Kingdom, the Department for Business, Innovation and Skills (BIS) is creating a depositary of evaluation evidence. The main funding bodies for scientific research in the United Kingdom also collate information on research outcomes, including evaluation of impact, through an online system and are working to increase the degree of commonality in the way that information on research outcomes is gathered across the research community (Box A1).

Going beyond national initiatives, use of a common metadata template across countries would make it far easier to track down information on the various evaluations of industrial policy that have been carried out worldwide. A standard online template for the required metadata (for example, policy area, instruments, precise focus of the evaluation, methods used, links to substantive evaluation materials) would facilitate online searching and enable users to do their own sorting, archiving and assessment of evaluation material. An internationally standardised metadata template would be even more effective as a way of making evaluations more accessible and there may be a role for an organisation like the OECD to coordinate agreement on a suitable template.[32]

Box A1. Example of a knowledge management system for evaluation

In the case of R&D funded by the Research Councils in the United Kingdom, RCUK, the strategic partnership for the UK's seven Research Councils, asks researchers and/or institutions to submit information about their research through an online system. Most of the research councils use the Research Outcomes System (ROS), a web-based system which collects data on the outputs, outcomes and impacts of research, such as:

- publications
- other research outputs such as new materials, exhibitions and websites
- staff development
- collaborations and partnerships
- communication and dissemination activities
- summaries of impact.

During 2013, the Research Councils undertook the first phase of a project to increase the degree of commonality in the way that research outcomes are gathered, in response to feedback from the research community and as part of a wider move to greater harmonisation across the Councils.

Source : http://www.rcuk.ac.uk/

There are however some pitfalls in making evaluation results more easily available. Systems of metadata can only provide an imperfect filter. Important evaluation studies may be missed, either because they have not been classified, or misclassified, or because they are carried out by institutions other than Government. Moreover, for those evaluations included, it should be noted that the findings of even quite rigorous evaluations may not always be applicable more widely. Evaluation findings are often context-dependent, varying, for example, with time, place, institutions, stage of the economic cycle, and other policy settings. Meta-evaluation studies and other forms of systematic review are also required in order to identify broader lessons. This is particularly important for longer term decisions about spending allocation, priorities for support and the strategic direction of industrial policy.

d) Organising the evaluation framework

In addition to the principles set out above for making more effective use of evaluation, it is important to follow good practice in the way that evaluation is organised and the institutions and frameworks used in the public sector. Based on good practice around the world, it is recommended that the following principles are followed:

Managing evaluation activities strategically

Resources available for evaluation will be used more effectively if a strategic approach is taken, focusing on the major areas of spend and tracking the indicators and impacts that will guide strategic decision-making on issues such as the implementation of major projects, future policy priorities, and the allocation of funding. It is good practice to set out an evaluation strategy that will make clear what the evaluation priorities are, describe the framework, and promote greater openness, awareness and credibility of the evaluation activities. Box A2 gives an example of an evaluation system for R&D spending in Spain. Several government departments in the United Kingdom now publish their evaluation strategy (for example, BIS, 2010; DfID, 2013; DfT, 2013).

It may be appropriate in some circumstances to systematise and institutionalise evaluations in key policy areas where the cost of collecting data is high and information limited. However, it will often be the case that a more flexible approach will produce better results and prevent evaluations from becoming academic exercises. Special attention should be given to the evaluation of *complicated* interventions that cut across many organisations. Central government agencies play an important role in managing the evaluation process; however, the actual evaluations can be decentralised to different actors at all levels of government. Ideally, major programmes should have funding earmarked for *ex post* evaluation as part of the programme budget. However, special funds for financing evaluations available from 'the centre' (in-house strategy units, co-ordinating agencies, or Departments of Finance) can serve as an important incentive for evaluating public policies.

The main purpose of the industrial policy evaluation system should be to provide evidence on what works to help inform future policy design and strategic economic policymaking. In some countries, however, evaluation systems tend to put more emphasis on transparency and accountability in fund allocation and expenditure rather than on lesson-learning for strategic economic policymaking and development of the national industrial and innovation system. Some feedback from evaluation into policymaking may be built into the overall system, but the design of the evaluation process generates mainly descriptive content, rather than a focus on economic impact and the implications for the strategic direction of policy. While such systems may have value in demonstrating to external stakeholders that money is well spent, they may be less effective in ensuring that policy learning takes place.

Choosing the right evaluator

One of the choices that has to be made in setting up an evaluation framework is between internal and external evaluators and their degree of independence. Evaluation can be carried out by those responsible for managing the programme, by central teams within the funding department (for example, central analysis, audit or strategic planning teams), by other Ministries with an oversight function (for example, Finance or Planning Ministries) or by external bodies (for example research institutes, management consultants or academics).

Comparing these types of evaluation, OECD (1998) concluded that:

> *"Self evaluation by an organisation is appropriate when the main objectives are organisational learning and improved implementation. However, the time and skills of staff may be insufficient, the range of issues covered may be limited and the credibility of findings may also be questioned. Evaluation by central management agencies is appropriate when the objective is improving budget priorities and when it is important that the evaluator has close links with decision-making processes. Evaluation by external evaluators is appropriate when the objective is to provide new perspectives on public policies or when there is a need for specialised evaluation skills. However, these evaluators may have limited understanding of the substance and the culture of the evaluated policy or organisation and offer theoretical evaluations."*

One difficulty in choosing the right evaluator is the potential tension between the use of evaluation for accountability and its use for lesson-learning. Independent evaluation is more appropriate when the objectives are to improve accountability and transparency. However, experience suggests that policy managers and other decision-makers may be reluctant to accept the findings and recommendations. Lesson-learning – the closing of the feedback loop between evaluation and policy development– may be easier to achieve if there is an evaluation culture embedded within the organisation. Managing the tension to ensure that evaluation advice is both objective and influential is one of the hardest challenges in the management of *ex post* evaluation. A good evaluation framework will require a combination of in-house and external evaluation, with checks and balances to ensure that in-house evaluations have credibility and that the organisational culture is open to accepting the results of independent evaluation.

