

# Evaluation

<http://evi.sagepub.com/>

---

## **Evaluation of Cluster Policy: A Methodological Overview**

Claudia Schmiedeberg

*Evaluation* 2010 16: 389

DOI: 10.1177/1356389010381184

The online version of this article can be found at:

<http://evi.sagepub.com/content/16/4/389>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



The Tavistock Institute

**Additional services and information for *Evaluation* can be found at:**

**Email Alerts:** <http://evi.sagepub.com/cgi/alerts>

**Subscriptions:** <http://evi.sagepub.com/subscriptions>

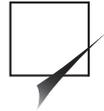
**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://evi.sagepub.com/content/16/4/389.refs.html>

>> [Version of Record](#) - Oct 18, 2010

[What is This?](#)



# Evaluation of Cluster Policy: A Methodological Overview

*Evaluation*  
16(4) 389–412  
© The Author(s) 2010  
Reprints and permission: sagepub.  
co.uk/journalsPermissions.nav  
DOI: 10.1177/1356389010381184  
<http://evi.sagepub.com>  


**Claudia Schmiedeberg**  
University of Hamburg, Germany

## Abstract

Cluster policy is increasingly becoming part of many governments' economic policy strategies. At the same time, evidence-based policy-making is gaining importance, bringing about a call for policy evaluation. Since the quality of the evaluation results depends highly on the method used, data, assumptions and techniques must be adequate for the specific evaluation question. This holds for cluster policy evaluation in particular, given the complexity and indirect nature of cluster policy interventions. This article provides an overview of evaluation methods suited to the ex-post analysis of cluster policy, covering both micro- and macroeconomic approaches.

## Keywords

innovation clusters; policy evaluation; regional policy

## Introduction

The concept of regional industry clusters has received increased attention over the past few years (see e.g. Lagendijk, 1999, for the development of the concept). Since the seminal work of Porter (1990) it has become popular as a lever to increase the competitiveness of regions, not only in industrialized countries (see Bachtler et al., 2005; OECD, 1999), but also in countries and regions lagging behind (Camagni, 1995; Rosenfeld, 2002; Schmitz and Nadvi, 1999). Accordingly, cluster policy is seen as a powerful instrument at the intersection between regional and industrial policy.

However, due to financial constraints, careful selection of where to spend the money is essential, which makes the evaluation<sup>1</sup> of policy measures necessary (for a literature overview on policy evaluation, see e.g. Georghiou and Roessner, 2000; Hansen, 2005). Regarding evaluation techniques, a large variety of tools is available, differing in terms of their rationale, complexity, data requirements and underlying assumptions; hence, evaluators face the task of choosing the appropriate method for a specific evaluation study (Foss-Hansen, 2005). In general, these considerations hold for each type of policy evaluation; but the complexity and the indirect nature of the interventions in the cluster policy approach pose particular difficulties and require adequate analytical methods. And while a body of literature exists which develops and proposes evaluation methods in general, a distinct evaluation concept or toolkit with focuses on cluster policy is not yet available.

---

## Corresponding author:

Claudia Schmiedeberg, University of Hamburg, Department of Economics, Chair of Economic Policy and Industrial Economics, 20146 Hamburg, Germany.  
Email: [schmiedeberg@econ.uni-hamburg.de](mailto:schmiedeberg@econ.uni-hamburg.de)

There is some literature on cluster policy evaluation with a focus on particular topics, like participatory evaluation (Angeles Diez and Esteban, 2000) or performance indicators (Arthurs et al., 2009). A different strand of literature (e.g. Learmonth et al., 2003; Fromhold-Eisebith and Eisebith, 2008) develops specific evaluation models. Raines (2002b), for instance, presents a comprehensive evaluation model for cluster policy covering the multiple dimensions of policy effects, but without going into methodological details.

The purpose of this article is, hence, to fill this gap by providing an overview of existing cluster policy evaluation methods. It discusses briefly the strengths and weaknesses of these methods and in which setting which method is an appropriate instrument. Due to the broad perspective, this overview is limited in depth regarding the methodological discussion, but it provides references to theoretical and empirical literature as well as to implementation manuals.

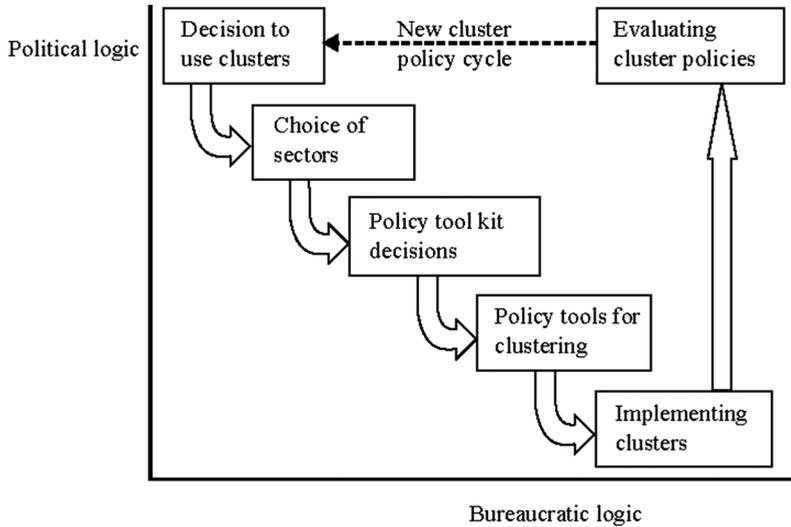
In the next section a definition of clusters and cluster policy and a basic understanding of evaluation are provided. Then, particularities of cluster policy evaluation as compared with other policy areas are discussed. The next section presents the evaluation methods, classified in five categories. The final section summarizes and concludes.

## Clusters, Cluster Policy and Evaluation

The definition of the characteristics and constituting elements of clusters exhibit considerable variation (Benneworth et al., 2003; Gordon and McCann, 2000; Martin and Sunley, 2003; for an overview of cluster definitions see Lublinski, 2002), and the lines between cluster policy and traditional industry, innovation, and regional policy are blurred (Boekholt and Thuriaux, 1999; Raines, 2002a). For the present discussion, a cluster is defined as a group of proximate firms 'interlinked by input/output, knowledge and other flows that may give rise to agglomerative advantages' (Lublinski, 2003: 454). This definition emphasizes the difference between clustering and co-location of firms, meaning that firms join a cluster intentionally in order to profit from spatial proximity, while co-location may be a result of external factors or contingent historic developments.

The advantages of agglomeration are stated in the literature (an overview is given e.g. by OECD, 2007), including in particular the existence of specialized suppliers and labour pools (Florida, 2002; Marshall, 1920; Porter, 1990), lower transaction costs (Krugman, 1991) and knowledge spillovers (Krugman, 1991; Malmberg and Maskell, 2002). From these factors, the competitive advantage of clusters is derived, but the cluster may in some cases require upfront support from policy. Based on these considerations, cluster policy is regarded as a promising approach to strengthen the innovative capacities of regional systems, leading to greater competitiveness of a region and its actors. Applying the hierarchical cluster concept (Litzenberger and Sternberg, 2005), cluster policy means upgrading the cluster from the mere agglomeration to a regional (innovation) system with beneficial implications for the clustered firms (see European Commission, 2002). The implementation of such policies has been demonstrated by a number of case studies (e.g. Styria (Hartmann, 2002), Basque Country (Aranguren et al., 2006)) and manuals (DTI, 2004; GTZ, 2007). On the other hand, as Formica (2003) argues, an interventionist cluster policy is not necessarily beneficial, due to the high potential of political institutions to have links with bureaucracy, patronage systems and unresponsiveness which increase transaction and compliance costs and potential inefficiencies. Besides, cluster policy can adopt a wide range of measures (as listed e.g. by Pfähler and Lublinski, 2003), while it is not definite which form of cluster policy is most effective. Hence, if and to what extent the measures actually are fruitful remains to be proven.<sup>2</sup>

The comprehensive literature on policy evaluation provides an overview of the reasons to undertake evaluation and a basic methodological understanding. The key evaluation question is:



**Figure 1.** Cluster Policy Cycle  
Source: Benneworth and Charles, 2001.

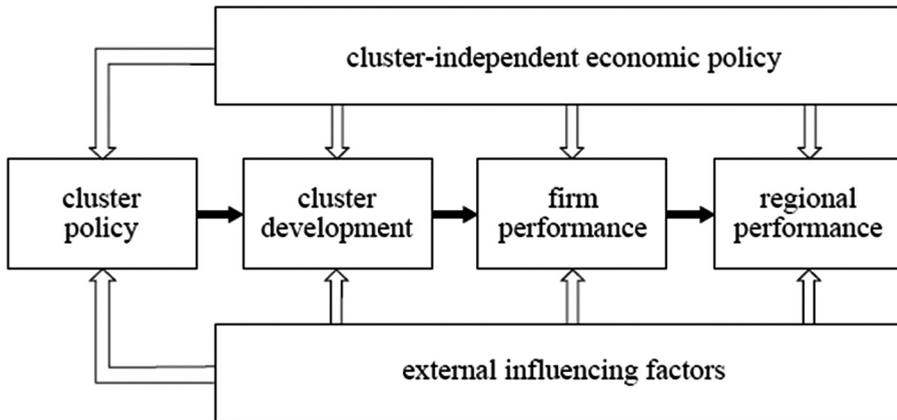
‘Does the policy programme work?’ The function of evaluation is twofold. First, it legitimates policy actions by proving their effects, and second, it deepens the understanding of the mechanisms of the measures supporting future decision-making (Guy, 2003). However, evaluation may refer to the effects of measures in various ways. Focusing on *effectiveness*<sup>3</sup> or *efficacy*, the evaluator analyses if the measures had a positive result. While efficacy aims at showing that in principle a measure is able to achieve the desired effect, an effectiveness proof shows that in practice the effect can be achieved by the measure. In contrast, the analysis of *efficiency* relates the effects of the policy actions to their costs. The latter perspective is the most comprehensive type of evaluation, but as it requires precise and detailed data both on outcome/output and input, it is not frequently applied.

