Inter-American
Development Bank

Competitiveness and
Innovation Division,
Institutions for
Development

TECHNICAL NOTE

No. IDB-TN-551

Evaluating the Impact of Cluster Development Programs

Elisa Giuliani
Alessandro Maffioli
Manuel Pacheco
Carlo Pietrobelli
Rodolfo Stucchi

July 2013

Evaluating the Impact of Cluster Development Programs

Elisa Giuliani
Alessandro Maffioli
Manuel Pacheco
Carlo Pietrobelli
Rodolfo Stucchi

Inter-American Development Bank
2013

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Evaluating the Impact of Cluster Development Programs / Elisa Giuliani … [et al.].
p. cm. (IDB Technical Note ; 551)
Includes bibliographical references.
1. Industrial clusters. 2. Business networks—Research—Methodology. I. Giuliani, Elisa. II. Maffioli, Alessandro. III. Pacheco, Manuel. IV. Pietrobelli, Carlo. V. Stucchi, Rodolfo. VI. Inter-American Development Bank. Competitiveness and Innovation Division. VII. Title. VIII. Series.
IDB-TN-551

http://www.iadb.org

Abstract[*]

Do the programs that aim to promote and develop industry clusters (also known as Cluster Development Programs, or CDP) work? Do they have an impact on enterprise development? This paper offers an insight into the methods that can help answer these fundamental questions through solid quantitative evidence. In general, results will depend on the level of coordination that is achieved and on the actions undertaken as a result of improved coordination and strategy-setting of the relevant actors. The techniques of Social Network Analysis (SNA) can be employed to assess the evolution of coordination among cluster actors, with the requirement that network indicators are observed before and after the implementation of the CDPs. While this particular analysis can assist in monitoring and assessing the process of coordination and its changes throughout the program, other qualitative and contextual information can also assist in interpreting the data and, thus, increase the reliability of results. However, in order to properly assess the impact of CDPs, their causality needs to be explored further by the application of additional quantitative methods. In fact, the effects cannot be attributed to the program itself, unless a proper counterfactual is built in, such as what would have happened to the beneficiaries in the absence of the program. By definition, this particular counterfactual cannot be observed, but the application of experimental and quasi-experimental techniques can help construct control groups of non-beneficiaries to approximate the counterfactual and assess the evidence with econometric techniques. Furthermore, a detailed observation of cases and specific interviews can help regarding the interpretation of results derived from these methods. The quantitative tools discussed herein are indeed complementary and not alternatives, with each applied as a means to strengthen the explanatory capacity of the other. Each tool requires specific and challenging data analysis that can be achieved with careful resource planning and the appropriate team skill set. The overarching objective is to build new and solid evidence on the effectiveness of CPDs and their respective policies.

JEL Codes: O22, O25, O29, Z13
Keywords: cluster development programs, policy evaluation, social network analysis

[*] This publication was made possible by funding from the IDB's "Institutional Capacity Strengthening Thematic Fund."

Table of Contents

Acronyms

3ie	International Initiative for Impact Evaluation
AFIP	Administración Federal de Ingresos Públicos (Federal Administration of Taxes
ANSES	Administración Nacional de Seguro Social (National Administration of Social Security
BADE	Base para el Estudio de la Dinámica del Empleo (Base for the Study of Employment Dynamics
CACyDP	Centro de Abastecimientos Comunitarios y Desarrollo de Proveedores (local community center for supplies and supplier development)
CDP	Cluster Development Program
CIA	Conditional independence assumption
ISIC (CIIU)	International Standard Industrial Classification (Clasificación Internacional Industrial Uniforme)
CMM	Capability Maturity Model
CN	Collaboration network
CNPJ	Tax payer firm registry number
EAM	Encuesta Annual Manufacturera (Colombia's manufacturing survey)
EIA	Encuesta Industrial Anual
FOMIN	Fondo Multilateral de Inversiones (Multilateral Investment Fund)
GDN	Global Development Network
GIGO	Garbage In, Garbage Out
IBGE	Instituto Brasileiro de Geografía e Estatística
ICT	Information communication technology
IDB	Inter-American Development Bank
IN	Information network
ISLE	Industry-specific local externalities
IV	Instrumental variables
KCP	Knowledge and capacity-building project

LATE	Local average treatment effects
LSMS	Living Standards Measurement Study
MAR	Marshall-Arrow-Romer
MIF	Multilateral investment fund
MTE	Ministério do Trabalho e Emprego (Brazilian Ministry of Employment and Labor)
NIH	National Institutes of Health, U.S.
OEDE	Observatorio de Empleo y Dinámica Empresarial (Center for Employment and Business Dynamics)
OI	Operational inputs
PDP	Productive development policies
PIA	Pesquisa Industrial Annual (Brazil's annual industrial survey)
PROCAMPO	Programa de Apoyos Directos al Campo (Farmers' Direct Support Program)
PSM	Propensity score matching
PSU	Principal sampling units
RAIS	Relação Anual de Informações Sociais (Brazil's annual social information report)
RCT	Randomized control trial
SII	Servicio de Impuestos Internos (Chile's internal revenue service)
SNA	Social Network Analysis
SAOM	Stochastic actor-oriented models (for network dynamics)
Stocnet SIENA	A suite of computer programs for the statistical modeling of social network data, currently containing the SIENA programs
TA	Technical assistance
TC	Technical cooperation
TDF	Technology Development Funds
TFP	Total factor productivity
UA	Unit of analysis
USPTO	United States Patent and Trademark Office

Introduction

This publication provides some practical guidelines on how to measure the effectiveness of programs that aim to promote and develop industry clusters (also known as Cluster Development Programs, or CDP). The framework of this paper has two distinctive objectives. The first is to focus exclusively on the quantitative approach, which does not imply that the qualitative approach is not an important factor, when evaluating CDPs. While it is strongly believed that both approaches are complementary, it is essential that specific recommendations on how to measure the effect of CDPs be made. The second section centers on the effectiveness of CDPs.

The second objective is a focus on CDPs effectiveness. There are important hypotheses that should be considered when evaluating CDPs, such as their relevance, effectiveness, and consistency. However, evidence of the effectiveness of CDPs requires immediate attention, both in terms of knowledge and accountability. This introduction addresses three key issues: (i) definition of Cluster Development Programs; (ii) what results are worth measuring; and (iii) how can the findings of this study be attributed to CDPs.

Cluster Development Programs: Definition

A methodological recommendation depends on an accurate definition of the subject for analysis; in this case, CDPs. To do this, an industry cluster first needs to be defined, which can be challenging in itself (Martin and Sunley, 2003), and then the public interventions required to promote its development need to be justified. Several definitions of industry clusters have been proposed over time.[1] Nevertheless, a baseline definition interprets an industrial cluster to be a geographic group of interconnected firms and associated organizations, specialized in the same or related producer activities. This brings differing industry- and location-specific externalities. Firms, for instance, may take advantage of the concept of economies of agglomeration to increase their innovativeness, given the exchange of information and technological spillover that can exist, due to formal and informal networks between them and their associated organizations.

[1] Becattini (1989: 132) defines the concept of industrial district as "a localized social and productive 'thickening'", held together by a "complex and tangled web of external economies and diseconomies, of joint and associated costs, of historical and cultural vestiges, which envelops both inter-firm and interpersonal relationships." A simpler definition is derived from the work of Michael Porter (1990), who defines industrial clusters as "a geographic concentration of competing and cooperating companies, suppliers, service providers, and associated institutions." For a comprehensive review of the industrial cluster concept's definitions, see Giuliani (2005).

Successful clusters display certain characteristics. One is that organizations can provide specialized services to a specific group of firms within the cluster, especially in terms of supporting business planning, training for standards compliance, and/or accessing specialized research that is provided by institutions within the same cluster. A significant element of clusters is the development of non-market-related networks that can stem from entrepreneurs and workers from within the same culture and society, within a specific region. This is expected to improve economic coordination and reduce transaction costs.

Economies of agglomeration and clustering are particularly relevant in the tourism industry. By definition, this industry is geographically concentrated, due to its dependence on the natural and/or cultural attractions of a specific region. In addition, the strong complementarity of its products and the services it provides boosts the effect of the industry's externalities, making the coordination among local agents even more critical. Moreover, the tourism industry requires an infrastructure that is able to facilitate the movement of travelers, as well as the utilities to make their stays possible. The sustainability of a tourism cluster, however, may be hindered by a negative agglomeration of externalities, particularly when it is dependent on the natural heritage of a region.