While external evaluation is in many circumstances a good discipline, it should be noted that it would be wrong to equate external evaluation with independent evaluation. It would be reasonable to regard external evaluation undertaken by academics or research institutes, independently funded and subject to peer review, as independent. But there is anecdotal evidence that some external consultants feel under pressure to produce evaluation results in line with what they think the commissioning body wants to hear. One government agency in an OECD member country has brought its evaluation work back in-house precisely because they felt that they were not getting objective advice from their external evaluators.

Enhancing credibility

Whether evaluation is done in-house or externally, those who commission evaluation of industrial policy need to appreciate that building and maintaining a reputation for sound, objective and well-focused evaluation is an important objective. Lack of credibility will seriously undermine the use of evaluation findings, no matter how robust the underlying framework. Factors influencing credibility include the competence and authority of the evaluator, mutual trust between the evaluator and those evaluated, regular consultation with stakeholders and good processes for communicating findings.

Maintaining professional and ethical standards and the methodological quality of evaluation (encompassing issues such as relevant criteria, adequate data and evidence and reliable and clear findings)

also have an effect on the credibility of evaluation. Evidence of good practice in *ex post* evaluation of industrial policy, for example in the form of true economic impact evaluation or the use of state-of-the-art experimental methods, is one way to build credibility. Robust systems for quality assurance and open and frank dialogue can also improve credibility by exposing and rectifying potential weaknesses in evaluations.

Fostering an evaluation culture

The government has the most important role to play in fostering an evaluation culture that encourages innovation and adaptation to a changing environment. The basic message should be that, to stay relevant, organisations need to continue learning from feedback about results. Support for evaluations is demonstrated through leadership from the top and in particular the willingness of politicians, policy managers and central management agencies (for example central research Ministries and/or the Ministry of Finance), to make effective use of policy advice generated in evaluations.

Demand for evaluation needs to be generated, specified and articulated by internal and external stakeholders. Evaluations without "ownership" by stakeholders are unlikely to have impact. Institutional barriers to evaluation such as internal resistance can be reduced through consultation, aimed at creating mutual trust. Evaluations are also more likely to be influential if the culture is favourable to evidence based policy making in general.

Box A2. An Evaluation Tool for R&D in Spain

Integral Monitoring and Evaluation System (SISE): Monitoring instrument for public policies on R&D in Spain

The Integral Monitoring and Evaluation System (SISE) was introduced in Spain in 2008 as a tool for controlling the management of public funding RD&I programmes, making them more transparent and giving the general public and Spanish society a better understanding of the activities being financed with public funds.

The SISE was integrated into the Spanish National Plan for RD&I 2008-2011 as a mechanism for following up and evaluating research and innovation policies, incorporating the ex-post evaluation of the results of R&D programmes into existing processes for reviewing current activities and identifying the need for new initiatives.

The main goals identified for the system were:

- To establish a logical framework capable of linking, on a rational, global and hierarchical basis, the goals set out in the National Plan with the instruments identified for achieving them.
- To put in place a monitoring and evaluation system that covers administrative aspects as well as technical and strategic aspects.
- To identify a battery of indicators for each instrument that enables a simple analysis to be made of the degree of achievement of the goals established and assessment of the outcomes and their impact on the Spanish science and technology system.

For *ex post* evaluation, the plan called for the final outcomes of projects to be incorporated into the SISE tool in computerised form, enabling immediate online use, with the aim of applying criteria of productivity, additionality, efficiency and effectiveness to science and technology policy, and elaborating a Results Evaluation Programme, known as PROEVAR.

Analysing general indicators of the National Plan, INGENIO 2010 and the National Strategy for Science and Technology, as well as of the evaluation of the activities started up under them, the system was designed to help define and put forward proposals for new activities and funding and to provide mechanisms for the ad hoc supply of information needed by decision-makers.

Source: Fernandez de Labastida (2008).

Finally, it is important to have good provision of training and opportunities for professional dialogue for evaluators and for the commissioners and users of evaluation. Competent evaluators, well-informed commissioners and enlightened and enthusiastic users all contribute to a positive evaluation culture.

e) Creating institutions for embedding better use of evaluation

The creation of independent institutions with the formal role of collating and synthesising evidence and evaluation for use in policymaking is one way of embedding better use of evaluation more formally into the policy making cycle. There are a number of models for this around the world, with varying degrees of independence, for example the Productivity Commission in Australia, the National Bureau of Economic Analysis (CPB) in the Netherlands, the Agency for Growth Policy Analysis in Sweden, the new What Works Centres initiative in the United Kingdom and the Washington State Institute for Public Policy in the United States.

In Australia, the Productivity Commission is the Government's independent research and advisory body on a range of economic, social and environmental issues. The Productivity Commission was created as an independent authority by an Act of Parliament in 1998, to replace the Industry Commission, Bureau of Industry Economics and the Economic Planning Advisory Commission. Its role is to help governments make better policies, in the long term interest of the Australian community. The Commission's research on productivity and quantification of barriers to trade is drawn on internationally, with its analysis of regulatory policy and applications also attracting particular interest. The Commission has recently been exploring the use of a structured hypothesis testing approach in its work on the evaluation of the impact of regulations on business and the wider economy.

In the Netherlands, the Dutch National Bureau of Economic Analysis (CPB), established in the late 1940s, is tasked both with macroeconomic forecasting and the evaluation of government and opposition policies. Although formally part of the Ministry of Economic Affairs, the CPB have over the years earned a worldwide reputation for independence and objectivity.