## Challenges of Cluster Policy Evaluation

The main challenges of evaluation, such as the definition of a control group, the identification and measurement of effects and side effects, and the calculation of overall programme costs, are independent from the type of policy to be evaluated. But evaluation of cluster policy programmes faces particular challenges, due to the hybrid character of cluster policy which combines elements of various policy areas (Raines, 2002b), and the multidimensional, systemic concept of clusters. These are discussed in the following sections.

### The Organization of Evaluation

A first, rather practical question is how the evaluation study is organized. In general, there is a principal-agent problem in evaluation as evaluators are commissioned by the policy-makers and therefore are likely to try to satisfy the principal instead of being independent (Schilder, 2000). This holds for cluster policy evaluation as well. But in addition, cluster policy involves multiple actors from different political and administrative levels as well as several industries and regional entities. For instance, the aeronautics cluster in North Germany spans four federal states and a



**Figure 2.** Factors affecting economic performance of firms and regions

broad spectrum of related industries from cabin systems to maintenance and industrial design (Lublinski, 2002). Accordingly, cluster policy has to coordinate the decisions, actions and resource flows of several private and political partners, which are loosely organized without any clear hierarchy. Typically the coordinating entity will also conduct or commission the evaluation of the programme. But as the coordinator's authority within the non-hierarchical network of actors is limited, the evaluation might be hindered, for example, by differing data availability or lacking engagement in surveys. The structure of cluster policy challenges the evaluation in this regard, as the evaluators must involve and interact with at least the major stakeholders (Rossi et al., 2004).

### Defining Performance

The indirect nature of cluster policy interventions has been stressed previously (see Guinet, 2003: 158). The primary objective of cluster policy is not cluster formation but the assumed benefits of clustered firms in terms of the firms' performance and more generally in terms of regional performance, for example, increasing returns to scale and comparative advantages (see Buendia, 2005). This is illustrated by Figure 2.

Therefore the question arises on which policy outcome the evaluation should focus. Evaluation could investigate the development of the cluster, for example, the growth of the cluster or the number and intensity of interfirm connections, or on the region of the cluster concentrating on macroeconomic factors (e.g. employment rates); alternatively the focus could be on the individual firm whose performance should be strengthened by the development of the cluster. In addition, a distinction can be made between genuine economic indicators such as profit growth, productivity or (regional) GDP growth and technological (i.e. rather intermediate) indicators like R&D expenditure, patenting activities, collaborative agreements, etc. These decisions have to be made with care, as they will influence the evaluation results.

A further difficulty in this context is the definition of the cluster (see Raines, 2002b). It is characteristic for clusters that they integrate multiple industries and link different players, so that typically neither general industry classifications nor administrative regions are adequate to capture the boundaries of a cluster. The geographic proximity of the members, which is seen as a central feature of a cluster, entails considerable difficulties regarding definition and measurement (Litzenberger

and Sternberg, 2005). The drawbacks of a measurement based on secondary data have been highlighted by Lublinski (2003). He shows that a multitude of indicators must be used to define the boundaries of the cluster, but which still remain to a certain extent blurred. Besides, the boundaries might be evolving so that an evaluation based on a too rigid cluster definition might overlook a part of the development. Finally, as clustering is more than co-location, the evaluation must not be limited to measurements of firm density, industry-specific regional economic growth or the like; rather it should encompass the strength of agglomeration advantages to capture how the programme has changed the quality of the cluster.

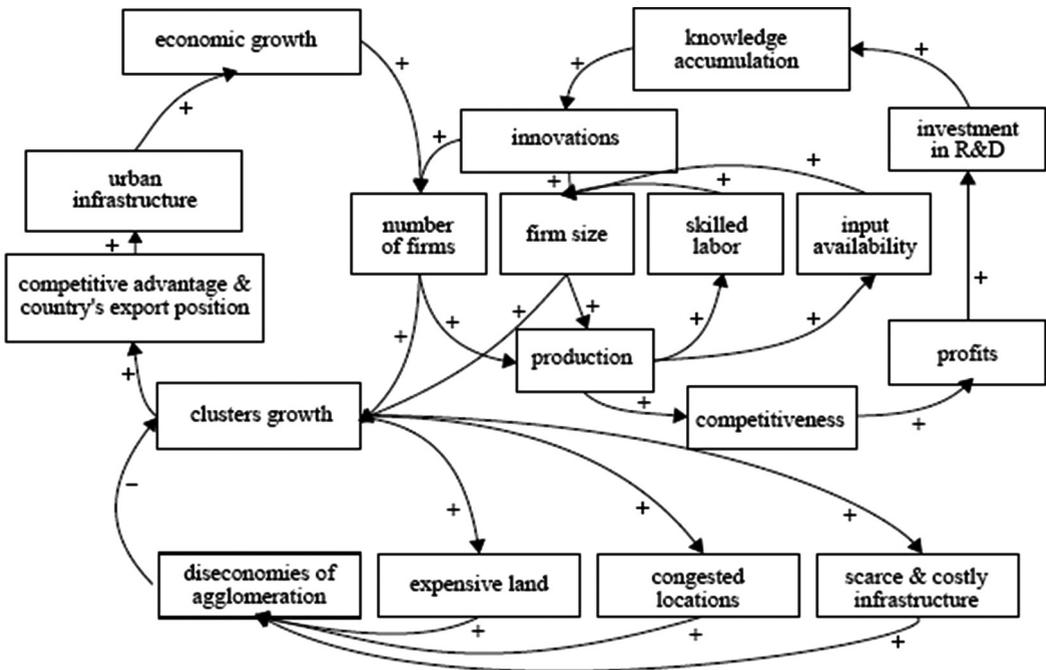
### *Attributing Impacts*

Cluster policy is mainly indirect (i.e. facilitative instead of pushing, according to Porter, 1998) and system oriented and typically not targeted to single projects or firms (Rip, 2003). What is more, a large number of different interventions are often combined, which increases the complexity of the policy approach (Boekholt, 2003: 257). For evaluation this poses particular difficulties. Buendia (2005) has captured the complexity of influencing factors and the mutual causality links in the context of cluster policy and performance (see Figure 3). The development of clusters is affected by a large number of (often unobservable) factors which are in many cases beyond the reach of cluster policy. Besides, cluster policy is embedded in a particular socioeconomic and institutional context, which must be taken into account (Angeles Diez and Esteban, 2000). Non-cluster policy instruments may influence the development of the cluster or the behaviour of the clustered firms, for example, through unintended effects of technology policy (Sternberg, 2003: 359). On the other hand, a policy intervention is likely to have several outcomes, both intended effects and unforeseen secondary actions (Schmidt, 1999). The effects may crop up only after some time (van der Linde, 2005: 29), so that the lag between policy intervention and policy impact may hide the causal link. Finally, cluster policy may not only affect the clustered firms but (e.g. via knowledge spillovers) also firms outside the cluster. Taken together, these factors lead to an attribution problem (Boekholt, 2003: 257), which requires adequate analytical (econometric) techniques and reliable data to detect causal relations which can be interpreted as actual impacts of policy tools.

### *Data Availability*

A third group of challenges relates to data restrictions which influence the choice of evaluation techniques. Particular limitations regarding cluster policy evaluation arise both due to the difficulties of measuring cluster performance and the regionally bounded perspective. On the one hand, the typical dimensions of clusters make macroeconomic modelling difficult; on the other hand, applications of microeconomic methods are confronted with small numbers of observations (if comparing e.g. similar clusters) and/or insufficient data quality. Data on innovation and network activities can be measured only partly, using indicators such as patent statistics, R&D expenditure or number of linkages (for an overview of innovation indicators, see e.g. Arthurs et al., 2009). To be applied in the cluster context, the data must contain detailed information on both industry and region of the firms and, for instance, their cooperation partners, which is not always the case in official statistics. Thus, evaluation must rely on imperfect data, which restricts the use of some methods.