In general, economies of agglomeration are the result of a set of positive externalities that are, simultaneously, industry- and location-specific, and have been well documented since the influential work of Marshall (1920).[2] These externalities are mainly due to the presence of a localized knowledge surplus, as well as input-sharing and labor-market pooling. Clusters are groups that are highly conducive to knowledge spillovers, because they allow for the transfer of tacit knowledge, which is not easily shared. Tacit knowledge is embodied in people and is difficult to articulate through codified knowledge; effective transfer of tacit knowledge requires personal contact and regular and direct interaction in order to be transmitted and absorbed (i.e., usefully employed). Hence, informal conversations between technicians and/or workers, labor mobility, and imitative behavior are examples of the ways tacit knowledge is communicated. The

[2] Because of Marshall's influential work, this phenomenon is often referred to as "*Marshallian externalities.*" In more generic terms, related literature has also referred to the concept of industry-specific local externalities (ISLE). Henderson, Kuncoro, and Turner (1995) refer to these types of industry-specific externalities that arise from regional agglomeration as "localization externalities," in particular, when firms operate in related sectors and are closely located.

gains from knowledge spillover may only be recognized from within, or by, a related sector of locally concentrated firms.[3]

With goods suppliers being in proximity with, and in the same industry as the end-users (Fujita, Krugman, and Venables, 1999) and, likewise, technology producers with the users (Lundvall, 1988), benefits can be drawn from a group of input suppliers within the specific industry. This will increase returns in the production of intermediate goods. This scenario is more likely to occur in the case of industry-specific services, such as business consulting, machine repair, and quality standards certification. In addition, labor pooling within a specific location in a particular sector allows local workers to train in that specific industry, without fear of unemployment caused by irrelevant skills. The centralization of specifically skilled human resources, thus, increases the concentration of firms to one locality.

Aside from purely *Marshallian* externalities, scholars have emphasized the role of local social embeddedness. Firms in clusters may participate in social networks, which can reduce transaction costs and increase efficiency (Granovetter, 1985), escalate the origination and sharing of tacit knowledge (Powell et al., 1996; McDermott et al., 2009), reduce uncertainty by strengthening bargaining and lobbying power (Guillen, 2000; Khanna; Rivkin, 2001; Mesquita and Lazzarini, 2008), and enable cooperative action (Schmitz, 1995).

Key questions are when is, and why is, public intervention needed to promote industrial clusters. It is well known in the study of economic theory that markets often fail to allocate resources optimally, when there is a collapse in coordination. As originally indicated by Rosenstein-Rodan (1943), investment decisions are interrelated, and the investment of one firm can have a positive effect on the profitability of another's. This is particularly important when geographic proximity and industry complementarities create economies of agglomeration. As an example—and with reference to the tourism industry mentioned above—a company may have a profitable investment by building a restaurant, following the investment in a hotel by another firm. Coordination failures, therefore, can lead a local economy or an industry cluster into low

[3] This argument was first introduced by Marshall (1920), followed by Arrow (1962), and then Romer (1986) and, later, formalized by Glaeser et al. (1992) as the Marshall-Arrow-Romer (MAR) model. The endogenous growth theory (Romer, 1986; Lucas, 1988; Krugman, 1991) underlines the importance of knowledge spillovers and externalities to bring about self-reinforcement, an increase of returns-to-scale within a region, and support to the agglomeration of economic activities. A stream of empirical scholarly articles on this topic began with the influential contribution by Glaeser et al. (1992), who sought to examine the respective roles of specialization and diversity with regard to local and regional development (Greunz, 2004).

investment equilibrium, if local businesses and governments fail to coordinate investment decisions.[4]

In this context, we can define CDPs as those public interventions that foster the beneficial effects of economies of agglomeration by creating a set of incentives to overcome the coordination failures that hamper the development of some industries in specific localities. As mentioned above, the idea of promoting the formation and development of clusters is based on the assumption that business-level performance benefits from agglomeration and, in turn, agglomeration facilitates coordination.

The objective defined above may be achieved through different methods, differentiating CDPs significantly in terms of both conceptualization and implementation. However, most CDPs usually involve the following activities:

- Motivation and mobilization of cluster stakeholders (firms, support institutions, and public agencies) to analyze and compare performance, capacity, and the potential of the local production systems.
- Development of a vision and consensus to enhance innovation, productivity, and knowledge, among firms within the cluster.
- Development and implementation of an integrated set of interventions to promote innovation and combined learning, overcome technological and environmental constraints, strengthen local innovation, and promote joint investment in local public and collective goods.

These interventions clearly aim to improve coordination at the local level. This methodological toolkit will provide a comprehensive approach to CDPs, often adopted by governments and international organizations, and will include some or all of the activities listed above.[5] The rationale for this kind of intervention is based on the assumption that coordination failures emerge at a very preliminary stage in the development of an industry cluster, and that public support is required to facilitate interaction and coordination among private-private and public-private organizations. The **initial stage** of intervention usually includes the preparation

[4] For a review on coordination problems in development, see Hoff (2000). For clusters and coordination failures, see also Pietrobelli, Casaburi, and Maffioli (2013).
[5] The Inter-American Development Bank (IDB) adopts this comprehensive approach. See Pietrobelli and Stevenson (2011).

and adoption of a CDP to coordinate and prioritize the various investment decisions of a cluster's stakeholders for a systematic path towards development, where industry- and location-specific externalities are accounted for. As a first step, public intervention can reduce those transaction costs that hinder coordination among agents within a cluster.[6]

The activities included in a CDP are implemented during the **second stage**. Investments often play a key role in the creation of knowledge and the adoption of technology, two of the major sources of industry and location-specific externalities. For this purpose, cluster technology development centers, generally, are financed as a group. At this stage, CDPs normally include the co-financing of public infrastructure and club goods. Therefore, public funding is not only used to reduce coordination costs, but also as a catalyst for new investment projects, both private and public. Sometimes these programs also co-finance the purchase of machinery, equipment to be collectively managed and used in co-financed technology centers by groups of firms located in each cluster. Thus, industry-specific production assets become club goods for the businesses involved in the cluster development plan.

Measuring the Results of CDPs

The evaluation of the two-stage intervention described above—the first mainly focused on solving pure coordination problems, and the second centered on implementing a set of coordinated private and public investments—involves the use of various kinds of measures. The assessment of the first stage requires measures that can capture the creation and/or strength of different kinds of interrelationships between firms and/or other stakeholders during the preparation and adoption of a CDP. A fundamental approach for this purpose is the Social Network Analysis (SNA). As this paper shows later, SNA applies graph theory to measure networks in quantitative terms (Wasserman and Faust, 1994) and to examine their structural properties, as well as to identify the position that differing agents occupy within networks.

The second stage requires measures relating to the expected effects of the implementation of the activities included in a CDP. Since the goal of a CDP is always related to fostering local economic development, this toolkit will focus on the outcomes related to increased output. Concentration will, thus, primarily be aimed at productivity and its related outcomes, such as

[6] Maffioli (2005) illustrates different roles that public agents can play in reducing coordination costs and inducing the formation of socially desirable networks that would not spontaneously arise otherwise. In particular, a public actor can play the role of coordinator, and modify the structure of firms' payoffs announcing that it is willing to cover part of the connection and coordination costs.

innovation, investment, export, and market share. Given that incentives to foster coordination are put in place at this stage, the strength and variety of firms' relationships with other businesses and the relevant cluster's organizations will be a key intermediate outcome to be evaluated. As already mentioned, the SNA tool provides a way of measuring that can be used for this purpose. This toolkit mainly relates to measurement at the business level (without disregarding measurement at the level of specific clusters and localities), and includes attention to potential spillover effects and externalities.

Attributing Observed Results to CDPs

This toolkit will especially focus on the question of whether CDPs work or not. In other words, are CDPs able to deliver expected results? In seeking an answer, two parallel evaluation processes will take place. One will assess the progress of the first stage of a CDP, focusing on whether and how coordination among cluster stakeholders changed over time. The other will focus directly on impact evaluation to establish whether the CDP is the key driver of the observed results. Assessment of the process of coordination is easier, in terms of data requirement and analyses, than evaluating the causal impact of CDPs on cluster performance, although both are important complementary dimensions of the CDP evaluation process. An appraisal of the coordination among cluster stakeholders requires that network indicators are observed before and after the implementation of a CDP, but it does not necessarily require a control group. It is obvious that this analysis will only assist in monitoring the process of coordination and its changes along the project's lifetime while, at the same time, qualitative and contextual information can be collected to aid in the interpretation of data and increase the reliability of results. Sections 2 and 3 will illustrate, in detail, how the analysis can be carried out.

To properly assess the outcomes of a CDP, more sophisticated methods are essential to determine causality, in particular, one has to deal with the fundamental problem of attribution effects. For this purpose, several techniques have been developed, according to literature relating to policy evaluation. Most of these techniques are based on the idea of constructing a proper counterfactual scenario that represents what might have happened to the beneficiaries in the absence of a program. Because this counterfactual scenario, by definition, is not apparent (a firm cannot simultaneously be in a position to receive or not receive a specific benefit), control groups of non-beneficiaries have been constructed to approximate the counterfactual scenario (see Section 4 for an in-depth methodological discussion).

Structure of the Toolkit

The toolkit is structured as follows. Section 1 will define the typical research questions and specific objectives of a CDP evaluation study. Section 2 will describe a set of key indicators to be used for measuring coordination and evaluating the impact of a CDP on performance. Section 3 will include data requirements and the methods used for data collection. Section 4 will discuss the attribution problem and outline the methods for assessing the CDP's impact on performance. Finally, Section 5 will conclude with a discussion relating to the practical issues involved in a CDP evaluation study.

1. The Logic of CDP Research Questions in Evaluation

In defining what the specific objective of the analysis is, discussion will center on the justification, intervention logic, and structure of a typical CDP. Given the variety of approaches relating to a CDP that have been developed over time by policymakers around the world, this will prove simple. However, as mentioned in the above Introduction, it is an essential step towards defining the scope of this methodological proposal. Moreover, it will reflect a specific type of CPD that has been widely adopted in the Latin America and the Caribbean (LAC) region, often with the support of the Inter-American Development Bank (IDB).