The Swedish Agency for Growth Policy Analysis has a mission to develop a "knowledge infrastructure" for decision in central Government on growth policy-measures for sustainable growth in all parts of Sweden[33]. An independent body within the Ministry of Enterprise, Energy and Communication, it has been commissioned by the Government to carry out evaluation and analysis of Swedish growth policy and to be responsible for strategic intelligence on international economic developments. It manages a database containing both corporate and individual data that can be used for evaluation and regularly publishes its own analytical reports and evaluation studies on aspects of growth policy in Sweden.

The United Kingdom government established in 2013 a network of "What Works Centres" covering a diverse range of social and economic policy areas, including local economic growth. Together these centres cover over GBP 200 billion of public spending and are intended to be a source of robust and high quality synthesis of the research evidence on the effectiveness of interventions in each field. Box A3 sets out the core functions for these centres.

In the United States, the Washington State Institute for Public Policy, has since the 1990s, been directed by the Washington State legislature to identify well researched and evidence-based public policies that can, with a high degree of certainty, lead to better policy outcomes and a more efficient use of taxpayer dollars.[34] While its emphasis is more on social programmes rather than on industrial policy, the Institute is much admired for its marshalling of evaluation evidence and the publication of benefit-cost results in a form designed to facilitate comparison across programmes.

Box A3. What Works Centres in the United Kingdom

Each What Works Centre will be independent of Government, with a clear and relevant policy focus. Each will:

Generate evidence synthesis

- Undertake systematic assessment of relevant evidence and produce a sound, accurate, clear and actionable synthesis of the global evidence base which:

    1. assesses and ranks interventions on the basis of effectiveness and cost-effectiveness;
    2. shows where the interventions are applicable;
    3. shows the relative cost-effectiveness of interventions; and
    4. shows the strength of evidence on an agreed scale.

Translate the evidence

- Produce and apply a common currency for comparing the effectiveness of interventions.
- Put the needs and interests at the heart of its work

Evidence absorption

- Publish and disseminate findings in a format that can be understood, interpreted and acted upon.

Promote good evidence

- Identify research and capability gaps and work with partners to fill them.
- Advise those commissioning and undertaking innovative interventions and research projects to ensure that their work can be evaluated effectively.

Source : https://www.gov.uk/what-works-network.

NOTES

1 More generally, Cook et al. (2013) assess the distribution of benefits from three different programmes in the United Kingdom: an area-based technology support programme, an export support programme and a regional European Commission programme. They find that in all three cases the top 20% of businesses surveyed accrued well over 90% of the benefits.

2 Monge-González and Rodríguez-Álvarez (2013) also present an evaluation of two business support programmes in Costa Rica: PROPYME and CR Provee. The first sought to increase innovation in SMEs, and the second aimed to develop backward linkages between SMEs and multinational companies operating in Costa Rica. The impacts of each programme were measured in terms of real wages, employment demand and the probability of exporting. Fixed effects and propensity score matching techniques were used to correct for selection bias. Both programmes were found to have positive impacts on SME performance. However, firms that received both programmes simultaneously experienced a higher increase in average real wages than those that only received the CR Provee programme.

3 For instance, at the outset, an open and transparent competition could be arranged for solution designs for a particular problem, with the preferred designs being retained. In a second phase, a competitive tender would take place for prototype development for the preferred designs, and the most promising prototypes retained. In the final phase, the most promising prototypes would compete through the development of a test series. At this stage the procurer would possess information allowing technical comparison between the new product/service and the existing (or alternative) procured product/service. However, the new product/service would still have to be rolled out commercially at sufficient scale in order to compete in a new procurement of the final product/service (and this commercial roll-out might not be undertaken by the same enterprise that won the pre-commercial procurement).

4 The OECD's Directorate for Science, Technology and Industry is currently engaged in an EC-funded project aimed at assessing the quality of existing data sources on innovation procurement and setting out methodologies for measuring this form of procurement.

5 Rothwell (1984) and Geroski (1990) held that under some circumstances innovative public procurement could be more effective than subsidies in stimulating private investments. The point is perhaps best argued by Geroski (1990), but the empirical evidence is not systematic.

6 Evaluation of loan guarantee schemes and other forms of support for lending to SMEs is not covered in detail in this paper but has been the subject of related work in the OECD Working Party on SMEs and Entrepreneurship (WPSMEE).

7 Under the ZIM programme, SMEs receive grants for application-oriented R&D and innovation projects of between 35-50%, depending on the size of the SME and its location. Although not an equity scheme, ZIM is included here as an example of an evaluation of a financing intervention established in response to the 2008-09 crisis.

8 See Warwick (2013) for a fuller discussion of some of the definitional issues in the industrial policy debate.

9 Examples can be found of sector approaches where firms are randomly selected within the sector, for example innovation credits for the creative industries sector (Bakhshi et al, 2011), discussed in Chapter 4 below.

10 Industrial policy in Japan is a mixture of horizontal and selective policies. Each ministry evaluates its own policies, and this is reflected in budget allocations, but the evaluations are mainly qualitative. Lechevalier et al (2010) is one of the few independent evaluations using rigorous techniques. Recently, however, there

has been a greater focus on evaluation by academic researchers. The Japanese policy think tank, RIETI (the Research Institute for Economy, Trade and Industry) has a programme of research on the experience of industrial policy.

11 See, for example, Low et al (2005) for an evaluation of one such scheme.

12 The study by Lechevalier et al (2010), reviewed above, is an example.

13 Devlin and Moguillansky (2009) briefly review the experience of ten countries (Australia, the Czech Republic, Finland, Ireland, Korea, Malaysia, New Zealand, Singapore, Spain, and Sweden) with national-level strategic public-private alliances aimed at long term economic development.