Besides, the fundamental difficulty of analysing clusters in comparison to each other must be considered. Innovation clusters seem hardly comparable to each other, due to path dependencies and the cumulative nature of knowledge, but also as to the uniqueness of cluster structure, which makes the cluster a 'singularity in economic space' (Guinet, 2003: 154). As cluster policy seeks to



**Figure 3.** The Evolution of Industrial Clusters

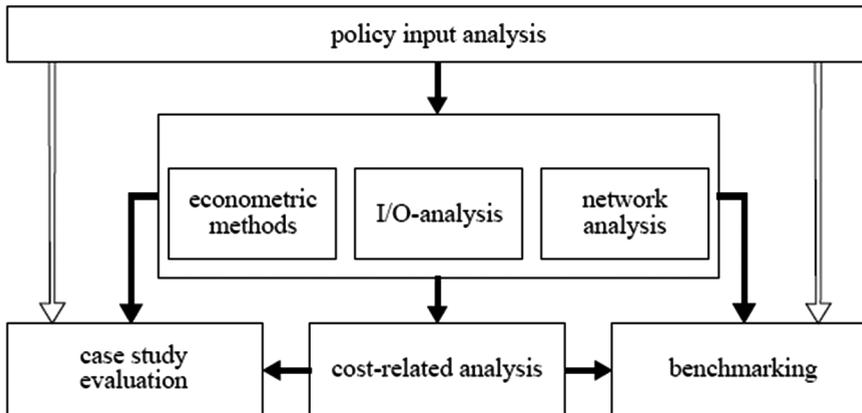
Source: Buendia, 2005

focus on high-potential clusters, ‘picking winners’ instead of spreading equally over all regions or even supporting lagging regions (see Cheshire, 2003), even greater peculiarities are likely. If cluster policy is well targeted its effects will only add on the above-average development that would have been expected without the policy intervention. Hence, evaluation must identify the additional effect of the intervention separate from the cluster’s innate potential, and comparisons to other clusters tend to have only limited analytical power.

## Evaluation Methods

In the following, a range of evaluation methods is presented. All of them can be applied to different research or evaluation questions, for example, in social and labour market policy or management, and typically their origins lie outside cluster policy. The discussion proceeds from rather simple, intuitive approaches to more complex and comprehensive methods. Figure 4 gives an overview of the methods and illustrates possible combinations of tools constituting an evaluation model similar to the evaluation model proposed by Raines (2002b).

A complete evaluation would encompass a qualitative and quantitative analysis of policy input, output and outcome on firm, cluster and regional level. But for the value of evaluation to stand in relation to cost and effort, typically only parts of this process will be realized. In practice, most evaluation studies rely on single methods, accepting thereby that results will be selective depending on the focus and the method applied. Accordingly, choosing the optimal evaluation method is crucial. The following discussion describes the possibilities and limitations of the methods included in Figure 4.



**Figure 4.** Evaluation Methods

### *Policy Input-oriented/Reporting Methods*

Policy input-oriented methods, or reporting methods, as they will be called in the following, though rarely mentioned in the academic literature, are frequently applied in various policy contexts, for instance education (e.g. US Dept of Education, 1999), social (e.g. TMSFG, 2008) or science and technology policy (e.g. European Commission, 2004). An example of cluster policy evaluation of this type is given by Aranguren et al. (2006). Instead of analysing policy impacts they report on the execution of the programme, including the chronological progress, faced difficulties and procedural failures as well as subjective perceptions of participating parties. Their main objective, hence, is accountability and transparency (see Rip, 2003) rather than enhancing the understanding of if and how the measures work.

Reporting methods can be based both on quantitative and qualitative data; often a combination of qualitative and quantitative information is used. Reporting is related to performance auditing, which starts with the question on the financial implementation of the programme, i.e. the correct and efficient commitment of funds (as defined e.g. in Council Regulation (EC) No 1260/1999, Art. 38). In addition to financial data, information on the structure and the components of the programme are reported, such as the timeframe of activities, description of operations, as well as indicators like the number of beneficiaries, handling time of applications. For a more comprehensive investigation an implementation analysis is included. It focuses on what was done when executing the programme, taking into account also qualitative information on how the measures were executed. For this purpose, official (quantitative) data are gathered as a documentation of the implementation process, for example, operations and activities (Nagarajan and Vanheukelen, 1997). Surveys and interviews for additional qualitative data can be targeted at customers (e.g. for customer satisfaction, as applied in Buchinger and Wagner, 2003), beneficiaries or the staff involved in programme implementation.

The appeal of reporting methods obviously lies in their practicability: attribution problems due to hidden causalities or inadequate control groups are unlikely. In addition, requirements regarding the evaluator's methodological capabilities and resources are moderate compared to output oriented techniques and, at least as regards auditing, data can be made available easily from the public accounting system. Also, reporting can provide immediate evidence (see Corbett and Lennon, 2003) by which it enables early intervention in case of a clear failure of the policy measures or their

implementation. However, reporting methods are not able to provide any information on the effectiveness of the programme, but only on the efficiency of the utilization of resources within the programme and the implementation. Therefore, they should be seen rather as a supporting tool. A number of criteria for the evaluation of cluster policy initiatives are discussed by Jappe-Heinze et al. (2008), who also demonstrate the value of preparatory reporting.

### Case Studies

Case study evaluation is a research strategy, rather than a method (Yin, 2003: 14). Being a rather open and flexible approach (see Ruegg and Feller, 2003: 34 ff.), case studies can involve several techniques both of a qualitative and quantitative nature ('triangulation', Jick, 1979). Qualitative/descriptive case studies aim to trace the historical process, i.e. the development of the analysed case, by the description, explanation and interpretation of data drawn from multiple information sources (see Ghauri, 2004). For cluster policy evaluation, the focus can be either on single participants of the programme or on the development of the entire cluster. Similar to qualitative case studies, benefit–cost case studies are based on multiple data sources. But in addition to the context analysis, the monetary effects and costs are to be estimated. For the quantification of costs financial business analysis methods are applied (which goes back to Mansfield et al., 1977), including measures such as benefit–cost ratio, net present value and rate of return (see Pelsoci, 2005). But in contrast to a full cost–benefit-evaluation, benefit–cost case study evaluation only analyses some exemplary cases from which conclusions on cost drivers and efficiency of the measure are derived.

Case studies are concurrently popular as well as highly controversial (see e.g. Flyvbjerg, 2004). Their appeal certainly lies in the intuitive understanding and plasticity of the results (see Ruegg and Feller, 2003: 34) as well as in the fact that case studies provide a comprehensive in-depth analysis of the subject in its context (Ghauri, 2004). In addition, due to its openness to qualitative information case study evaluation can take into account also aspects that cannot be expressed in quantitative or monetary terms (Ruegg, 2006).

On the other hand, the lack of objectivity, validity and generalizability are criticized (Flyvbjerg, 2004). One main objection regarding case studies is that analysing only the participants of the policy programme does not allow for conclusions on causality. This argument is based on the counterfactual definition of causality which as evidence that A being a cause of B requires two conditions: A and B coincide, and B would not have happened without the occurrence of A (Mohr, 1999). To examine the counterfactual case, i.e. to check whether B would have occurred if A had not happened, a control group is necessary. Case studies that compare participants and non-participants (e.g. clusters with and without political support or institutions within and outside the cluster) rely on this concept of causality. But if only participants are analysed, an alternative definition is required. Corresponding approaches have been developed, like the *modus operandi* (Scriven, 1976), process tracing (George and Bennett, 2005) and pattern matching (Campbell, 1966) which take as evidence for causality the existence of a 'signature' (Mohr, 1999) which links the cause with the result. In this manner, the case study must rule out concurring plausible causes and step by step show the 'characteristic value chain' (Scriven, 1976: 105), i.e. the mechanism how the cause A leads to the effect B. In social sciences the causes for an event are often intangible and thus difficult to detect, but nevertheless this method can produce (internally) valid results (Mohr, 1999).<sup>4</sup> However, the more indirect a policy instrument is, the more difficult is it to track the mechanisms.

A second objection is related to the degree to which these results can be generalized: the smaller the sample, the higher is the risk that the observed subjects are (positive or negative) particular cases (Mohr, 1999).<sup>5</sup> This implies the necessity to document the circumstances of the

cases to gain explanatory power; but then case studies, which target the understanding of the causal process, may even have an advantage over large sample econometric studies, as Mohr (1999) argues. In the context of cluster policy evaluation, where cases can be focused for instance on single players and their linkages as well as on specific projects of the programme, generalization issues are demanding. Not only the circumstances of the cluster development must be taken into account, but also particularities of the analysed players and their position and function in the cluster. In this regard case studies can be used for efficacy rather than for effectiveness analyses. They may show that under certain conditions a specific policy works, without proving its general effectiveness.

Hence, for case study evaluation the first step appears to be crucial: the choice of the case(s). They should give insights into the functioning of the programme and enable learning instead of only illustrating success stories in the manner of ‘cherry-picking’. (In particular the analysis of unsuccessful projects, for instance, can be instructive, according to Shipp et al., 2005.) Then, it seems to be a suitable approach capable of taking into account both the idiosyncrasies of clusters and the complexity of cluster policy. In fact, a considerable number of case studies have been conducted for evaluation purposes (e.g. Fromhold-Eisebith and Eisebith, 2008).