1.1. CDP in Practice: Some Elements of the IDB Experience[7]

1.1.1. The First Stage: Strategy Design and Coordination

As mentioned above, a CDP typically starts with an initial phase, where public intervention is needed to coordinate and prioritize the various decisions of cluster actors within a proposed cluster or area. This takes place through the preparation and adoption of cluster development plans, which set out, in detail, agreed initiatives to promote its development. In some cases, the coordination of activities may be less formal, and may include marginalized actors in local communities. Box 1 illustrates how the first phase of a CDP is defined.

Box 1. Example of the Design and Coordination Stage of a Typical CDP

Cluster Competitiveness Support Program for Minas Gerais, Brazil

A specific set of activities was designed within a program, in the Brazilian state of Minas Gerais, to assist in the coordination of different clusters, to promote institutions through local government entities (EGLs), as well as prepare proposals for improved competitiveness (*Planos de Melhoramento da Competitividade, or* PMCs). It followed a participatory process engaging the clusters' firms and related institutions. The PMCs had to include, at least, (a) a definition of the cluster's type of business and the firms and related institutions that comprised it; (b) a medium- and long-term strategy for positioning the business within the domestic and international markets; (c) a diagnostic study of the current status of the cluster, including its social and environmental approaches and strategy to achieve the various objectives proposed; and (d) a proposal for the services and contributions under each of the program's lines of support, with justification for each line of support. The last component was the source of finance required for the diagnostics, workshops, surveys, studies, and TA (technical assistance) under the PMCs for all the clusters in order to strengthen the EGLs.

---

[7] This section draws on Pietrobelli and Stevenson (2011) (see http://www.iadb.org/en/publications/publication-detail,7101.html?id=26526%20&dcLanguage=en&dcType=All), where these elements are analyzed in greater detail.

This first phase encompasses the key mechanism at the base of a CDP in resolving coordination problems and creating or strengthening the different types of inter-organizational networks. While seeking to resolve coordination problems is a constant throughout the overall implementation of a CDP, it is clear that cluster competitiveness development proposals play a key role in defining how coordination can occur and what incentives there are to encourage investment. In this regard, the effort to determine and strengthen cluster *governance* in the initial stage is especially relevant.

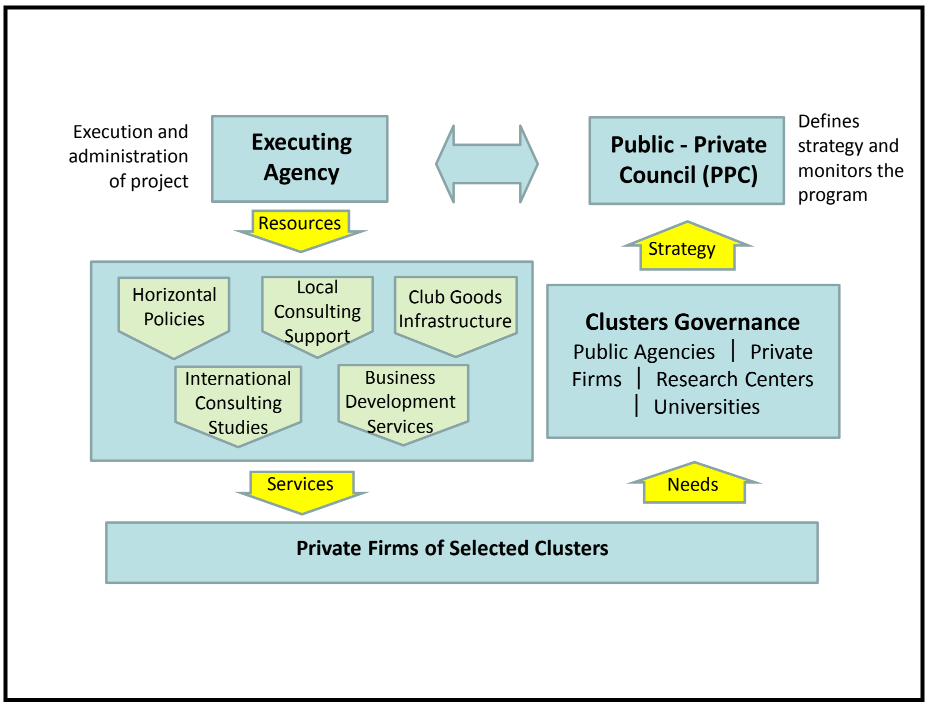
Strategy design and relevant coordination requires thorough research, analysis, and consultation. Essential to this are the selection of clusters and the identification of key stakeholders (bold entrepreneurs, as well as local leaders, civil society, and committed public officials). Subsequently, workshops and consultation should take place—usually beginning at the policy level, with the selected clusters and sectors, and then within each cluster—to establish the priorities and activities. During the diagnostic stage, continuous analyses and consultation with stakeholders are carried out to identify the main barriers hindering the competitiveness of the industry and of individual firms.

Another essential component in the initial phase relates to the supervision of a program as a means to promote public and private dialogue, both at the program and cluster levels. In this regard, the institutional capacity of public institutions engaged in this type of program and the degree of liaison with the private sector are crucial to the success of a CPD.

Included in the inception phase are a detailed analysis of each cluster to identify the gaps; a benchmark, created by comparing other sectors and industries in comparable countries; and a list of the main activities to be carried out during implementation of the program. The aim is to avoid the failures that some markets have experienced and ensure coordination in private and public activities. In order to maximize impact, a series of parallel activities to improve the business climate is necessary. This may include removing the regulatory bottlenecks for starting and managing a firm, investing in key infrastructures to reduce market access costs, or implementing training programs that improve the human resource capacity to serve the local industry.

As mentioned above, one of the most challenging aspects of a CDP is the complexity of its implementation and its supervision. Usually a national or local public institution administers the program through an executing agency. This can be a government agency in charge of the promotion of private sector development and competitiveness at the national or local level, or a unit created by the relevant ministry. In addition, the executing agency is usually in charge of the coordination between the different public entities, both at the national and subnational levels. Guidelines, strategy, and monitoring of the program's results are typically the responsibility of a public-private strategy council. The selection and number of participants and the legal nature of the council are defined on a case-by-case basis, with stakeholders from the public and the private sectors included. In some cases, a cluster will have its own governance structure, presenting its proposals to the council. The council, in turn, will consider the governance structure representing the interests of the local actors and, thus, determine which entity is better placed to administer it. It could be a university, an association, or a government agency.

Figure 1. Typical Institutional Structure of an IDB Cluster Program



Source: Pietrobelli and Stevenson (2011).

Figure 1 exhibits the typical structure of a cluster development program. The executing agency interacts with a public-private council that establishes the strategic objectives of a program, as well as monitors it. Resources are allocated by the executing agency to carry out the various activities that are instrumental to promoting the competitiveness and development of the cluster, such as the provision of business development services, investment in the infrastructure of club goods, or contribution of TA. Private firms participate in the program, and collaborate to create and oversee the administration of the cluster, together with other relevant organizations, such as government agencies, research centers, and universities.

1.1.2. The Second Stage: Implementation of a CDP

The second phase of a CDP includes the implementation of specific activities that may vary, depending on the program's objectives and on the analysis and proposals developed in the initial stage. Firms may have access to the resources of a CDP, according to a pre-defined set of criteria, and these can be used to finance the projects that are selected to support the strategic objectives of the proposal for enhancing the competitiveness of a cluster. These projects are presented to the various participants in the cluster, and must prove to be sustainable, both in financial, environmental and social terms, and innovation, benefitting the cluster as a whole and generating externalities and local public goods. The main components of a CDP, therefore, usually include investment in knowledge creation and technology, two of the major sources for industry and location-specific externalities. In addition, a program may include components for not only institutional strengthening, but also for the promotion of overall business cooperation and integration, and for reforms to enhance the business climate. Some programs may also include a factor for access to finance.