14 To different degrees, and varying from one location to another, the economic drivers of cluster formation in particular industries can include, among other variables: *i) proximity to markets*. Despite low-cost international transportation, being near to markets can be important, especially for products that are not easy to transport or that require continuous interaction with customers; *ii) supplies of specialised labour*, such as occur around many universities; *the presence of input and equipment suppliers*. A high frequency of exchanges between capital goods producers and users has been pointed to as an important underpinning of the innovative performance of firms in many industrial districts; *iii) the availability of specific natural resources*; *iv) economies of scale in production*. Such economies may allow only a small number of efficient-scale plants in a given market; *v) the availability of infrastructure*. Some types of infrastructure may also be quite specific, such as with certain transport or tourist facilities, further encouraging agglomeration; *vi) low transaction costs*. When firms and their suppliers operate near to each other, and the frequency of interaction is high, the costs of negotiation and contract enforcement might be reduced. This effect may be reinforced by social norms affecting entrepreneurs belonging to overlapping social groups; and *vii) superior access to information*.

15 And the opportunity for financial intermediaries to acquire detailed understanding of firms in a given cluster, on account of frequent exchanges with enterprises of a similar type, may well facilitate efficient lending.

16 Rosenfeld (2001) notes that: networks allow firms access to specialised services at lower costs, whereas Clusters attract specialised services to a region; networks have restricted membership, whereas Clusters have open "membership"; networks are based on contractual or other forms of agreements, whereas Clusters are based on market dynamics; networks make it easier for firms to engage in complex production, whereas Clusters generate demand for more firms with similar and related capabilities; networks are based on cooperation, Clusters require competition.

17 Clusters are also dynamic and can decline. For instance, Porter (2008) notes, the manufacture of golf equipment in the United States shifted from New England, where clubs were built from steel and wood, to California when the use of advanced and synthetic materials became a possibility.

18 Furthermore, data from the United Kingdom suggest a low responsiveness of firms to subsidies aimed at encouraging location in low-income regions. Devereux et al (2007) show that a grant of GBP 100 000 increases the probability of a firm relocating to such a region by just 3%.

19 In a related manner, Criscuolo et al (2007) show that in the United Kingdom the policy of Regional Selective Assistance had positive effects on employment and investment in firms in lagging regions, but no impact on productivity. The authors consider that subsidy for less efficient firms could hinder aggregate productivity growth by slowing reallocation from less efficient plants.

20 TAFTIE (The European Network of Innovation Agencies) is a group of 28 organisations from 25 European countries who collaborate and share best practice in the implementation of national technology programmes. TAFTIE Task Forces operate for a maximum of two years, during which they prepare reports to the TAFTIE Board and wider membership on their results and conclusions.

21 Denmark, Finland, the Netherlands, Norway, Sweden and the United Kingdom.

22 No attempt was made, however, to measure spillovers to other firms.

23 Impacts on Manchester-based creative services businesses were not tracked.

[24] The distinction follows Rogers (2008).

[25] Guerzoni and Raiteri (2012) discuss the challenge of "hidden treatment" in evaluation when there are multiple treatments in place, specifically the role of R&D tax credits, other R&D subsidies and innovation-oriented public procurement and possible interaction effects.

[26] See: http://www.oecd.org/sti/maximising-the-benefits-of-r-d-tax-incentives-for-innovation.pdf.

[27] http://www.oecd.org/cfe/smes/englishfrenchitalianversionsofbolognaconferenceproceedingsandworkshoppapers.htm.

[28] Michael Porter, in his book *On Competition*, provides an example of how various players in an industry may need to upgrade their operations simultaneously if overall productivity gains are to be had: "In the wood products cluster, for example, the efficiency of sawmills depends on a reliable supply of good-quality timber and the ability to maximise the utilisation of timber in either furniture (highest quality), pallets and boxes (lower quality) or wood chips (lowest quality). Portuguese sawmills suffered from poor timber quality because landowners would not invest in timber management. Hence most timber was processed for use in pallets and boxes, a lower value use that limited the price paid to landowners. Substantial improvement in productivity was possible, but only if several parts of the cluster changed simultaneously. Logging operations, for example, had to modify cutting and sorting procedures while sawmills had to develop the capacity to process in more sophisticated ways. Co-ordination to develop standard wood classifications and measures was an important enabling step. Such linkages can be recognised and captured more easily within clusters than among dispersed participants."

[29] As an example, some clusters in the United States have become a forum for targeting education and training programs. Regional Skill Alliance grants from the United States' Department of Labor have funded projects that organise firms in regions to enable them to better identify common skill sets and standards, to reduce training costs, and to learn from one another. Some regional institutions of higher education have concentrated their resources on local industry clusters.

[30] Some of these principles are developed more fully in the Annex on Evaluation in the Policy Cycle.

[31] This section and the one that follows are based on OECD (1998) and also draw on World Bank (2006, 2009) and DfID(2013).

[32] The template could be designed just for industrial policy, or for industrial and innovation policy, or for all evaluations and reviews conducted by, and for, governments. In the latter case, industrial policy per se would be just one sub-set of the metadata. A more general template is more likely to maximise learning opportunities and would offer much greater scope to maximise the returns on the investment, albeit modest, required to develop such a template. On the other hand, it may be that a generic template would be more difficult to agree and/or be insufficiently flexible for the needs of a particular policy area.

[33] http://www.tillvaxtanalys.se/en/home/about-growth-analysis.html.

[34] http://www.wsipp.wa.gov/BenefitCost.

REFERENCES

Andreoni, A. (2013), "Industrial Policy Evaluation Frameworks in the Future High Value Manufacturing Landscape", presentation to the OECD Expert Group on Evaluation of Industrial Policy, 15 January 2013.