### *Econometric Models*

In contrast to case study analysis, (micro-)econometric impact analysis is based on the mentioned counterfactual definition of causality (White et al., 2006). It has been gaining attention over the last years via a set of elaborate and informative techniques (see Augurzky and Kluge, 2004; Heckman, 2004), its strength lies in the distinction between significant policy impacts and concomitant circumstances independent of the policy measures (White et al., 2006). But, as the analysis is focused on the clustered firms (instead of the entire cluster), rather than the policy impact on the cluster, the indirect effects on the members of the cluster are captured. Therefore, the indirect (i.e. through cluster development) effects of cluster policy on firm performance must be taken into account explicitly in the underlying theoretical model. Accordingly, indicators must be found that adequately represent the effects of the policy measure on firm performance. As companies benefit from innovation clusters in various ways – be it knowledge flows, a specialized labour pool or specialized suppliers – intermediate indicators such as innovation performance can be taken as well as success measures like profits, firm growth or entrepreneurship (for an overview of indicators see Arthurs et al., 2009). The choice of indicators is not trivial, since all of them are criticized for having their deficiencies. Input indicators such as R&D spending are criticized as being uninformative regarding firm performance. If using output indicators like patents, profits or firm growth, on the other hand, due to the large number of potential influencing factors and the time lag between the intervention and measurable impact, a clear attribution of policy impacts is difficult. In addition, quantitative indicators capture reality only incompletely; measuring, for example, linkages within the cluster by the number of cooperative agreements does not consider the intensity, quality or the success of the cooperations (Ahuja, 2000).

Quantitative (microeconomic) evaluation explicitly aims at answering the question: ‘what would have happened without the intervention?’ This question has been formalized in the ‘Potential Outcome Model’ (POM, see Holland, 1986). The policy impact can be written as the difference between the case of policy intervention  $Y_i(1)$  and the case without intervention  $Y_i(0)$ :

$$\Delta Y_i = Y_i(1) - Y_i(0)$$

Where  $Y$  can be any indicator as already described, and  $i$  is the observed unit respective to the aggregation level, for example, a firm or household. The analysis can be based on a sample of  $i = 1, 2, \dots, N$  units as well as on one element in  $i = 1, 2, \dots, T$  different points in time. The value of  $\Delta Y_i$ , as formulated, is not observable, since for each  $i$  only one of the cases has actually materialized, while the second case is an unobserved counterfactual situation.

Instead of individual level effects, evaluation typically focuses on the average treatment effect on the treated (see Keilbach, 2005), assuming that the intervention does not have any impact on non-participating elements (White et al., 2006). These average treatment effects result from the (hypothetical) comparison of the ‘treated’ elements, which have undergone the intervention (i.e.  $M = 1$ ), with the counterfactual case that these elements would not have participated in the programme:

$$\tau = E\{\Delta Y | M = 1\} = E\{Y(1) | M = 1\} - E\{Y(0) | M = 1\}$$

Still, the counterfactual  $E\{Y(0) | M = 1\}$  cannot be observed, so the counterfactual average must be substituted by an alternative, observable average. This is done by constructing a control group (outside the policy programme); according to the type of control group, a number of approaches can be distinguished.

The *before–after comparison* focuses on only the treated elements, but at two different moments so that the performance of the treated units previous to the intervention ( $t = 0$ ) can be used for comparison with the performance during or after the programme ( $t = 1$ ).

$$\tau = E\{Y_{t=1}(1) | X, M = 1\} - E\{Y_{t=0}(0) | X, M = 1\}$$

The main underlying assumption here is that (while controlling for observable factors  $X$ ) unobserved factors remain constant over time, both regarding environmental conditions and unobserved properties of the elements like management talent of firms (Grossman, 1994). But, strategic behaviour of programme participants, for example, postponing eligible investments to the programme period, cannot be detected, which leads to overestimation of the policy impact (see Ashenfelter, 1978). Considering cluster policy evaluation as field of application, an additional bias of before–after comparisons might arise from the fact that a linear development of clusters independent of business cycle influences and path-dependencies cannot be assumed. Thus, the before–after estimator which compares each clustered firm to itself attributes also the effects of changed unobservable characteristics to the cluster.

The *with/without comparison* contrasts participants with non-participants:

$$\tau = E\{Y(1) | X, M = 1\} - E\{Y(0) | X, M = 0\}$$

This implies that the participants would have performed similarly to the non-participants if they had not been subject to the intervention, conditional to additional observable characteristics ( $X$ ). But if programme participation is not random – which can be assumed for cluster policy, which is explicitly targeted at high-potential clusters – the with/without comparison is vulnerable to self-selection bias, since factors leading to programme participation might also influence the performance of the participants irrespective of the intervention. In addition, the comparison might be biased by exogenous factors which influence the control group in a different way than the participant group (see Blundell and Costa Dias, 2000).

The *difference-in-difference estimator* can be seen as a combination of the two mentioned approaches as it compares the development of the performance of both participants and non-participants.

$$\tau = E\{Y_{t=1}(1) - Y_{t=0}(1)|X, M = 1\} - E\{Y_{t=1}(0) - Y_{t=0}(0)|X, M = 0\}$$

However, bias due to self-selection and strategic behaviour of the participants persist also in this method, even if controlling for observable characteristics (see Heckman and Smith, 1999). Besides, panel data on both participants and non-participants is required.

Thus, also if these control group methods appeal by their intuitive understanding and more or less simple implementation, they are likely to be biased by neglecting unobservable factors. Nevertheless both approaches can produce very informative and reliable results, provided that an adequate control group is chosen to simulate a natural experiment. Opportunities for a simple control group comparison arise often in case of a structural break, for example, due to changes in policy conditions or regional differences. In these cases, neither selection bias nor strategic behaviour interferes with the identification assumption.

Alternatively, *selection models* can be applied which explicitly consider the participation decision and thus eliminate potential selection bias. The first type, *instrumental variables estimation* is based on the assumption that some variables ( $Z$ ) explain the participation status, but are uncorrelated with unobservable characteristics affecting the outcome  $Y$ . Then, the treatment variable  $M$  can be regressed on  $Z$ , so that the coefficient of the resulting instrument  $\hat{M}$  can be interpreted as the treatment effect. The choice of instruments, however, is difficult, since both weak instruments and the correlation with omitted variables will lead to inconsistent estimation results (Angrist and Krueger, 2001). As a second type, the *Heckman Selection Correction* (Heckman, 1979) takes into account both observable and unobservable factors when modelling the participation selection. In a two-stage approach, a (binary) participation estimation is specified which is subsequently used for the outcome regression (in form of the inverse Mills ratio). By this, the estimated effect of the expected participation status conditional on observable and unobservable characteristics is interpreted as the treatment effect. But, as has been shown by LaLonde (1986), the results may be biased due to specification errors, regarding both the choice of control variables and the distribution of the error term.

In contrast, *matching* as a quasi-simulation approach tries to overcome the shortcomings of observational data by building 'twin' pairs of treated and non-treated elements, i.e. assigning to each participant a non-participant with similar observable characteristics. After the matching procedure, the subsamples of treated and not-treated elements should be comparable so that the average treatment effect on the treated can be estimated:

$$\tau = E\{Y(1)|X, M = 1\} - E\{Y(0)|X, M = 0\}$$

As finding a perfectly identical twin is difficult (and impossible in case of continuous control variables as argued by González and Pazó, 2008) and requires large data sets, a certain degree of dissimilarity between the matched elements is allowed; for instance, in nearest neighbour or radius matching from the variables  $X$  a propensity score (Rosenbaum and Rubin, 1983) is estimated which is used as matching criterion afterwards (see Becker and Ichino, 2002).

The matching method relies on two assumptions:

- The unconfoundedness assumption  $\{Y(0), Y(1)\} \perp M | X$  means that treatment and potential outcomes are independent, conditional on the observed variables  $X$  (Rosenbaum and Rubin, 1983).
- Overlap:  $0 < p(M = 1 | X) < 1$ . This condition ensures that statistical twins can be found, since the propensity score of both treated and untreated firms lies within the same interval  $[0; 1]$  (Imbens, 2004).

Unlike two-stage estimation methods, matching permits an intuitive understanding and direct interpretation of the results (Dar and Gill, 1998). On the other hand, matching is sensitive to the choice of control variables as well as unobservable characteristics. In particular, the plausibility of the unconfoundedness assumption is doubted (Imbens, 2004). Furthermore, all methods alike face the difficulty of identifying not-treated firms, which should be part of a comparable cluster, but not participate in any (cluster) policy programme; otherwise the analysis would overstate the treatment effect by confounding agglomeration advantages and policy impact if the firms in the treated cluster were compared to not-treated and not-clustered firms.

Apart from the econometric pitfalls of all the mentioned methods, the explanatory power of econometric methods depends on the data available (Schmidt, 1999). The applications are based on micro-level data focusing on firms, households or persons for which data on a large number of comparable observations are available.<sup>6</sup> A systemic view on the entire cluster, however, seems hardly compatible with these econometric models; and by only comparing treated and not-treated firms the whole process of cluster policy and development is seen as a black box. Thus, these approaches are likely to miss the core aspect of cluster development of a collective action (Raines, 2002b). Furthermore, also if applied on the firm-level, the quantitative precision of the results produced by econometric methods may in some cases be misplaced (Corley, 2007) and create an overstated impression of objectivity. After all, the results depend on the specification so that even small changes regarding method or variables alter the measured treatment effects. This has to be borne in mind as well if comparing evaluation studies on different programmes.