1.2. Key Questions Relating to the Evaluation of a Typical CDP

1.2.1. Effectiveness in Terms of Increased Coordination and of Program Delivery

As mentioned above, the main objective during the first stage of development of a CDP is the coordination of the different actors in a cluster. In the CDP discussed above, it is expected that coordination has evolved and become strengthened as the project moves towards implementation in the second stage. Hence, a key goal of a CDP impact evaluation study should be the measurement of actor coordination over the duration of a project and beyond.[8] Focus should, therefore, be placed on identifying the changes in the indicators that relate to increasing coordination, which would determine modification of the relationships or—to use a more

[8] It should be noted that each CDP generates a number of coordination activities that start during the first stage and that are meant to last and change during the implementation stage and beyond that.

common term in the area of international development—linkages among cluster actors, as the program progresses. The overall key question at this stage should be: *How did the CDP affect the coordination among the different cluster actors?*

More specifically:

1. *How did the CDP affect the structure of the linkages (i.e., network) among cluster actors during the first and second stages of the project?*
2. *Are these changes, if any, in the network a sign of increased coordination?*

These two questions imply both the **quantity** and **quality** of the linkages among cluster actors. Specifics relating to changes in coordination will depend on the design and objective of each CDP. Hence, program evaluators will need to model their coordination research questions according to the specific objectives of each CDP. For example, some CDPs may require strengthening horizontal coordination between firms that operate at the same stage of the value chain. If policymakers intend coordination to be stronger as one of the results of a CDP (i.e., the number of linkages between relevant firms is increased), then an important question would be: *Did the density of inter-firm linkages increase during and after the CDP?*

Other CDPs may focus more on a vertical coordination between a group of small firms and one or two larger corporations (such as subsidiaries of multinational corporations or large domestic firms) or major organizations (such as a food laboratory for agro-industrial producers, a design institute for garment producers, or a metrology/quality control/certification organization in the case of metalworking producers). In this case, program evaluators should not only seek whether overall connectivity has increased but, rather, whether such coordination has moved towards a different form of governance. The appropriate question to raise would be: *How has the configuration of network ties, reflecting the governance mode within the cluster, changed during the first and second stages of the CDP?*

Box 2 offers examples of CDPs that targeted different kinds of networks, showing that they often target both horizontal and vertical linkages among cluster actors (and others in their proximity). For appropriate research and evaluation question(s), program **evaluators should have a clear knowledge of what the coordination objectives are in a CDP,** which should have been included at the policy stage. CDPs often aim at increasing coordination without a specific concept in place. On the other hand, it is possible to extrapolate more specific objectives, as those illustrated earlier. The more evaluators there are to clarify this point, the more they will be able to develop key questions relating to the coordination evaluation.

Box 2: CDPs That Target Different Forms of Relationships

The CDP in Cali, Colombia, strengthened collaboration among more than 200 producers of *Uva Isabella,* a company of Isabella grape products. The program's goal was to improve trade relations with larger international markets. In terms of linkages among cluster actors, this indicated stronger relationships between many small producers and a few larger agro-industrial producers (vertical linkages), as well as an improved horizontal coordination among small producers to introduce an electronic monitoring and evaluation system and better quality control. This would reinforce the bargaining power of the small producers *vis à vis* the larger agro-industrial producers. Similarly, the CDP of the Cordoba software cluster in Argentina aimed at not only strengthening horizontal collaboration among producers for TA to gain Capability Maturity Model (CMM) certification and lobby for sector regulation, but also strengthening vertical collaboration with tertiary training institutions for the required technical skills. In the CDP of Santiago de los Caballeros, Dominican Republic, furniture producers developed stronger horizontal relationships to jointly obtain inputs and technical assistance, as well as jointly bid for tenders. The CDP of Atuntaqui, Ecuador, included the enhancement of the technological progress of the garment manufacturers in Atuntaqui through, among other things, stronger linkages with technical service centers that were also supported by the project.

Note: The above examples are drawn from various projects of the Multilateral Investment Fund/IDB and United Nations Industrial Development Organization. Details can be found in Dini et al. (2007).

1.2.2. Effectiveness in Terms of a Firm's Increase in Performance

Key questions related to the expected outcomes of CDPs on the performance of firms are the following:

1. *What is the effect of the CDP on the performance of firms?*
2. *What is the effect of the CDP on the performance of firms, due to a change in their respective social network?*

For a response, the causality needs to be addressed more importantly than the correlations; that is, what is the change in performance *due to* the CDP. Of interest is the effect of the CDP, both on the firms participating (direct beneficiaries), as well as on the firms that do not participate in the CDP, but receive spillover from participants (indirect beneficiaries).

There are other questions that program evaluators will need to seek responses to. First is the effect of the CDP on the different measures of performance, such as on a firm's productivity or sales growth. Second is the heterogeneity of the effect—whether the effect of the CDP is larger (or smaller) for small firms, or whether it is different for firms in different industries, or for firms in a different stage of development. A third question relates to the dynamics of the effect. For example, is the effect of the program a result of its initial years of undertaking, or will it be necessary to wait some years before the effect can be identified? Further, still, is the question that relates to some CDPs using different instruments, such as credit and/or TA. In this case, what is the effect of each type of instrument or combination of instruments? It is also possible to combine questions, such as whether the effect on small firms is more rapid than that on larger ones, or whether a particular instrument has a consequence for small or newly established businesses.

It is necessary to note that addressing the heterogeneity of the effect, itself, or the effect of different instruments, requires a large number of treated and untreated businesses firms. If there are not enough treated (or control) firms in some categories of firms or type of instruments, it is not possible to detect the effect of the CDP (even if there is one). Section 4 of the toolkit will include the ideal quantity of firms needed for evaluation.

2. Indicators

2.1. Social Network Analysis: Measuring CDP Effects on Coordination

Given that strengthening coordination is one of the objectives of most CDPs, this toolkit will provide guidance to program evaluators in the employment of quantitative methods to measure achievements. The methods are based on the application of SNA, mentioned above, and are detailed in more depth in Giuliani and Pietrobelli (2011). SNA uses graph theory to measure coordination, which, in the context of this publication, refers to the inter-organizational networks formed by cluster actors. SNA methods can be applied for at least two purposes: first, to measure the position of individual actors within a network and, second, to measure the way linkages are distributed within a network (i.e., the network structure). These are further described below.

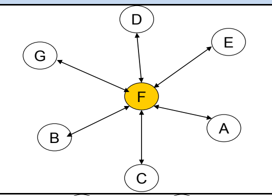
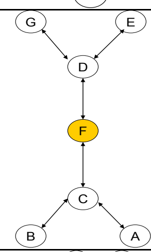
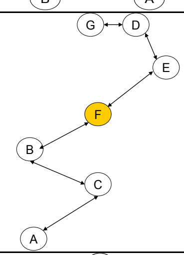
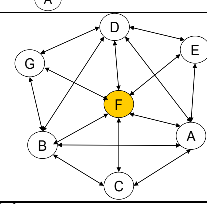
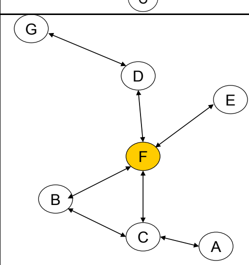
2.1.1 Measuring an Actor-level Position Within a Network

There are different measures that can be employed by CDP evaluators to establish where actors are positioned within a network, depending on the objectives of the program. Table 1 illustrates some of the most popular concepts and measures. Each measure can be applied to observe how the position of key actors within a cluster changes during the different stages of a CDP. Degree centrality (Table 1 (a)) is measured by the number of direct ties an actor maintains with others within the network, and it signals the access to the resources that are transferred through the network, such that highly centralized actors have better access to resources than those who are disconnected or on the periphery.

In other cases, it is not the number of direct links that results in a positional advantage. Centrality, in certain cases, empowers, and empowerment is a result of being connected to those who are weaker, as being connected to powerful others that have many potential trading partners reduces one's bargaining power. Hence, the power of one actor within the network may be related to the many direct links that actor has, as well as to the little ties of its direct contacts (Table 1(b)). However, the opposite may be true if the “central” actor, such as a large multinational firm, is expected to distribute services (assets) to the other cluster firms.

Another measure of centrality, “betweenness” centrality, relates to the degree to which one actor is vital in a network in connecting actors that otherwise would not be connected (Table 1 (c)). Betweenness centrality is critical if program evaluators wish to identify and follow, over time, the bridging actors; that is, actors that are indispensable in maintaining parts of the network (e.g., two different sectors) connected to each other. This may be the case in terms of firms, agencies, and buyers, which play the role of "brokers" within a cluster.

Table 1. Examples of Network Positions, Beneficial Effects, and Limitations

SNA concept	Brief description	Illustration	Advantages/ benefits	Limits
(a) Degree centrality	Number of direct ties an actor has with others within the network.		Easy access to information, knowledge, and any type of resource.	Too many connections can be time-consuming, not always rewarding.
(b) Bonacich centrality	Centrality of an actor, dependent upon the centrality of its direct contacts (alters).		Power (if alters have low centrality), access to resources (if alters have high centrality).	Too many connections may overload the actor.
(c) Betweenness centrality (see also "structural holes")	Degree to which an actor is able to connect others that will be otherwise disconnected.		Gatekeeping, influence, dependence, control.	If there are only a few actors with high betweenness centrality, they may easily disrupt the network (vulnerability risk).
(d) Closed ties	High local connectivity between an actor's alters.		High trust, quality knowledge, joint problem-solving, reduction of transaction costs.	Too much closure is detrimental and leads to lock-in.
(e) Structural holes (see also "betweenness centrality")	When an actor's alters are not/are poorly connected to each other.		High level of knowledge diversity, high opportunities for creativity and radical innovation, efficiency and control in links.	Does not have the advantages of network closure.

Source: Giuliani and Pietrobelli (2011).

Positional advantages also depend on the degree to which the alters of actors are (or not) connected to each other. Concentrated local connectivity of alters (Table 1 (d)) often conveys a high level of trust and quality of knowledge to relationships, while the lack of connectivity between alters (technically, the presence of "structural holes" (Table 1 (e)), brings access to diverse knowledge and resources, since actors that are not connected to each other tend to be different.