Aschhoff, B. and W. Sofka (2009), "Innovation on Demand: Can public procurement drive market success of innovations?", *Research Policy*, Vol. 38, No. 8, pp. 1235-1247.

Audretsch, D, B., A.N. Link, , and J.T.Scott (2002), "Public/private technology partnerships: evaluating SBIR-supported research", *Research Policy* 31(Vol. 1), pp.145-158.

Bakhshi, H., Edwards, J. Roper, S. Scully, J. and D. Shaw (2011), "Creating Innovation in SMEs: Evaluating the Short-term Additionality of the Creative Credits Innovation Vouchers Pilot", London: NESTA.

Bakhshi, H., A. Freeman and J. Potts (2011), "State of Uncertainty - Innovation policy through Experimentation", *NESTA Provocation No. 14*, April 2011.

Banarjee, A. and E. Duflo (2011), "Poor Economics: A radical rethinking of the Way to Fight Global Poverty", Public Affairs Books, New York City.

Beason, R. and D. E. Weinstein (1996), "Growth, Economies of Scale, and Targeting in Japan (1955-1990)", *Review of Economics and Statistics*, 78:2 286-95.

Beaudry, C. and S. Breschi (2000), "Does 'Clustering' Really Help Firms' Innovative Activities?", http://www.cespri.uni-bocconi.it/papers/wp111.PDF.

Bellego, C. and V. Dortet-Bernadet (2013), "The French cluster policy and the R&D spending of SME and intermediate-sized enterprises", Document de travail N° G2013/06, INSEE, April 2013.

Bérubé, C. and P. Mohnen (2009), "Are firms that receive R&D subsidies more innovative?", *Canadian Journal of Economics*, Vol. 42, pp. 206-225.

Bianchi, P. and S. Labory (2011), "Industrial Policy after the Crisis: Seizing the Future", Edward Elgar, Cheltenham, United Kingdom.

Binks, J. (2006), "Using Public Procurement to Drive Skills and Innovation", Local futures.

BIS (2010), "Evaluation Strategy; The Role of Evaluation in Evidence-based Decision-making", Department of Business, Innovation and Skills, London, August 2010.

Bloom, N., R. Griffith and J. Van Reenen (2002), "Do R&D Tax Credits Work?", *Journal of Public Economics*, Vol. 85, pp. 1-31.

Boekholt, P. (2014), "Benchmarking effectiveness & impact of innovation instruments, Towards a reference model for evaluation", presentation to OECD Expert Group on Evaluation of Industrial Policy, 16 January 2014.

Bound, K. and R. Puttick (2010), "Buying Power? Is the Small Business Research Initiative for procuring R&D driving innovation in the UK?", *NESTA Research report*, National Endowment for Science, Technology and the Arts, United Kingdom.

Cappelen, Å., A. Raknerud and M. Rybalka (2012), "The Effects of R&D Tax Credits on Patenting and Innovations", *Research Policy*, Vol. 41(2), pp. 334-345.

Copeland, A. and J. Kahn (2013), "The Production Impact of 'Cash-For-Clunkers': Implications for Stabilization Policy." *Economic Inquiry* 51, no. 1 (2013).

Cook, J., B. Macdonald and R. Pates (2013) "The Pareto Principle: The Importance of the Vital Few in Business Support Programmes" SQW Working Paper. 9 (http://www.sqw.co.uk/insights-and-publications/the-pareto-principle/).

Cowling, M., J. Siepel, G. Murray, P. Nightingale (2014) "Evaluation of UK Government Equity Schemes". Forthcoming, Department for Business Innovation and Skills, London.

Criscuolo, C., R. Martin, H. G. Overman and J. Van Reenen (2012), "The Causal Effects of an Industrial Policy", *CEP Discussion Paper*, No 1113, January 2012.

Criscuolo, C., R. Matin, H. Overman and J. Van Reenen (2007), "The Effect of Industrial Policy on Corporate Performance: Evidence from Panel Data", Discussion paper available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.785&rep=rep1&type=pdf.

Cunningham, P., A. Gök and P. Laredo (2013), "The Impact of Direct Support to R&D and Innovation in Firms", *NESTA Working Paper*, No.13/03.

Czarnitzki, D. and g. Licht (2006), "Additionality of public R&D grants in a transition economy: The case of Eastern Germany", *Economics of Transition*, 14, pp. 101-131.

Czarnitzki, D., P. Hanel and J.M. Rosa (2011), "Evaluating the Impact of R&D Tax Credits on Innovation: A Microeconometric Study on Canadian Firms", *Research Policy*, Vol. 40, Issue 2, pp. 217–229.

Czarnitzki, D. and H. Hottenroot (2011), "Financial Constraints: Routine Versus Cutting Edge R&D Investment", *Journal of Economics and Management Strategy*, Vol. 20(1), pp. 121-157.

Danish Agency for Science, Technology and Innovation (2011), "The impacts of cluster policy in Denmark – an impact study on behaviour and economic effects of Innovation Network Denmark". Available at: http://en.fi.dk/publications.

Davies, P. (2012), "The state of evidence-based policy evaluation and its role in policy formation", *National Institute Economic Review*, No. 219, ppR41-R52, January 2012, NIESR London

Deming, W. E. (1986), *"Out of the Crisis",* MIT Center for Advanced Engineering Study. ISBN 0-911379-01-0.

Den Hertog, P. (2014), "Evaluation of the Innovation in Dialogue approach: experiences from sectoral innovation programmes in the Netherlands", presentation to OECD Expert Group on Evaluation of Industrial Policy, 16 January 2014.