### Systemic Approaches

As an alternative to the microeconomic-level analysis, also a number of quantitative methods taking a systemic point of view are available. These tools provide a static descriptive view on the cluster, so that for evaluation the analysis must include two or more points in time to investigate how the cluster has developed regarding its size and structure. This can be seen as a before/after comparison, but without being able to attribute the changes to the policy programme via significance tests.

*Input–output models* study the relations between industries and spatial units on the basis of commodity flows (see Schaffer, 1999). Based on regional input–output tables, linkages between industries and the importance of industries on the economic development of the region can be approximated. This method is commonly used to identify clusters (e.g. Larreina, 2007), but also for cluster policy evaluation (Learmonth et al., 2003).

Input–output models can be implemented at any regional level, so that the aggregation can to a certain degree be adapted to the dimensions of the cluster. But full input–output tables on a small regional scale are rarely available and difficult to construct, in particular if the analysis focuses on single sectors on a disaggregated level (see Gabriel (2001) for Hamburg and

Larreina (2007) for the Rioja region). To determine the regional level of the model, assumptions on the geographical boundaries of the cluster are necessary; if the cluster is spread over administrative borders a realistic delimitation is difficult. Similarly, the delineation of the cluster along industry lines is problematic as typically the clustered firms belong to more than one industry, so that the results given by input–output analysis are somewhat arbitrary (Lublinski, 2001). Furthermore, input–output models only consider commodity flows whereas immaterial linkages such as flows of (both codified and tacit) knowledge are not captured (Lublinski, 2001). Hence, input–output analysis will produce – at best – an incomplete picture of the linkages within the cluster so that in the majority of cases additional investigations will be necessary, be it network analysis (as proposed e.g. by van den Hove et al., 1998) or micro-economic methods.

*Network analysis* explicitly takes a systemic perspective as well, but instead of value-added chains it relies on communication and interaction linkages between the players within the cluster (Dybe and Kujath, 2001). In network analysis, the cluster as a social system is represented by a network of vertices and edges which represent the actors within the cluster and the ties between the actors, respectively. It is based on an interaction matrix (possible matrix types are similarity or distance matrices, see Hawe et al., 2004) containing data on the relationships between the members of the network.<sup>7</sup> The required data can be drawn from surveys asking the actors about their relations with other actors (for a description of survey methods and difficulties see Marsden, 2005); also communication flows measured by email traffic (e.g. Gloor et al., 2008), co-authorship in the academia (e.g. Newman, 2004) or collaborations between firms (e.g. Breschi and Cusmano, 2004) can be used to identify the existence and strength of ties between the members of the network.

Depending on the research question and the scale of the network, three (distinct, but partly complementary) types of analyses can be used. Graphical visualizing is predominantly used as a tool for explorative analysis; it can give intuitive insights in network structure, though its analytical power is limited to small networks (Newman, 2003: 169–71). Quantitative analysis applies a number of (descriptive) statistical metrics on network properties, such as density, connectivity, clustering, resilience or community structure (see Wassermann and Faust, 2005). Similarly, the position of single actors within the network can be studied, considering for example their centrality, hierarchical position, and structural holes.

Cluster development can be defined as network growth, i.e. the addition of vertices and/or edges, as well as changes in the properties of the cluster as measured by indicators like resilience and transitivity or regarding the simulated performance of the network. A number of models also take into account dynamic network evolution explicitly (for an overview see Snijders, 2005). By setting network development in a relationship to the cluster policy measures, the analysis can provide insights – though without statistical significance – on how the policy influenced the cluster.

As cluster policy explicitly focuses on the interaction of players, network analysis seems to be an adequate tool, at least for exploratory analysis on cluster development. However, the value of a cluster is determined not only by network size and strength, but by the economic value the firms draw from these relationships (Raines, 2002b). Thus, by relying only on network analysis, the evaluation remains limited to intermediate outputs, but cannot draw conclusions on the real economic benefits of the policy.

The third systemic approach, *Benchmarking*, is known as a management tool rather than an policy evaluation method. It can be used in particular if cluster policy is spread over a number of clusters, so that the success of the single projects can be evaluated in comparison with each

other. This can help to detect success factors and give insights into why certain measures did (not) work by placing successful interventions alongside failed concepts; similarly, the implementation of one specific measure in several clusters can be analysed by benchmarking cluster development. It should be kept in mind, though, that the results are sensitive to contextual factors which should be taken into account (Gebel, 2006). As multiple indicators, both of a qualitative and quantitative nature, can be included, a detailed picture can be drawn (Schütz et al., 1998), depending on the quality of performance indicators (Gebel, 2006). In particular, indicators can be taken from preceding evaluation steps, combining for instance network indicators or policy input measures in one benchmarking study. An overview of indicators for cluster analyses is given for example by Koschatzky and Lo (2007).

In short, the benchmarking process contains the following steps (Gebel, 2006; Schütz et al., 1998; Tornatzky, 2003): (1) finding a benchmarking group, (2) defining dimensions and indicators of performance, (3) identifying 'best-in-class' organizations or programmes, (4) determining the performance gap, and (5) describing best practices.

To measure the performance gap, a number of methods have been proposed (Jones, 2004). If a combination of indicators is used, they can be integrated in a Radar-chart analysis, which reduces complexity and thus permits an intuitive analysis (Schütz et al., 1998). Starting from a graphical representation of the indicators (normalized to the interval [0; 1]) in a radial chart, the SMOP8 (Surface Measure of Overall Performance) can be calculated for each cluster:

$$\text{SMOP} = \{(x_1 \ x_2) + (x_2 \ x_3) + (x_3 \ x_4) + \dots + (x_n \ x_1)\} \sin \frac{360^\circ}{n},$$

Where  $x_1, \dots, x_n$  are the distances of the respective corner to the center of the chart, corresponding to the value of the respective indicator. The higher the SMOP-value, the better is the performance of the respective project, so that projects can be ranked according to their performance. Note that the SMOP value is sensitive to the order of the indicators; therefore the values can be interpreted only in relation to each other, but not regarding their absolute values. The graphical analysis, in addition, can be used to identify strengths and weaknesses of measures and trade-offs between policy targets (Schütz et al., 1998). However, a further analysis of best practices is required to find the reasons for the different performance.

### *Cost-related Approaches*

So far, the article has presented evaluation methods that take into account only the impact of the policy measure. But to achieve a full assessment of the success of an intervention, the cost should be considered (Dar and Gill, 1998), in order to assess if the effect of the programme is worth the resources spent. Instead of effectiveness of the policy measure, cost-oriented methods investigate efficiency (Schmidt, 1999). These approaches are an extension of the presented methods; they make use of the estimation of impacts and add the calculation of costs. This implies, however, that the drawbacks of the methods applied to capture programme impacts persist, so the quality of the results depends centrally on the underlying 'pure' impact analysis (e.g. econometric methods or network analysis).

A number of methods are available that differ in the way they capture policy impacts: *Cost-benefit analysis* expresses the estimated impact in monetary terms in order to measure the net benefit or the

rate of return of an intervention (Levin and McEwan, 2000). Values are discounted over time in order to set lagged impacts in relation to immediate expenditures. This point is of particular importance if long-term effects of programmes are studied (as is the case in cluster policy). But the social discount rate, being a key parameter in the cost–benefit analysis, is difficult to define (Spackman, 2007). Moreover, in many cases the quantification of benefits in monetary values is questionable, particularly if ‘soft’ effects like communication flows in a firm network are measured (Levin and McEwan, 2000). For a full financial appraisal of benefits, both direct and indirect beneficiaries as well as unintended side effects (Stufflebeam, 1999) and also negative effects of the programme such as potential displacement effects, i.e. the shifting of economic activity from other regions, must be considered (Dar and Gill, 1998).

In contrast, *cost-effectiveness analysis* directly sets the effects in relation to costs by building cost-per-unit ratios (e.g. based on regression coefficients of the preceding impact analysis). By this, assumptions on valuation and discount rates can be avoided, and any outcome indicator can be used. But the intuitive and simple character of the analysis vanishes as soon as more than one policy outcome is analysed. This compound of effects can be expected for most cluster policies which typically combine a range of measures (Raines, 2002b). Furthermore, these calculated values do not allow for conclusions on the efficiency of a programme, as long as they are not contrasted with another (comparable!) intervention (Levin and McEwan, 2000).

The calculation of costs is identical for both methods, bearing difficulties similar to the estimation of financial benefits. First, the evaluator must be sure to have included all relevant costs – direct implementation costs of the programme as well as social costs and opportunity costs (Schmidt, 1999). Again the proper discount rate must be defined, giving rise to over- or underestimation. Besides, as many of the required values (e.g. opportunity costs) can only be approximated, the calculation relies on many, sometimes questionable (Stufflebeam, 1999: 21) assumptions.