2.1.2 Measuring a Network's Structural Properties

The advantages and limitations of different network positions discussed, so far, refer to the individual firms and may not hold true for a cluster, as a whole. In fact, the high centrality of some businesses, in some cases, may even be detrimental for other firms within the cluster. To understand the overall advantages of a network, it is necessary to study its structural properties on a case-by-case basis, and consider the underlying business and industry strategies.

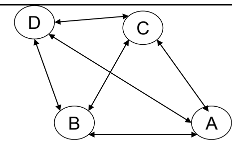
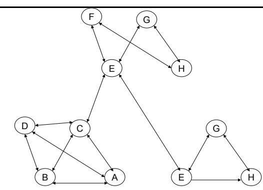
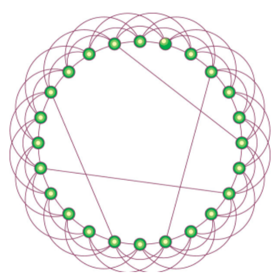
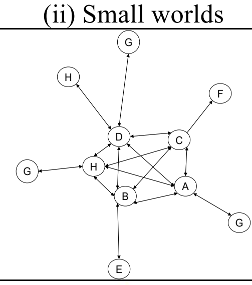
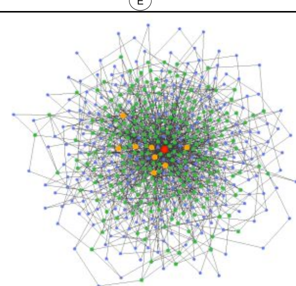
Table 2 presents examples of measures for network structure. The study of network structures often includes a process for identifying the subgroups of actors that display a higher-than-average connectivity than the rest of the actors within the network: "…subset of actors among whom there are relatively strong, direct, intense, frequent, or positive ties" (Wasserman and Faust, 1994: 249). Single cohesive subgroups are known as "cliques," which incorporate at least three actors connected to each other within a network. A giant clique could include hundreds of actors, each connected to the other (Table 2 (a)). Provided that the links have some valuable content, a network with cliques has the advantage of ensuring a cooperative environment, where social monitoring, trust, and resource-sharing would be high. Furthermore, a cliquish network can be considered to be a non-hierarchical space, where resources are distributed in a highly egalitarian manner. In reality, however, very few networks are fully cliquish. This is why most CDPs, generally, aim to increase network density and to achieve highly cliquish structures. However, it is important to bear in mind that networks that are too densely connected may have drawbacks, given that many links are costly to maintain and are not always rewarding. As an example, in the case of an export consortium with a multitude of participants, more ambitious coordination activities (such as brand creation, collective quality control, and insurance) can be harder to implement.

Contrary to previous belief, other network structures—while being less dense—may also bring advantages for firms and their clusters. In small networks (Table 2 (b)), cliques can be connected to each other by sparse or weak links (Table 2 (b) (i)). In larger networks, this structural feature has become known as "small world" (Table 2 (b) (ii)). What characterizes "small world," is that actors have dense connections to their neighbors (local cliques) and, at the same time, have sparse connection to other distant actors (clique-spanning links). This particular structure benefits from a high level of local trust, and is conducive to a cooperative environment. At the same time, it ensures that local cliques do not remain isolated, given that some of its members are also connected to distant actors. Small worlds are, therefore, efficient structures, minimizing the number of links and, at the same time, allowing diverse knowledge coming from distant actors to be accessed or exchanged by local actors. This occurs when local clusters openly engage in relationships and collaborate with distant actors, buyers, and global value chains—a determinant of the remarkable industrial success of Taiwanese firms (Guerrieri et al., 2003).

Frequently, cluster networks appear to be organized along a "core-periphery" structure (Table 2 (c)), characterized by a densely connected core (a clique-like subgroup) and a set of "hangers-on" actors (i.e., the periphery), which are loosely connected to the core and very loosely connected among themselves (Borgatti and Everett, 1999). Core-periphery structures tend to signal the presence of an elite group (the core), which exchanges resources and shares assets with great frequency, while peripheral firms are often at a disadvantage. This structure reveals a lot to policymakers, since it highlights the existence of an elite group of firms and how they relate with the other firms or organizations on the periphery. A CDP that reduces the exclusion of producers from a local network should change the core-periphery structure into a less hierarchical one. It is essential, therefore, to understand whether the latter structure is present within a cluster and, if so, whether, and in what direction, it has changed during the various stages of the CDP. There could be an indication of a process of selective creative destruction, gradually marginalizing the less efficient and dynamic firms in the cluster. This, nevertheless, could be considered positive, if it coexists with a dynamic group of new firms.

To reiterate, program evaluators need to ensure that the evaluation research undertaken is relevant to the type of structure of the network. The Chilean wine cluster of Colchagua is a valid example of this structure (Giuliani and Bell, 2005). The firms with the strongest knowledge base in the cluster constitute the core, while the weaker ones tend to be positioned at the periphery. Giuliani (2013) found that the core firms reinforced their position over time, while many of the weakest firms either exited the industry or remained peripheral. This core-periphery structure bears some advantages: it enables the circulation of high-quality and constructive knowledge among the densely connected core firms that have considerable potential to upgrade the knowledge base and facilitate the transfer of this knowledge. At the same time, the continuity of a core-periphery structure will minimize the risk that transferred knowledge can be "downgraded" by firms with a weaker knowledge base, since these latter firms are consistently in marginal network positions.

Table 2. Examples of Network Structures: Advantages and Limitations

SNA concept	Brief description	Illustration	Advantages/benefits	Limits
(a) A single cohesive set (clique)	A dense network, where (almost) all actors are connected to each other.		High level of trust, cooperation, support, and social monitoring.	Redundant linkages, high opportunity costs, risk of "getting trapped in their own net."
(b) Small worlds	Non-overlapping cliques (high local closeness), connected by a few links with distant actors.	 (i) Small cliquish structure  (ii) Small worlds	Efficient structure, local dense links (trust and cooperation), and distant links (competitive advantage, search for diversity).	Success is dependent on actors with local and distant links.
(c) Core-periphery	A core of densely connected firms and a periphery with a few connections to the core and little intra-periphery links.		Core actors, as well as actors connecting the core to the periphery, may have advantages.	Hierarchical structure; peripheral actors may suffer exclusion; uneven network structure.
(d) Scale-free	Few hub firms holding all the connections, "orchestrating" a network.		Hierarchical and organized management of the value chain.	Very uneven structure; polarization of power and resources in a few actors; vulnerable to attacks to hubs.

Source: Giuliani and Pietrobelli (2011).

Finally, there are structures that are even more hierarchical than the core-periphery one described above, such as the “scale free” networks (Table 2 (d)). These are characterized by a few actors behaving as hubs and holding an outstanding number of connections, while the vast majority of actors are poorly connected (Barabasi and Albert, 1999). This type of structure is, generally, the result of a rich-get-richer mechanism of network growth, by which some actors tend to gain more linkages over time at the expense of others, and become more powerful and central within the network. Such networks tend to be very hierarchical and distribute resources and knowledge in a very uneven and polarized way. Clusters that are characterized by scale-free networks reflect the well-known “hub-and-spoke” cluster typology (Markusen, 1996: 298); that is, a cluster typified by “(…) a business structure is dominated by one or several large vertically-integrated firms surrounded by suppliers (…)” Detecting structures that fit, or are close to fitting, a scale-free network is important for those CDPs that aim to create a hierarchical structure, led by one or a few leading actors, and aim to monitor how its changes during and subsequent to the CDP implementation stage. Scale-free networks need not be anchored to a firm. In other cases, policymakers may have an interest in promoting the centrality of a public or public-private organization, which is expected to facilitate the upgrading of a firm (see the case of the Mendoza wine cluster in McDermott et al. [2009]).

2.1.3 Considerations Regarding the Use of SNA Measures and their Meaning

Caution is advised when determining the advantages and disadvantages of different network positions; they need to be considered in terms of where the networks are formed, including the quality of the linkages (their nature and value). In addition, the characteristics of the individual actors within the network will reveal to what degree an actor is able to leverage its positional advantage and what the overall benefits of the network will be. For example, a network of unskilled workers is less likely to generate innovation compared to one with qualified engineers, even if the two networks are alike in terms of structure. Likewise, two actors in the same network position (e.g., high betweenness centrality) may take advantage of their position, according to their respective skills and those of the other actors with whom they each are linked.

Moreover, impact evaluators should recognize that it is not always possible to find a network position or structure that is optimal because this may vary from case to case. The discussion above highlighted the potential benefits and the drawbacks of each type of position and structure. SNA can be used to track changes along the duration of the program's implementation, as well as subsequently. The advantages or disadvantages that such changes may expose will need to be assessed, on a case-by-case basis, and in the context of when the program had been implemented. For example, the promotion of a dense network may be advantageous in a case where a CDP is meant to create or increase a cluster's social capital. In an artisan cluster, where small entrepreneurs do not trust each other and live in constant conflict, it may be beneficial to increase the network's density, so as to encourage trust between them.