Department of Enterprise, Trade and Employment (2009*)*, "Using Public Procurement to Stimulate Innovation and SME Access to Public Contracts: Report of the Procurement Innovation Group", Dublin, Ireland.

Devereux, M.P., R. Griffith and H. Simpson (2007), "Agglomeration, Regional Grants and Firm Location", *Journal of Public Economics* 91, pp 413-435.

Devlin, R. and G. Moguillansky (2009), "Public -private alliances for long-term national development strategies", *CEPAL Review* 97, April 2009.

DfID (2013), "International Development Evaluation Policy", Department for International Development, London, May 2013.

DfT (2013), "Monitoring and Evaluation Strategy", Department for Transport, London, March 2013.

Donselaar, P. (2013), "Evaluation of the WBSO scheme: Ministry of Economic Affairs, The Hague, Department of Innovation and Knowledge", presentation to the OECD Expert Group on Evaluation of Industrial Policy, Paris, 15 January 2013.

Duranton, G. (2011), "California Dreamin': The Feeble Case for Cluster Policies", *Review of Economic Analysis*, Volume 3, pp. 3-45.

Edler, J., L. Georghiou, E. Uyarra and J. Yeow (2012), "Public procurement and innovation in an evolutionary framework: evidence from public sector suppliers", *Presentation delivered at the conference 'Demand, Innovation and Policy Undeprinning Policy Trends with Academic Analysis'*, Manchester Institute for Innovation Research, 22-23 March 2012.

Engel, D., T. Mitze, R. Patuelli and J. Reinkowski (2012), Does Cluster Policy ATrigger R&D Activity? Evidence from German Biotech Contests. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1988305.

Estonian Ministry of Economic Affairs and Communications (2013), *Annual Report 2013.*

European Commission (2010), "Communication "Europe 2020 Flagship Initiative Innovation Union", *COM (2010) 546 final.*

European Commission (2010a),"An Integrated Industrial Policy for the Globalisation Era Putting Competitiveness and Sustainability at Centre Stage, *COM(2010) 614 final* .

European Commission (2011), "Horizon 2020 – The Framework Programme for Research and Innovation", *COM(2011) 808 final.*

European Commission (2011a), "Single Market Act – Twelve levers to boost growth and strengthen confidence", *COM(2011) 206 final* .

European Commission (2013), "Evaluation in the field of State Aid: draft methodological guidance paper - concepts and recommendations", November 2013.

Fabiani, S., G. Pellegrini, E. Romagnano, and L.F. Signorini (2000), "Efficiency and Localisation: the Case of Italian Districts", in M. Bagella and L. Becchetti (eds.), *The Competitive Advantage of Industrial Districts: Theoretical and Empirical Analysis*, Heidelberg: Physica-Verlag.

Falck, O., S. Heblich and S. Kipar (2010), "Industrial Innovation: Direct Evidence from a Cluster-Oriented Policy", *Regional Science and Urban Economics*, 40(6). pp 574-582.

Falk, R, J. Borrmann, N. Grieger, E. Neppl-Oswald and U. Weixlbaumer (2009), "Tax Incentive Schemes for R&D, Part 4 of the Evaluation of Government Funding in RTDI from a Systems Perspective in Austria, Vienna: Austrian Institute for Economic Research. Available at: http://www.bmvit.gv.at/service/publikationen/innovation/forschungspolitik/downloads/systemevaluierung/report4.pdf.

Feldman, M.P., and M.R. Kelley (2003), "Leveraging Research and Development: Assessing the Impact of the U.S. Advanced Technology Program", *Small Business Economics*, Vol. 20, pp. 153-165.

Fernandez de Labastida, J.M. (2008), presentation to OECD Symposium on international comparison of the budget cycle in R&D and innovation policies, Madrid, July 2008. http://www.oecd.org/gov/budgeting/symposiumoninternationalcomparisonofthebudgetcycleinrdandinnovationpoliciesjuly2008.htm.

Foray, D. (2012), "Smart Specialisation and the New Industrial Policy Agenda", Innovation for Growth *Policy Brief* N° 8, European Commission.

GAO (2012), "Designing evaluations: 2012 revision" (GAO-12–20–8G). United States Government Accountability Office, Washington, DC.

Garcia-Quevado, J. (2004), "Do public subsidies complement business R&D? A meta-analysis of the econometric evidence", Kylos 57, pp.87-102.

Gayer, E. and E. Parker (2013), **"**Cash for Clunkers: An Evaluation of the Car Allowance Rebate System", *Economic Studies at Brookings*, 31 October 2013.

Geroski, P.A., (1990), "Procurement policy as a tool of industrial policy", *International Review of Applied Economics* 4 (2), S.182–S.198.

Glennerster, R. (2012*),* "The Power of Evidence: Improving the Effectiveness of Government by Investing in More Rigorous Evaluation", *National Institute Economic Review* No. 219, ppR4-R14, January 2012, NIESR London.

González, X., J. Jaumandreu and C. Pazó (2005), "Barriers to Innovation and Subsidy Effectiveness", *RAND Journal of Economics* 36, pp. 930-950.

Guellec, D. and B. van Pottelsberghe de la Potterie (2003), "The Impact of Public R&D Expenditure on Business R&D", *Economics of Innovation and New Technology*, Vol. 12(3), pp. 225-243.

Guerzoni, M. and Raiteri, E. (2012), "Innovative public procurement and R&D Subsidies: hidden treatment and new empirical evidence on the technology policy mix in a quasi-experimental setting", Working

Paper 18/2012. Department of Economics and Statistics Cognetti de Martiis & BRICK, University of Turin.

Haegeland, T. and J. Moen (2007), "Input additionality in the Norwegian R&D tax credit scheme", reports 2007/47, Statistics Norway 2007

Hall, B.H. and J. van Reenen (2000), "How Effective are Fiscal Incentives for R&D? A Review of the Evidence", *Research Policy*, (29), pp 449-469.