Applications of all named approaches can be found in various policy areas (illustrative examples on education are listed by Levin and McEwan, 2000).<sup>9</sup> However, the use of cost-related methods is limited by their complexity and high data requirements, in particular in policy areas with intricate structures. Accordingly, cluster policy is rarely evaluated using cost-oriented approaches, mainly due to the complexity of the field; as determining impacts is quite difficult, most evaluations are limited to this task, avoiding additional time and effort for financial quantification of benefits and costs.

## Summary: Choosing Evaluation Methods

In the preceding discussion, a range of methods has been presented which can be applied to undertake cluster policy evaluation. It has been shown that these approaches are different in their focus and the underlying assumptions, building on different theoretical models and taking different perspectives. Which method is appropriate depends on the purpose of the evaluation and the structure and scope of the programme, but also on limitations regarding time, financial resources and methodological capacities. In Table 1 an overview of relevant characteristics of the techniques is provided, according to the criteria guiding the decision on the actual evaluation design in a specific case. The first group of criteria regards the basic conditions of the evaluation and gives clear guidance: If short-term evaluation is requested, the analysis is limited to policy input reporting, while for learning purposes qualitative studies including case

studies are advisable. Besides, a good match between the policy strategy and the evaluation method is important in order to produce informative evaluation results. For instance, if the programme primarily pursues a branding strategy, an I/O-analysis on commodity flows will be off target. In the second group of criteria data requirements are summarized. Data availability has been identified as one of the central problems in evaluation of cluster policy, so that a thorough appraisal of existing data, information sources and options of collecting data must precede the evaluation design. Data limitations, which preclude the application of corresponding methods, may arise if resources required for data collection are lacking, data (e.g. financial data for cost-related approaches) are not made available by the programme coordinators and participants, or the sample size (e.g. the number of firms in the cluster) is too small. The third group of criteria shows what can be expected of the methods. Certainly, the more complex methods have the higher explanatory power, but on the other hand if the results are difficult to understand or do not allow for practical conclusions, the impact of the evaluation on policy-makers will be low.

A short summary of the presented evaluation methods follows:

- *Reporting methods*, which are the least challenging instrument regarding the timeframe, data requirements and complexity, should be included in every evaluation. Besides making general information on the setting available to the evaluators, reporting can be used as a controlling tool.
- The strength of *case studies* lies in their intuitive understanding, flexibility and in-depth view. They can show the mechanisms of cluster development in detail, but generalizing results of case studies is difficult. Answers on the question ‘did the programme work?’ may be ambiguous.
- *Econometric methods* can quantitatively test the effects of cluster policy (mainly on single actors within the cluster), which increases the credibility of the results. Requirements regarding data and methodological capabilities are high, and often significant and positive results will be found only several years after the policy programme. The evaluation does not take into account soft facts and details.
- *Systemic approaches* in particular take the cluster idea into account, instead of focusing on single members of the cluster. I/O-analysis and network analysis provide quantitative results on cluster performance, but data requirements are high. In contrast, benchmarking can be used to analyse best practices and critical aspects in cluster policy.
- *Cost-related approaches*, finally, should be included in each evaluation to give an answer on the efficiency question: ‘was it worth it?’ But in particular due to data restrictions, reliable cost-related evaluations are difficult. They can be based on the above-named quantitative methods, and should ideally be combined with one of the qualitative methods to enable learning.

It can be seen from this list that using only a single evaluation method will provide a very limited view on the cluster policy programme. This makes it even more important to choose an evaluation method or a combination of methods which is appropriate for the specific evaluation question. On the other hand, as time and effort to undertake a full quantitative and qualitative evaluation might be beyond the budget of the cluster policy programme, resource requirements, explanatory power and practical value of the evaluation method should be balanced and adapted to the target group.

**Table 1.** Overview of Evaluation Methods

Approach Criterion	Reporting	Case studies	Econometric methods	Systemic approaches		Cost-related approaches
				I/O-analysis	Network analysis	
<b>1. Evaluation conditions</b>						
Time horizon	short	Medium /long	Medium/long	Medium /long	Medium /long	Medium /long
Type of cluster policy	All types	Combination of multiple instruments	clearly defined policy targets (with focus on firm performance)	focus on buyer-supplier-relationships	network oriented policy strategy targeting several clusters	depending on type of impact analysis
Purpose of evaluation	preparing further evaluation steps; transparency	legitimating and efficacy control	legitimating and effectiveness control	Legitimating and effectiveness control	learning	decision support; legitimating and efficiency control
Target group	tax payers; policy makers	Policy makers	Economic research; policy makers	Economic research; policy makers	Policy makers	economic research; tax payers; policy makers
<b>2. Data requirements</b>						
Collect own data	not necessarily: use of official project reports and statistics	yes: detailed and specific information on the cases required	not necessarily: use of existing firm level databases	yes: I/O-tables on cluster-level usually not available	yes: complete data on all network participants required	not necessarily: depending on type of impact analysis
Indicators	a) information from project reports, e.g. activities, financial resources b) opinions of participants	Multiple indicators (quantitative and qualitative)	(quantitative) firm level indicators: productivity, firm growth, etc.	regional commodity flows (from regional I/O-tables)	firm level data, e.g. patents, communication flows, collaboration links	unit-costs, net benefit; complete cost accounting, estimation of opportunity costs

Table 1. (Continued)

Approach Criterion	Reporting	Case studies	Econometric methods	Systemic approaches		Cost-related approaches
				I/O-analysis	Network analysis	
Sample size	a) no sample b) only participants (interviews, surveys)	Small sample	large sample: comparison group outside the cluster and/ or observation over time	Aggregate (regionalized) data	large sample required: complete representation of the network	depending on type of impact analysis
<b>3. Evaluation results</b>						
Perspective	Policy input	single aspects, of e.g. firms' benefits, programme execution	Impact on firm performance	impact on cluster performance; impact on territorial performance	Impact on cluster performance impact on territorial performance	depending on type of impact analysis
Explanatory power	low: no impact measurement	exploratory, conclusions on best practices	statistical significance of impacts; cluster treated as a black box	only information on commodity flows, no significance test if changes are caused by policy	descriptive results on network behavior, no significance test on policy impacts, no conclusions on economic impact	information on both impacts and costs; statistical significance only in combination with econometric methods
Interpretation of results	shows shortcomings in programme execution	illustrative, concrete examples	abstract, needs explanation	abstract, needs explanation	Graphical illustration possible	clear interpretation of indicators

## Notes

1. Throughout this article, the terminus evaluation is used according to the definition of the UK Cabinet Office (Spencer et al., 2003) which distinguishes (ex-post) evaluation from (ex-ante) appraisal (see also Rip, 2003).
2. The function of evaluation as an integral part of the cluster policy cycle is shown in Figure 1.
3. The differentiation between effectiveness, efficacy and efficiency stems from medicine, where therapeutic effects are assessed according to these definitions (see Cochrane, 1972).
4. In contrast to its controversial status in the field of evaluation, this kind of causal inference is used widely in other disciplines such as medical diagnosis, history or engineering (Scriven, 1976).
5. From an epistemological viewpoint, the difference between case studies and large sample observational studies is only secondary: From single observations – regardless of their number – one cannot derive universal conclusions, so both types of empirical evidence are subject to the ‘problem of induction’ (Popper, 2002: 3).
6. In fact, the bulk of contributions can be found in the literature on firm R&D subsidies (e.g. Autio et al., 2008; Gabriele et al., 2006) and labour market programmes (e.g. Heckman et al., 1997; Ichino et al., 2008).
7. Social network analysis studies typically use complete network data instead of random samples (see Breiger, 2004), which increases the efforts to be made for data collection as compared to empirical analyses on individuals. Apart from that, for data collection the same methods are applied as for individual level studies (see Hawe et al., 2004).
8. An application of SMOP not on cluster policy but for benchmarking clusters can be found in Pfähler and Lublinski (2003).
9. In particular cost–benefit analysis is used not only for (ex-post) evaluation, but also for ex-ante appraisal (see e.g. European Commission, 2006).

## References

- Ahuja G (2000) Collaboration networks, structural holes, and innovation: a longitudinal study. *Administrative Science Quarterly* 45(3): 425–55.
- Angeles Diez M, Esteban MS (2000) The evaluation of regional innovation and cluster policies: looking for new approaches. Paper presented at the 4th EES Conference, Lausanne.
- Angrist JD, Krueger AB (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* 15(4): 69–85.
- Aranguren MJ, Larrea M, and Navarro I (2006) The policy process: clusters versus spatial networks in the Basque Country. In Pitelis C, Sugden R, and Wilson JR (eds) *Clusters and Globalization*. Cheltenham and Northampton: Edward Elgar, 258–80.
- Arthurs, D, Cassidy E, Davis CH, and Wolfe D (2009) Indicators to support innovation cluster policy. *International Journal of Technology Management* 45:3–4: 263–279.
- Ashenfelter O (1978) Estimating the effect of training programs on earnings. *Review of Economics and Statistics* 60(1): 47–57.
- Augurzyk B, Kluge J (2004) Assessing the performance of matching algorithms when selection into treatment is strong. RWI Discussion Paper No. 21, Essen: Rheinisch-westfälisches Institut für Wirtschaftsforschung.
- Autio E, Kanninen S, and Gustafsson R (2008) First- and second-order additionality and learning outcomes in collaborative R&D programs. *Research Policy* 37(1): 59–76.
- Bachtler J, Yuill D, and Davies S (2005) *Regional Policy and Innovation*. EoRPA Paper 05/5, prepared for the twenty-sixth meeting of the EoRPA Consortium, Ross Priory.