CDPs, on the other hand, may have other objectives. A group of densely connected aspiring entrepreneurs may already exist before a CDP is put in place, in which case the policy objective is to increase their connectivity with weaker firms. This would call for a core-periphery structure, where the core is represented by the aspiring entrepreneurs and the periphery is composed of the weaker firms. In a recent impact evaluation study, carried out in a cluster of the electronics industry in Cordoba, Argentina, Giuliani and Matta (2012) discovered that this CDP had had an impact on the development of new linkages, subsequent to policy treatment, although they did not contribute to discounting the pre-existing network structure, which had been measured two years after the start of the CDP. The authors observed a highly centralized structure, somewhat resembling a core-periphery structure (where the core firms played a critical role in maintaining the network connections over time, as well as ensuring its structural stability in a path-dependent manner). The authors concluded, for the above reasons, that this type of network structure could be beneficial for members—and the cluster, as a whole—provided the core firms were prepared to collaborate with new entrants and with the existing actors on the periphery. Failing this would result in those latter firms not benefiting from spillovers and geographic proximity.

To summarize, it is essential that policymakers and team members discuss together the basic features of a CDP's potential network structure at the end of program treatment. In doing so, they must take into account all the factors relating to the firms within the cluster.

2.1.4 Stochastic Actor-oriented Models for Network Change

Using SNA, it is possible to study the effects of a CDP on network changes. Networks are dynamic by nature, and relationships tend to change over time, with new ones being formed and others cemented or disbanded. Such changes may be the result of concurrent effects and random residual influences. The non-random effects that drive network change may be classified into three types: (1) those that relate to the structural position of the actors within the network (i.e., structural or endogenous); (2) those that depend on the actors' own attributes or characteristics (i.e., actor covariates); and (3) those that relate to variables depending on a pair of actors (i.e., dyadic covariates). Stochastic actor-oriented models (SAOM) for network dynamics are considered to be the most favorable models that allow for the statistical assumption of network dynamics, by simultaneously analyzing the impact of these different types of effects on network change. Excluding the technical aspects of SAOM methodologies (Snijders, 2001; 2005; Snijders et al., 2010) these approaches can be used to determine whether a firm's participation in a CDP (including within a set of CDP initiatives) relates to the formation of new linkages, controlling for a set of other structural, covariate, and dyadic covariate effects that could simultaneously influence the formation of new linkages.

To apply this model, at least two studies will need to be undertaken (two-wave panel analysis) to collect the relevant data prior to, and after, implementation of the CPD. It is also recommended to have a third one-wave collection undertaken *during* program implementation, as well as to use full network data, although entry/exit and missing data are allowed in the model.

2.2. Measures of Efficiency (Productivity and Intermediate Outcomes)

To assess the impact of a CDP on the performance of firms, there are several indicators, which can be used as outcome variables.[9] These relate to productivity, exports, innovation, and employment (Box 3).

2.2.1. Productivity

There are several measures relating to the productivity of firms, two of which are to consider the productivity of one input, such as labor productivity, or to consider the productivity of all inputs, such as total factor productivity (TFP).

Measuring productivity is not a straightforward task. In the case of labor productivity, it is measured as the ratio of output to labor input, and requires some attention. Ideally, it should be measured as the quantity of produced goods, together with the hours worked to obtain those goods. In practice, however, businesses produce various goods, and labor may not be

[9] As noted by Gertler et al. (2011), good indicators have to be SMART; that is, Specific, Measurable, Attributable, Realistic, and Targeted.

homogeneous. In this case, in terms of output, the real value of sales (or the real value added) should be used. In terms of labor input, the number of employees (or the real labor cost) is normally used. Nominal variables need to be deflated to obtain real variables, and in order to do this effectively, a company-specific price index should be applied although, in practice, most studies use industry-level price indices, given that data on prices at the firm-level may not be available. For information on the use of company-level price indices, see Foster et al. (2008).

With regard to TFP, all matters related to the measurement of output and inputs are present, although there are additional ones that should be considered in relation to the methods applied. TFP is not observed but, generally, can be obtained as a residual; that is, change in output cannot be explained by change in input. Several methods exist to obtain TFP, all of which hint to varying assumptions relating to the production process and the degree to which the market is competitive. This suggests that each method has strengths and weaknesses.[10]

Box 3. Required Information for Measuring Outcome Indicators

(a) Firm-level productivity
- Annual value of production
- Number of employees (number of permanent and temporary employees)
- Materials (value of materials used in the production process)
- Capital stock (net book value of machinery and equipment)
- Labor cost.

(b) Export-related indicators
- Value of exports
- Number of products exported
- Number of countries to where exports are made.

(c) Innovation-related indicators
- R&D expenditures
- Product innovation
- Process innovation.

(d) Employment-related indicators
- Number of employees by type of education
- Wages.

[10] For further literature on estimating productivity at the firm or establishment level, see Hulten (2001), Bartelsman and Doms (2002), Van Biesebroeck (2007), and Syverson (2011).

The simplest way to measure the change in TFP is using the Solow Residual. Suppose the production function has the form $Y_{it} = A_{it} F(L_{it},K_{it})$, where Y_{it} is the annual real value added, L_{it} is the number of worked hours during the year, K_{it} is real value of the stock of capital, and A_{it} is TFP. Then, if F(.) is homogeneous of degree one (i.e., under constant return to scale), the change in TFP can be obtained by[11]

$$\Delta a_{it} = \Delta y_{it} - \alpha_{Lit} \Delta l_{it} - \alpha_{Kit} \Delta k_{it}$$

where lower cases represent the log of upper cases, Δ measures the annual change, and α_{Lit} and α_{Kit} are the average of the cost share of labor and capital in t and t-1. The cost share of labor, α_L, is the ratio between labor cost and total cost, and the cost share of capital is equal to $1-\alpha_L$.[12]

When the interest is in the productivity of some industry or of a region, it is also important to study the reallocation of resources between firms. If more productive companies gain market share, there is an improvement in the aggregate level of productivity, since the aggregate level of productivity is a weighted average of the productivity of each unit. Usually, the weight is the market share but, in some cases, the employment share is used as weight. The entry and exit of firms from the market also affect the growth of the aggregate level of productivity. If entrants are more productive than incumbents, there is an improvement in productivity. Similarly, if exiting firms are less productive than continuing ones, the exit of those firms implies an increase in the aggregate level of productivity. Therefore, to address the effect of a CDP at the aggregate level, it is necessary to evaluate its effect on the productivity and growth—measured by sales or employment—of each firm and on the firm's decision to enter or exit the market. There are several variables that are related to productivity, such as the value of exports, R&D expenditures, technology adoption, and innovation. In some cases, a CDP aims at improving these variables.

2.2.2. Export-related Indicators

The main link between international trade and productivity is the increase in competition caused by trade liberalization (see, for example, Pavcnik [2002], who researched the effect of trade liberalization on the productivity of Chilean plants). Van Biesenbroek (2005) also undertook a study on trade liberalization to see whether export helped African businesses to achieve higher

[11] If the production function is Cobb-Douglas, this expression can also be used to obtain level of TFP.
[12] In some cases, cost shares are averaged by industry or across time.

productivity levels. He considered other avenues that could lead to an increase in the productivity of the companies engaged in export activities; in particular, he discovered that economies of scale are an important conduit for expanding productivity, but credit constraint and contract enforcement problems prevented firms that produced for the local market from fully exploiting opportunities in economies of scale.

With regard to policy evaluation, Das et al. (2007) have developed a dynamic model of export supply, which embodies plant-level heterogeneity in export profits, uncertainty about the determinants of future profits, and market entry costs for new exporters. They applied the model to factory-level panel data relating to three Colombian manufacturing industries, and found that export revenue subsidies are more effective at stimulating exports than the policies that subsidize entry costs. More recently, Melitz and Ottaviano (2008), based on Melitz (2003), developed a monopolistically competitive model of trade with heterogeneous firms that is well suited to analyze trade policies.

There are several indicators that can be used to measure the effect of a CDP on exports, such as the value of exports, probability of becoming an exporter, number of goods exported, and number of export markets. Some evaluations of productive development policies have been made, using some of these indicators in Latin America. Adopting a firm-level dataset for Peru from 2001 to 2005, Volpe and Carballo (2008) found that export promotion activities can be associated with increased exports, both in terms of market and product. In Argentina, using firm-level data from 1996 to 2008, Castillo et al. (2013a) showed that participation in the PRE small- and medium-sized enterprise (SME) support program implemented in this country increased the firms' probability of becoming an exporter.

2.2.3. Innovation-related Indicators

Following Crepon et al. (1998), a distinction can be made between innovation-input (short-term) indicators, innovation-output (medium-term) indicators, and economic-performance (long-term) indicators. Innovation-input indicators are those more directly affected by intervention. As an example, in the case of an R&D subsidies program for small firms, an innovation-input indicator relates to total expenditures in R&D. Innovation outputs are variables, where a realization of an innovation can be observed, such as patents. In some cases, a firm is directly asked whether it has significantly changed its products, whether a new one has been introduced, or whether there has been a change in the production process.