Hall, R. and D. Jorgenson, (1971), "Application of the Theory of Optimum Capital Accumulation", in Fromm, G., "Tax Incentives and Capital Spending", Brookings Institution, Washington, United States, 1971.

Haynes, L, O. Service, B. Goldacre and D. Torgerson (2012), "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials", Cabinet Office Behavioural Insights Team, London, June 2012.

Hjalmarsson, D. (2014), "Evaluating public-private partnerships – a multi-method approach to evaluate RTOs: the case of RISE", presentation to OECD Expert Group on Evaluation of Industrial Policy, 17 January 2014.

HM Treasury (2011), "The Magenta Book: Guidance for Evaluation", HM Treasury, London, April 2011.

Impact Evaluation Expert Working Group (2012), "Dare to Measure: Evaluation Designs for Industrial Policy in The Netherlands", November 2012 (Theeuwes Report).

Jaumotte, F. and N. Pain (2005), "Innovation in the Business Sector", *OECD Economics Department Working Papers*, No. 459, OECD, Paris.

Karjalainen, K. and K. Kemppainen (2008), "The involvement of small- and medium-sized enterprises in public procurement: Impact of resource perceptions, electronic systems and enterprise size", *Journal of Purchasing and Supply Management, 14*, pp. 230-240.

Klier, T. and J.M. Rubinstein (2013), "Restructuring of the U.S. Auto Industry in the 2008-2009 Recession", *Economic Development Quarterly* 2013 27: 144.

Köhler, C., P. Laredo and C. Rammer (2012), "The Impact and Effectiveness of Fiscal Incentives for R&D", *NESTA Working Paper* No.12/01.

Kuznetsov, Y. and C. Sabel (2011), "New Open Economy Industrial Policy: Making Choices without Picking Winners", World Bank PREMNotes, Economic Policy No. 161, September 2011.

Lavis, J. N., D. Robertson, J.M. Woodside, C.B. Mcleod, & J. Abelson (2003). "How can research organizations more effectively transfer research knowledge to decision makers?" *Milbank Quarterly*, 81(2), 221—248.

Lechevalier, S., Y. Ikeda and J. Nishimura (2010), "The Effect of Participation in Government Consortia on the R&D Productivity of Firms: A Case study of Robot Technology in Japan", *Economics of Innovation and New Technology*, Vol. 19, No. 8, November 2010, pp. 669-692.

Lee, J. (1997), "Government Interventions and Productivity Growth in Korean Manufacturing Industries", *Journal of Economic Growth*. 1:3 391-414.

Lentile, D. and J. Mairesse (2009), "A policy to boost R&D: Does the R&D tax credit work?", *European Investment Bank Papers*, Vol. 14, No1. 2009.

Lerner, J. (1999), "The government as venture capitalist: the long-run impact of the SBIR program", *Journal of Business*, 72.

Lokshin, B. and P. Mohnen (2012), "How effective are level-based R&D tax credits? Evidence from the Netherlands". Applied Economics, Volume 44, pp. 1527–1538 http://www.tandfonline.com/doi/abs/10.1080/00036846.2010.543083.

Low, C., D. Hulls and A. Rennison (2005), "Public Private Partnerships in Scotland: Evaluation of Performance", report prepared by Cambridge Economic Policy Associates for the Scottish Government.

Martin, P., Mayer, T. and F. Mayneris (2010), "Public support to clusters: A firm-level study of French "Local Productive Systems", *Regional Science and Urban Economics*, Vol. 41, Issue 2, pp. 108-123.

Martin, P., Mayer, T. and F. Mayneris (2011), "Spatial Concentration and Plant-level Productivity in France", *Journal of Urban Economics*, Vol. 69, Issue 2, March, pp. 182-195.

Matthews, M. and G. White (2013), "Faster and Smarter: Using a Hypothesis-Testing Approach to Policy and Programme Evaluation", Technical Memorandum prepared for the OECD Export Group on the Evaluation of Industrial Policy, 27 August 2013.

Mazzucato, M. (2013), "The Entrepreneurial State: Debunking public vs. private sector myths", London: Anthem Press.

Mian, A. and A. Sufi (2012), "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program", *The Quarterly Journal of Economics (2012) 127 (3): 1107-1142*.

Monge-González, R and J. Rodríguez-Álvarez (2013), "Impact Evaluation of Innovation and Linkage Development Programs in Costa Rica: The cases of PROPYME and CR Provee", *IDB Working Paper Series*, No.IDB-WP-461, Inter-American Development Bank, Washington DC.

Nishimura, J. and H. Okamuro, (2011), "R&D productivity and the organization of cluster policy: an empirical evaluation of the Industrial Cluster Project in Japan", *The Journal of Technology Transfer*, April 2011, Volume 36, Issue 2, pp 117-144.

OECD (1998), "Best Practice Guidelines for Evaluation", *PUMA Policy Brief No. 5*, May 1998, OECD: Paris.

OECD (2005), "Public/Private Partnerships for Innovation", in *OECD Science, Technology and Industry Outlook 2004*, OECD Publishing.

OECD (2012a), "Recommendation of the Council on Regulatory Policy and Governance", OECD Publishing, Paris. Available at: http://www.oecd.org/gov/regulatory-policy/49990817.pdf.

OECD (2012b), "Revised Terms of Reference for the TIP Activity on Innovative Public-Private Partnerships for Science and Technology", internal working document.

OECD (2013a), "Supporting Investment in Knowledge Capital, Growth and Innovation", OECD Publishing. doi: 10.1787/9789264193307-en.