- Becker SO, Ichino A (2002) Estimation of average treatment effects based on propensity scores. *Stata Journal* 2(4): 358–77.
- Benneworth P, Charles D (2001) Bridging cluster theory and practice: learning from the cluster policy cycle, in OECD (ed.) *Innovative Clusters: Drivers of National Innovation Systems* 389–403. Paris: OECD.
- Benneworth P, Danson M, Raines P, and Whittam G (2003) Confusing cluster? Making sense of the cluster approach in theory and practice. *European Planning Studies* 11(5): 511–20.
- Blundell R, and Monica CD (2000). Evaluation Methods for Non-Experimental Data. *Fiscal Studies*, 21(4): 427–468.
- Boekholt P (2003) Evaluation of regional innovation policies in Europe. In Shapira P, Kuhlmann S (eds) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*. Cheltenham and Northampton: Edward Elgar, 244–59.
- Boekholt P, Thuriaux B (1999) Public policies to facilitate clusters: background, rationale and policy practices in international perspective. In OECD (ed.) *Boosting Innovation: The Cluster Approach*, 381–412. Paris: OECD.
- Breiger RL (2004) The analysis of social networks. In Hardy MA, Bryman A (eds) *Handbook of Data Analysis*. London: SAGE, 505–26.
- Breschi S, Cusmano L (2004) Unveiling the Texture of a European Research Area: Emergence of Oligarchic Networks under EU Framework Programmes. *International Journal of Technology Management* 27(8): 747–772.
- Buchinger E, Wagner P (2003) Evaluierung des Technologietransfer-Programms ‘TechnoKontakte Seminare’, Endbericht im Rahmen des Projektes 7.62.00083 im Auftrag des Österreichischen Bundesministeriums für Wirtschaft und Arbeit, Seibersdorf.
- Buendia F (2005) Towards a system dynamic-based theory of industrial clusters. In Karlsson C, Johansson B, and Stough RR (eds) *Industrial Clusters and Inter-Firm Networks*. Cheltenham and Northampton: Edward Elgar, 83–106.
- Camagni RP (1995) The concept of innovative milieu and its relevance for public policies in European lagging regions. *Papers in Regional Science* 74(4): 317–40.
- Campbell DT (1966) Pattern matching as an essential in distal knowing. In Hammond KR (ed.) *The Psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston, 81–106.
- Cheshire PC (2003) Territorial competition: lessons for (innovation) policy. In Bröcker J, Dohse D, and Soltwedel R (eds) *Innovation Clusters and Interregional Competition*. Berlin: Springer, 331–46.
- Cochrane AL (1972) *Effectiveness and Efficiency: Random Reflection on Health Services*. London: Nuffield Provincial Hospitals Trust.
- Corbett T, Lennon MC (2003) Implementation studies: from policy to action. In Lennon MC, Corbett T (eds) *Policy into Action: Implementation and Welfare Reform*. Washington, DC: Urban Institute Press, 1–14.
- Corley EA (2007) A use-and-transformation model for evaluating public R&D: illustrations from polycystic ovarian syndrome (PCOS) research. *Evaluation and Program Planning* 30(1): 21–35.
- Dar A, Gill IS (1998) Evaluating retraining programs in OECD countries: lessons learned. *World Bank Research Observer* 13(1): 79–101.
- DTI (2004) *A Practical Guide to Cluster Development: A Report to the Department of Trade and Industry and the English RDAs*. London: DTI.
- European Commission (2002) *Regional Clusters in Europe*. Observatory of European SMEs 2002/3. Brussels: EC.
- European Commission (2004) *Five-Year Assessment of the European Union Research Framework Programmes 1999–2003*. Luxembourg: EC.
- European Commission (2006) *Guidance on the Methodology for Carrying out Cost–Benefit Analysis*. DG Regional Policy Working Document, 4. Brussels: EC.

- Florida R (2002) *The Rise of the Creative Class*. New York: Basic Books.
- Flyvbjerg B (2004) Five misunderstandings about case-study research. In Seale C, Gobo G, Gubrium JF and Silverman D (eds) *Qualitative Research Practice*. London: SAGE, 420–34.
- Formica P (2003) Corporate governance of cluster development agencies: the case for market orientation. In Bröcker J, Dohse D, and Soltwedel R (eds) *Innovation Clusters and Interregional Competition*. Berlin: Springer, 241–71.
- Foss-Hansen H (2005) Choosing evaluation models: a discussion on evaluation design. *Evaluation* 11(4): 447–62.
- Fromhold-Eisebith M, Eisebith G (2008) Looking behind facades: evaluating effects of (automotive) cluster promotion. *Regional Studies* 42(14): 1343–56.
- Gabriel C (2001) Constructing regionalized input–output tables: a new simple-to-use method. In Pfähler W (ed.) *Regional Input–Output Analysis: Conceptual Issues, Airport Case Studies and Extensions*, 75–87. Baden-Baden: Nomos.
- Gabriele R, Zamarian M, and Zaninotto E (2006) Assessing the economic impact of public industrial policies: an empirical investigation on subsidies. Paper presented at EARIE annual conference, Amsterdam.
- Gebel M (2006) *Monitoring und Benchmarking bei arbeitsmarktpolitischen Maßnahmen*. ZEW-Dokumentation 2006-01, Mannheim: ZEW.
- George AL, Bennett A (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Georghiou L, Roessner D (2000) Evaluating technology programs: tools and methods. *Research Policy* 29: 657–78.
- Ghuri PN (2004) Designing and conducting case studies in international business research. In Marschan-Piekkari R, Welch C (eds) *Handbook of Qualitative Research Methods for International Business*. Cheltenham and Northampton: Edward Elgar, 109–24.
- Gloor PA, Kidane YH, Grippa F, Marmier P, and von Arb C (2008) Location matters? Measuring the efficiency of business social networking. *International Journal of Foresight and Innovation Policy* 4(3–4): 230–45.
- González X, Pazó C (2008) Do public subsidies stimulate private R&D spending? *Research Policy* 37(3): 371–89.
- Gordon IR, McCann P (2000) Industrial clusters: complexes, agglomeration and/or social networks. *Urban Studies* 37(3): 513–31.
- Grossman JB (1994) Evaluating social policies: principles and U.S. experience. *World Bank Research Observer* 9(2): 159–80.
- GTZ (2007) *Cluster Management: A Practical Guide*. Eschborn: Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH.
- Guinet J (2003) Drivers of economic growth: the role of innovative clusters. In Bröcker J, Dohse D, and Soltwedel R (eds) *Innovation Clusters and Interregional Competition*. Berlin: Springer, 150–60.
- Guy K (2003) Assessing RTD program portfolios in the European Union. In Shapira P, Kuhlmann S (eds) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*. Cheltenham and Northampton: Edward Elgar, 174–203.
- Hansen HF (2005) Choosing evaluation models: a discussion on evaluation design. *Evaluation* 11(4): 447–62.
- Hartmann C (2002) Styria. In Raines P (ed.) *Cluster Development and Policy*. Aldershot: Ashgate, 123–40.
- Hawe P, Webster C, and Shiell A (2004) A glossary of terms for navigating the field of social network analysis. *Journal of Epidemiology and Community Health* 58(12): 971–5.
- Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47(1): 153–61.
- Heckman JJ (2004) Micro data, heterogeneity and the evaluation of public policy. *American Economist* 48(2): 3–25.