There have been previous studies evaluating innovation and technology policies in Latin America, including Hall and Maffioli (2008), which summarized the appraisals of government Technology Development Funds (TDF) in Argentina, Brazil, Chile, and Panama. These appraisals were done at the firm level, using data from innovation and industrial surveys and the administrative records of the grantees. They were done in tandem with quasi-experimental econometric techniques to minimize the effects of any selection bias. TDF effectiveness depends on the financing mechanism used, presence of non-financial constraints, company-university interaction, and characteristics of the target beneficiaries. Four levels of potential impact were considered: R&D input additionality, behavioral additionality, increases in innovative output, and improvement in performance. The evidence suggests that TDF do not crowd out private investment, and that they positively affect R&D intensity. In addition, participation in TDF induces a more proactive attitude of beneficiary firms towards innovation activities. However, the analysis finds little statistically significant impact on patents or new product sales. Other studies for Latin American countries include Crespi et al. (2001); Chudnovsky et al. (2006); Benavente et al. (2007); Crespi et al. (2010); Alvarez et al. (2012); and Castillo et al. (2013b).

2.2.4. Employment-related Indicators

A fourth set of indicators is that related to employment. These indicators may include the number of employees, type of employees by level of qualification, or the wage employees are paid by the firms. For example, Castillo et al. (2013a) found that the PRE SME support program in Argentina increased both employment and wages.

3. Data

There is one principle that relates to *any* empirical analysis, which is the GIGO principle; simply put, Garbage In, Garbage Out. No matter how sophisticated the analysis, if the data is no good, the result cannot be good. This section provides the guidelines relating to the type of data that is required for the SNA application and the evaluation of a CDP.

One of the challenges met during the evaluation of CDPs in the Latin American and Caribbean region was that secondary data (collected for purposes other than for the appraisal of a particular program) were not available. Although meaningful surveys and statistics existed from which data could have been extrapolated for the monitoring and evaluation of cluster programs, they were not always readily available. When census information about the firms in a given context is not available, identifying a sample for primary data collection also becomes a challenge.

The section below will be divided into two parts, one relating to the advantages and disadvantages of secondary data and the other to providing guidelines for collecting primary data. In the latter case, separate discussions will be included relating to sampling for SNA analysis and sampling for impact evaluation.

3.1. Secondary Data

There are three sources of secondary data: surveys, censuses, and administrative records. Each one of these sources has advantages and disadvantages that need to be evaluated. Surveys have the advantage of being able to construct a panel of firms with annual information, which is essential when applying quasi-experimental techniques that allow control for selection in non-observables. Surveys also provide information on several variables that allows the evaluator to use matching techniques to find non-beneficiaries with similar characteristics to the beneficiaries. As mentioned above, this is important because the more information the dataset has, the more similar the firms in the control group will be.

The main limitation of surveys, however, is that they collect information on a sample of the population and, in many cases they include only a small proportion of beneficiaries. This limitation is particularly meaningful when appraising CDPs, given that it may be likely that several beneficiaries will not be included in the survey. In addition, given that CDPs are at the industry and regional level, and surveys usually represent the national level, the data on only a small proportion of firms within a particular industry and region may be collected. Another possible drawback in the use of surveys is that most of them cover the manufacturing sector and only a few include other sectors.

Censuses, on the other hand, collect information about the universe of firms in a country. If beneficiaries happen to be active during a census, they will be included in it. Furthermore, a census usually contains more information than in a survey, which is essential when applying the propensity score-matching model. The drawback, however, is that they may take place every 10 years, or various years, before or after policy implementation, rather than every year. This can be one of many factors that could affect an appraisal of a program's impact. Even if the difference-in-difference estimate is applied, the length of time from one observation to another would make it difficult to prove that the program impact can, indeed, be identified. To add to this, while businesses should have remained active within the10-year period between the census taking, it is likely that smaller firms, related to CPDs, have been in business for less. Box 4 describes some of the datasets that have been used in the evaluation of productive development policies (PDP). These include Brazil's annual industrial survey (Pesquisa Industrial Annual, or PIA), (Encuesta Industrial Annual, EIA) in Chile, Colombia's manufacturing survey (Encuesta Annual Manufacturera, or EAM), and the innovation surveys of Argentina, Chile, and Panama.

The third type of secondary data comes from administrative records. There is a wide range of institutions collecting useful information about firms for different purposes. Compared to surveys and censuses, the main advantage of the administrative datasets is that, in most cases, they provide annual information for an entire group of firms. The main drawback, however, is that they provide little information about the firms, themselves. As with censuses and surveys, they can only be accessed within the relevant institution under a confidentiality agreement. Datasets obtained in previous evaluations of productive development policies are the Base for the Study of Employment Dynamics (Base para el Estudio de la Dinámica del Empleo, or BADE) in Argentina, the annual social information report (Relaçao Anual de Informaçoes Sociais, or RAIS) in Brazil, and the Internal Revenue Service (Servicio de Impuestos Internos, or SII) in Chile (see, Box 5).

Box 4: Surveys Used in the Evaluation of CDPs

Pesquisa Industrial Anual (PIA) is an annual survey sampling conducted by Brazil's census bureau (Instituto Brasileiro de Geografía e Estatística *[IBGE]*) that relates to formally established Brazilian mining and manufacturing firms and their plants. The sample of firms in *PIA* is drawn from two strata: a non-randomized sample of all Brazilian mining and manufacturing firms with a labor force of 30 or more workers and employees and a random sample of small to medium-sized firms with a labor force of five to 29 workers and employees. A firm is eligible to be sampled in *PIA* only if at least half of its revenues stem from manufacturing, and if it is formally registered as a taxpayer with the Brazilian tax authorities. The methodology was changed in 1996 and 2006. The major change was in 1996, and includes changes in both the sampling scheme and in the information collected from the sampled firms. The change in 2006 was not as important because it was limited to include firms that employ less than five workers in the sampling scheme. The 2004 *PIA* sample covers 42,371 firms among 155,656 eligible ones. *PIA* contains three main groups of variables: (a) information about longitudinal relations across firms; (b) balance sheet and income statement information; and (c) economic information beyond the balance sheet and income statement. The main longitudinal information in group (a) is the taxpayer firm registry number (CNPJ code). This allows for linking observations longitudinally, as well as combining them with other sources, such as the Annual Social Information tool (Relação Anual de Informações Sociais [RAIS]). Among other variables in group (a) are the ones that indicate the state of activity of a firm in a given year (such as whether it operates all year, only part of the year, or exits) and its structural changes (such as whether it emerges from a pre-existing firm or whether it creates a spin-off firm). Variables in group (b) include cost, revenue, and profit information, detailed in a manner similar to a typical Brazilian income statement. On the revenue side, for example, it is possible to exempt non-operational revenues, while on the cost side it is feasible to identify intermediate inputs, among other details. Variables in group (c) go beyond the income statement and include data, such as investment flows by type of asset and numbers of workers and employees. Employment is broken down into production and non-production workers.

EAM data, in Colombia, are collected and managed by the National Administrative Department of Statistics (Departamento Administrativo Nacional de Estadística [DANE]). EAM is a census of all manufacturing plants with 10 or more employees, or output exceeding COL$500 minimum in wages, and is available from 1977 to 2007. This survey contains information on plant characteristics and performance variables, such as location, sector of activity, legal organization status, plant size, sales, employment, investment, and expenditures. An important characteristic of this survey is that it is possible to construct firm-level price indices (see, for example, Eslava et al. [2004)]. This dataset has been used by Crespi et al. (2010) to evaluate innovation support programs and Arráiz et al. (2012) to evaluate a partial credit guarantee program. Other surveys that have been used in the evaluation of PDPs are the Innovation Surveys, collected in several countries in Latin America.

The datasets mentioned above do not include network data, given that they are very scarce. It is less common to find network data at the regional or cluster level in developing countries, although it is sometimes possible to locate information regarding firms through patent office databases (such as the United States Patent and Trademark Office [USPTO] or co-publication databases (such as the Thomas Reuters Web of Knowledge, formerly known as the ISI Web of Knowledge), as well as other company information and business intelligence databases (such as Bureau van Dijk). Most of these databases, however, capture only formal connections between firms and/or other organizations and are, therefore, not likely to showcase informal linkages, which are essential in a regional cluster's network structure. Other secondary sources of network linkages may exist at the local level for evaluators to explore.

Box 5: Administrative Records in Argentina, Brazil, and Chile That Have Been Used in the Impact Evaluation of Productive Development Policies

In Argentina, a dataset called BADE, constructed and managed by the Center for Employment and Business Dynamics (Observatorio de Empleo y Dinámica empresarial [OEDE]) at the Ministry of Labor, Employment, and Social Security, includes data from the administrative records of several public institutions, including the National Administration of Social Security (Administración Nacional de Seguro Social [ANSES]), Federal Administration of Taxes (Administración Federal de Ingresos Públicos [AFIP]), and National Customs Administration (Dirección General de Aduanas). BADE is a matched employer-employee longitudinal dataset that contains annual information for all the firms disclosing employment information in Argentina after 1996. It covers the primary, manufacturing, and service sectors, and has firm-level information about age, location, industry, employment, wages, and value of exports. In 2008, the dataset included around 6 million workers and 570,000 firms. This dataset was used by Castillo et al. (2013a) in the evaluation of the PRE SME support program.

A similar dataset in Brazil is RAIS. This dataset is maintained by the Brazilian Ministry of Employment and Labor (Ministério do Trabalho e Emprego [MTE]). All registered taxpaying establishments are requested to submit, annually, information relating to every single worker employed by the establishment at any time during a specific year. RAIS is a matched employer-employee longitudinal dataset. This dataset contains information relating to industry, region, number of employees, wages, workers' occupations, and education. This dataset was used by De Negri et al (2006) and by Maffioli et al. (2011).