OECD (2013b), "Scoping Paper on Strategic Public/Private Partnerships in Science, Technology and Innovation", internal working document.

Oh, Inh, Jeong-Dong Lee, Gyoung-Gyu Choi and Almas Heshmati (2006). 'Evaluation of Credit Guarantee Policy using Propensity Score Matching', Seoul National University, Korea.

O'Sullivan, E., A. Andreoni, C. López-Gómez and M. Gregory (2013), "What is New in the New Industrial Policy? A Manufacturing Systems Perspective", *Oxford Review of Economic Policy*, 29, issue 2, p. 432-462.

Parsons, M., and N. Phillips (2007), "An Evaluation of the Federal Tax Credit for Scientific Research and Experimental Development", Department of Finance, Government of Canada, September 2007.

Patton, M.Q. (2006) "Evaluation for the Way We Work", *The Nonprofit Quarterly*. Vol. 13 (1): 28-33. http://www.scribd.com/doc/8233067/Michael-Quinn-Patton-Developmental-Evaluation-2006.

Patton, M. Q. (2010), "Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use", Guilford Press, New York.

Porter, M.E. (2008), "On Competition", Harvard Business Press.

Potter, J. and G. Miranda (eds.) (2009), "Clusters, Innovation and Entrepreneurship", OECD, Paris.

Rigby, J. (2013), "Review of Pre-commercial Procurement Approaches and Effects on Innovation, Manchester Institute of Innovation Research". Available at: https://research.mbs.ac.uk/innovation/.

Rodrik, D. (2004), "Industrial Policy for the Twenty-First Century", *KSG Working Paper* No. RWP04-047, Cambridge MA: Harvard University.

Rodrik, D. (2008), "Normalizing Industrial Policy", *Commission on Growth and Development Working Paper* No. 3, Washington DC.

Rogers, P. (2008), "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions".

Rogers, P. (2011): "Evaluating Complex Interventions – Projects, Programs, Policies, Strategies". presentation at Seminar at the University of Johannesburg, Department of Public Governance, Johannesburg, South Africa, 17 February 2011.

Romani, M., N. Stern and D. Zenghelis (2011), "The basic economics of low-carbon growth in the UK", *Grantham Research Institute on Climate Change and the Environment Policy Brief*.

Rosenfeld, S. A. (2001), "Backing into clusters: Retrofitting Public Policies. Integration Pressures: Lessons from Around the World". Paper presented at John F. Kennedy School Symposium, Harvard University, March 2001.

Rosenthal, S.S. and W.C. Strange (2004), "Evidence on the nature and sources of agglomeration economies", in Henderson, V. and Thisse, J.F. (eds), *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam.

Rothwell, R. (1984), Technology-Based Small Firms and Regional Innovation Potential: The Role of Public Procurement, *Journal of Public Policy*, 4, pp. 307-332.

Starzyńska, W. and A. Borowicz, (2012), "The pro-innovative effect of Poland's public procurement system on the country's economy", *Paper presented at the conference "Demand, Innovation and Policy. Underpinning Policy Trends with Academic Analysis"*, Manchester Institute of Innovation Research, MBS, University of Manchester, 22-23 March 2012 .

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

Stiglitz, J. E., J.Y. Lin and C. Monga (2013), "The Rejuvenation of Industrial Policy". *World Bank Policy Research Working Paper* No. 6628, 1 September 1 2013.

Potter, J. and D. Storey (2007), "OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes", OECD: Paris

Swedish Agency for Growth Policy Analysis (2011), "Competent capital?-Three countries, three trials", published 15 November 2011, Stockholm.

Technopolis (2011), "Trends and Challenges in Demand-Side Innovation Policies in Europe", *Thematic Report 2011 under Specific Contract for the Integration of INNO Policy Trend Chart with ERAWATCH (2011-2012).*

Uyarra, E. and R. Ramlogan (2012), "The Effects of Cluster Policy on Innovation", *NESTA Working Paper* 12/05, March 2012.

Viladecans-Marsal, E. and J.M. Arauzo-Carod (2012), "Can a Knowledge-based cluster be created? The case of the Barcelona 22@ district", *Papers in Regional Science*, Volume 91 Number 2, June 2012.

Wallsten, S. (2000), "The effects of government-industry R&D programmes on private R&D: the case of the Small Business Innovation Research Program", *Rand Journal of Economics, Vol. 31, N°1 (Spring 2000).*

Warwick, K. S. (2013), "Beyond Industrial Policy: Emerging Issues and New Trends", *OECD Science, Technology and Industry Policy Papers*, No. 2, OECD Publishing. http://dx.doi.org/10.1787/5k4869clw0xp-en

Westmore, B. (2013), "R&D, Patenting and Productivity: The Role of Public Policy", *OECD Economics Department Working Paper*, OECD, Paris.

White, G. (2013), "Evaluation of Industrial Policy in a Context of Uncertainty and Complexity", presentation to OECD Expert Group on Evaluation of Industrial Policy, 16 January 2013.

Wilson, K. and F. Silva (2013), "Policies for Seed and Early Finance: Findings from the 2012 OECD Financing Questionnaire", *OECD Science, Technology and Industry Policy Papers*, No. 9, OECD Publishing. http://dx.doi.org/10.1787/5k3xqsf00j33-en.

World Bank (2006), Impact Evaluation and the Project Cycle, May 2006.

World Bank (2009), "Making Smart Policy: Using Impact Evaluation for Policy Making Case Studies on Evaluations that Influenced Policy", PREM Thematic Group on Poverty Analysis, Monitoring and Evaluation, June 2009.

Ziegler, N. (2012), "National experience with the evaluation of industrial policy: the case of ZIM", presentation to OECD Expert Group on Evaluation of Industrial Policy, 3-4 September 2012.