- Heckman JJ, Smith JA (1999) The pre-programme earnings dip and the determinants of participation in a social programme: implications for simple programme evaluation strategies. *Economic Journal* 109(457): 313–48.
- Heckman JJ, Ichimura H, and Todd PE (1997) Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* 64(221): 605–54.
- Holland PW (1986) Statistics and causal inference. *Journal of the American Statistical Association* 81(396): 945–60.
- Ichino A, Mealli F, and Nannicini T (2008) From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics* 23(3): 305–27.
- Imbens GW (2004) Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics* 86(1): 4–29.
- Jappe-Heinze A, Baier E, and Kroll H (2008) *Clusterpolitik: Kriterien für die Evaluation von regionalen Clusterinitiativen*. Fraunhofer ISI Arbeitspapiere Unternehmen und Region 3/2008, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI.
- Jick TD (1979) Mixing qualitative and quantitative methods: triangulation in action. *Administrative Science Quarterly* 24(4): 602–11.
- Jones MK (2004) The dynamic benchmarking of labour markets. *Regional Studies* 38(5): 495–506.
- Keilbach M (2005) *Quantitative, Non-Experimental Approaches to the Microeconomic Evaluation of Public Policy Measures: A Survey*. Discussion Papers on Entrepreneurship, Growth and Public Policy 2005-30, Jena: Max Planck Institute of Economics, Entrepreneurship, Growth and Public Policy Group.
- Koschatzky K, Lo V (2007). *Methodological Framework for Cluster Analyses*. Fraunhofer ISI Arbeitspapiere Unternehmen und Region R1/2007, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI.
- Krugman PR (1991) *Geography and Trade*. Leuven: Leuven University Press.
- Legendijk A (1999) Innovative forms of regional structural policy in Europe: the role of dominant concepts and knowledge flows. In Fischer MM, Suarez-Villa L, and Steiner M (eds) *Innovation, Networks and Localities*. Berlin: Springer, 272–99.
- Larreina M (2007) Detecting a cluster in a region without complete statistical data, using Input–Output analysis: The case of the Rioja wine cluster. *CRIEFF Discussion Papers* 706.
- LaLonde R (1986) Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76: 604–620.
- Learmonth D, Munro A, and Swales JK (2003) Multi-sectoral cluster modelling: the evaluation of Scottish Enterprise cluster policy. *European Planning Studies* 11(5): 567–84.
- Levin HM, McEwan PJ (2000) *Cost-Effectiveness Analysis* (2nd edn). Thousand Oaks, CA: SAGE.
- Litzenberger T, Sternberg R (2005) Regional clusters and entrepreneurial activities: empirical evidence from German regions. In Karlsson C, Johansson B, and Stough RR (eds) *Industrial Clusters and Inter-Firm Networks*. Cheltenham and Northampton: Edward Elgar, 260–302.
- Lublinski AE (2001) Identifying geographical business clusters: a critical review and classification of methods using I/O data. In Pfähler W (ed.) *Regional Input–Output Analysis. Conceptual Issues, Airport Case Studies and Extensions*, 37–49. Baden-Baden: Nomos.
- Lublinski AE (2002) *Geographical Business Clusters: Concepts for Cluster Identification with an Application to an Alleged Aeronautics cluster in Northern Germany*. Dissertation, University of Hamburg.
- Lublinski AE (2003) Does geographic proximity matter? Evidence from clustered and non-clustered aeronautic firms in Germany. *Regional Studies* 37(5): 453–67.
- Malmberg A, Maskell P (2002) The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A* 34(3): 429–49.

- Mansfield E, Rapoport J, Romeo A, Villani E, Wagner S, and Husic F (1977) *The Production and Application of New Industrial Technology*. New York: Norton.
- Marsden PV (2005) Recent developments in network measurement. In Carrington PJ, Scott J, and Wasserman S (eds) *Models and Methods in Social Network Analysis*. New York: Cambridge University Press, 8–30.
- Marshall A (1920) *Principles of Economics*. London: Macmillan.
- Martin R, Sunley P (2003) Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography* 3(1): 5–35.
- Mohr LB (1999) The qualitative method of impact analysis. *American Journal of Evaluation* 20: 69–84.
- Nagarajan N, Vanheukelen M (1997) *Evaluating EU Expenditure Programmes: A Guide. Ex Post and Intermediate Evaluation*. Brussels: European Commission, DG Budget.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Review* 45(2): 167–256.
- Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* 101 (suppl. 1): 5200–5.
- OECD (1999) *Boosting Innovation: The Cluster Approach*. Paris: OECD.
- OECD (2007) *Competitive Regional Clusters: National Policy Approaches*. Paris: OECD Reviews of Regional Innovation.
- Pelsoci TM (2005) *Photonics Technologies: Applications in Petroleum Refining, Building Controls, Emergency Medicine, and Industrial Materials Analysis*. NIST GCR 05-879, Gaithersburg, MD: National Institute of Standards and Technology.
- Pfähler W, Lublinski AE (2003) *Luftfahrt-Cluster Hamburg Norddeutschland: Bestandsaufnahme, Perspektiven und Vision für die Zulieferindustrie*. Frankfurt/Main: Lang.
- Popper KR (2002) *The Logic of Scientific Discovery*. London: Routledge.
- Porter ME (1990) *The Competitive Advantage of Nations*. The Free Press, New York.
- Raines P (2002) Cluster policy – does it exist? In Raines P (ed.) *Cluster Development and Policy*. Aldershot: Ashgate, 21–33.
- Raines P (2002b) The challenge of evaluating cluster behaviour in economic development policy. Paper presented at the International RSA Conference ‘Evaluation and EU regional policy: New questions and challenges’, Aix-en-Provence.
- Rip A (2003) Societal challenges for R&D evaluation. In Shapira P, Kuhlmann S (eds) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*. Cheltenham and Northampton: Edward Elgar, 32–53.
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.
- Rosenfeld SA (2002) Creating smart systems: A guide to cluster strategies in less favoured regions. Paper prepared for the EU Regional Innovation Strategies Conference.
- Rossi PH, Lipsey MW, and Freeman HE (2004) *Evaluation: A Systematic Approach* (7th edn). Thousand Oaks, CA: SAGE.
- Ruegg R (2006) *Bridging from Project Case Study to Portfolio Analysis in a Public R&D Program: A Framework for Evaluation and Introduction to a Composite Performance Rating System*. NIST GCR 2006-891, Gaithersburg, MD: National Institute of Standards and Technology.
- Ruegg R, Feller I (2003) *A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade*. NIST GCR 2003-857, Gaithersburg, MD: National Institute of Standards and Technology.
- Schaffer WA (1999) *Regional Impact Models*. Regional Research Institute, West Virginia University, Morgantown, WV.
- Schilder A (2000) *Government Failures and Institutions in Public Policy Evaluation: The Case of Dutch Technology Policy*. Assen: Van Gorcum.

- Schmidt CM (1999) *Knowing What Works. The Case for Rigorous Program Evaluation*. IZA Discussion Paper 77, Bonn.
- Schmitz H, Nadvi K (1999) Clustering and industrialization: introduction. *World Development* 27(9): 1503–14.
- Schütz H, Speckesser S, and Schmid G (1998) *Benchmarking Labour Market Performance and Labour Market Policies: Theoretical Foundations and Applications*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung Discussion Paper 198-205.
- Scriven M (1976) Maximizing the power of causal investigations: the modus operandi method. In Glass GV (ed.) *Evaluation Studies: Review Annual*. London: SAGE, 1, 101–18.
- Shipp S, Chang C, and Wisniewski L (2005) *Evaluation Best Practices and Results: The Advanced Technology Program*. NISTIR 2005-7174, Gaithersburg, MD: National Institute of Standards and Technology.
- Snijders TAB (2005) Models for longitudinal network data. In Carrington PJ, Scott J, and Wasserman S (eds) *Models and Methods in Social Network Analysis*. New York: Cambridge University Press, 215–47.
- Spackman M (2007) Social discount rates for the European Union: an overview. In Florio M (ed.) *Cost–Benefit Analysis and Incentives in Evaluation: The Structural Funds of the European Union*. Cheltenham and Northampton: Edward Elgar, 253–79.
- Spencer L, Ritchie J, Lewis J, and Dillon L (2003) *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London: Government Chief Social Researcher's Office.
- Sternberg R (2003) New firms, regional development and the cluster approach: what can technology policies achieve? In Bröcker J, Dohse D, and Soltwedel R (eds) *Innovation Clusters and Interregional Competition*. Berlin: Springer, 347–71.
- Stufflebeam DL (1999) *Foundational Models for 21st Century Program Evaluation*. Evaluation Center Occasional Paper 16, Western Michigan University, Kalamazoo.
- Tornatzky LG (2003) Benchmarking university–industry relationships: a usercentered evaluation approach. In Shapira P, Kuhlmann S (eds) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*. Cheltenham and Northampton: Edward Elgar, 223–43.
- US Department of Education (1999) *Satisfaction with TRIO Programs: Final Report*. Washington, DC: US Dept of Education.
- van den Hove N, Roelandt TJA, and Grosfeld T (1998) *Cluster Specialisation Patterns and Innovation Styles*. The Hague: Ministry of Economic Affairs.
- van der Linde C (2005) Cluster und regionale Wettbewerbsfähigkeit. Wie Cluster entstehen, wirken und aufgewertet werden. In Cernavin O, Führ M, Kaltenbach M, and Thießen F (eds) *Cluster und Wettbewerbsfähigkeit von Regionen. Erfolgsfaktoren regionaler Wirtschaftsentwicklung*. Berlin: Duncker & Humblot, 15–33.
- Wassermann S, Faust K (1999) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- White H, Sinha S, and Flanagan A (2006) A review of the state of impact evaluation. Paper presented at the 5th meeting of the DAC Network on Development Evaluation.
- Yin RK (2003) *Case Study Research: Design and Methods* (3rd edn). Thousand Oaks, CA: SAGE.

Claudia Schmiedeberg is specialized in macro- and microeconomic empirical research in the fields of innovation, regional development and industry structure. Address: University of Hamburg, Department of Economics, Chair of Economic Policy and Industrial Economics, 20146 Hamburg, Germany. [Email: schmiedeberg@econ.uni-hamburg.de]