In Chile, there is also a longitudinal dataset covering the population of tax-paying firms. This dataset is maintained by Servicio de Impuestos Internos (SII) and contains information about industry, region, number of employees, sales, and taxes paid by the firms. This dataset was used in Arráiz et al. (2011) for the evaluation of the Supplier Development Program, PDP.

3.2. Primary Data

In the absence of secondary data, primary data should be collected, which requires the use of a specially tailored questionnaire. This can prove expensive, and tends to cover only a short period, unless systematic data collections are made on the same sample or population over time. With firms, historical accounting data may be available, although it may not be very indicative, since some of the information may not have been recorded, nor do all firms have proper accounting systems. The latter is particularly relevant to micro and small firms.

Several decisions require consideration prior to collecting primary data, especially with regard to the sample design, which is one of the most important steps in every study. These are

(a) unit of analysis and means of identification; (b) representative sample and size of sample; and (c) data collection process.

3.2.1. Unit of Analysis and Means of Identification

The most likely unit of analysis (UA) to be used for the impact evaluation of a CDP is the firm itself.[13] While it may appear straightforward, the criteria for firm identification should be carefully defined. Although a company is usually a unique unit of production with its own identifier (such as a fiscal code number), it may manage various production units (plants) that could have characteristics that are irrelevant to the study (for example, one of the plants may be outside of the cluster). On the other hand, the study may focus only on firms operating in certain industries, which will require criteria to identify the industries (the code relating to the International Standard Industrial Classification (ISIC) of All Economic Activities (Clasificación Internacional Industrial Uniforme, or CIIU) and a method to correlate a firm to a specific industry.

3.2.2. Representative Sample and Size of Sample

Designing a sampling strategy can be relatively complex, and any error at this stage could invalidate the entire study. It is essential, therefore, to analyze the following:

- *Sampling frame from which the sample will be drawn*. This is highly important, since the sampling frame will determine the generalizability (external validity) of the evaluation results.
- *Sample size*. Other things being equal, the lower the sample size, the less likelihood that the assessment will detect any impact, whether it exists or not. This is known as statistical power. In general, evaluations that have high statistical power are preferred.
- *Distribution of the sample between treated and control groups.* In terms of statistical power, the more balanced the groups are, the better it is for the evaluation exercise.
- *Levels at which the intervention is being implemented*. In technical terms, these are known as "clusters."

[13] In some cases, studies can be conducted at a more aggregate level, such as a municipality or other administrative unit.

(i) *Sampling based on power calculation for detection of minimal effects*

It is recommended that the sample used be designed by a specialized government agency, such as a country's national statistical office, which normally has access to census data. Such agencies also possess first- and second-hand information that could be useful for sample calculations. Alternatively, a public or private institution, such as the Chamber of Commerce or a business association, would have sufficiently reliable information to enable the identification of the population of firms within the cluster. Failing this, particularly in remote rural areas, the most accurate information possible should be sought from sources, such as the head of a village or a prominent entrepreneur. However, clusters are, by definition, geographically bounded areas, which makes it easier to identify the actors that belong to it.

Once the population of companies is known, the sampling should be carefully done in order to guarantee sufficient statistical power to determine a program's impact and to analyze the heterogeneity of impact among producers. Assessing the heterogeneity of impact is common in a CDP. Different programs may have varying effects on particular types of producers, so it may be in the best interest of the project team to identify and measure these heterogeneous outcomes separately. To accomplish this, stratifying the sample to assure representativeness within different groups in the population (strata) is advisable. For instance, it may be prudent to obtain estimates of the impact by region or by firm size. Stratification assures that the assumptions for every one of the subgroups are valid.

Statisticians tend to rely on the power formula to ensure that there is sufficient statistical power to identify program impact. The power formula, which depends on the sampling design, usually is composed of four pieces of information: (i) expected effect on outcome variable, (ii) outcome variable's standard deviation, (iii) confidence level (usually 95 percent), and (iv) statistical power.

To reduce sample dispersion and, therefore, administrative costs, it is common practice to obtain the sample in two stages. In the first, the principal sampling units (PSU) are randomly selected, and in the second, producers within the already selected PSU are randomly chosen. The PSUs are already defined geographic areas. However, employing this two-stage process involves correction for intra-cluster correlation, which increases the sample size.[14] Intra-cluster correlation correction is required, as the production units within the same cluster may be similar.

Finally, given that most of the data collection procedures for impact evaluations involve collecting a baseline and a follow-up survey for the same producers, a non-response/attrition adjustment may be needed. The attrition rate refers to the number of production units that cannot be interviewed in the follow-up survey because these are not found by the interviewers. The non-response and attrition correction varies among countries and regions. For example, a greater attrition correction could be adequate in a country that has a higher level of migration.

(ii) *Sampling for SNA analysis*

If SNA analysis is included in the study, the sampling strategy has some caveats. The collection of SNA data does not always follow the same sampling strategies as those used for general data collections (see Point (i) above). In fact, if the number of firms is limited and the evaluation budget allows for it, an attempt to interview all the actors of the cluster is recommended (Table 3 (i)). The ability to collect data on the whole population of cluster actors is ideal for undertaking both SNA and econometric analyses for an CDP impact study. The SNA tool will enable the collection of "full network" data, and the analysis of social structures, and measuring of all structural concepts of network analysis discussed earlier (Wasserman and Faust, 1994; Hanneman and Riddle, 2005). The collection of full network data is normally carried out by interviewing all cluster actors through a "roster recall" method. Whereas interviewing the whole population of cluster actors is preferable in order to carry out SNA analysis, it is not essential to an econometric analysis of impact evaluation. However, having population data does not hinder but, rather, improves the effectiveness of the econometric analysis.

A second strategy can be applied, when accessing the overall cluster population is not feasible and the only option remaining is to draw a population sample. This can be due to either the scope of the cluster population being too large or because firms are unwilling to collaborate in the survey (Table 3 (ii)). Sampling in SNA analysis does not generally follow the same rules applied to standard statistics. This means that, if the objective is to analyze the structure of networks into a cluster, random sampling or other standard sampling approaches (e.g., stratified sampling) are not generally endorsed. In fact, failure to include a central actor may generate a

[14] Note that "cluster," in this instance, is used in statistical terms.

bias into the measuring of the network. Likewise, non-respondents may severely distort data and introduce a bias, as a network map may be very misleading if the central actor is not included (Borgatti and Molina, 2003). One way to resolve this is to collect full network data about only a workable number of firms, which correspond to a sub-set of the cluster population, selected on the basis of a specific criterion. For example, it could be restricting the analysis only to firms occupying specific stages of the value chain (e.g., only assembly firms), or working in selected industries (in the case of several industries being present in the same cluster), or having other characteristics that can be distinguished *ex-ante*. While this is only a second-best approach for SNA analysis, it should be taken cautiously when applied to econometric impact evaluation studies, as it can introduce a selection bias that needs to be treated carefully. In particular, this approach implies that inferences can only be made about the sample of the population that has been selected and not about the overall cluster population (Wasserman and Faust, 1994) and, hence, their choice should be made considering the CDP evaluation objectives.

If this latter approach is not compatible with the objective of the impact study, it is possible to opt for a third way, in which a sample of focal actors (individuals or organizations, called "ego") are selected, based on the sampling approach discussed earlier (Point (i)), and who are interviewed about their direct linkages (Table 3 (iii)). This approach yields considerably less information about network structure because it is unable to map the full network, but it is often less costly and more practical (Hanneman and Riddle, 2005). The collection of data, in this case, includes interviewing the sampled focal actors about their linkages, often opting for a free recall method (Items 9.2. and 9.3. in Box 9 in Section 3.4.2). If possible, it is recommended to interview the focal actors' direct contacts to obtain full information about the structure of the **ego-networks** (Item 9.4. in Box 9 in Section 3.4.2). Ego-network data are less powerful than "full network" data, but can still be used to calculate different indicators of centrality and actor-level position in the network, such as degree of centrality, closure, and structural holes (Table 1). Moreover, if the sample has been drawn randomly, it is possible to explore some structural properties of the network by analyzing degree centrality distributions. However, many structural measures of the overall network (e.g., core-periphery) cannot be observed, if the network data has not been fully collected.

Table 3: SNA Sampling Strategies

Sampling Options	SNA	Econometric Impact Evaluation	Drawbacks
(i) Full population of cluster actors	Yes	Yes (provided that information for a control group is available).	High costs. Unfeasible for highly populated clusters or low-budget evaluations.
(ii) Selection of cluster sub-population (not random or stratified)	Yes, limited to the sub-population.	Yes, limited to the sub-population (provided that information for a control group is available).	Analysis and inference is possible only on a portion of the cluster population, not on the cluster as a whole.
(iii) Standard sampling techniques and ego-network data collection	Yes, limited to ego-centered analysis.	Yes (provided that information for a control group is available).	It is not possible to measure the structural properties of the cluster network, but only to analyze the size and characteristics of the network around each sampled actor.

3.2.3. Data Collection Process

With regard to the data collection process, the following applies:

- *Information that is collected and how it will be measured*. The design of a survey for a program's impact evaluation should be based on the theory of change. Therefore, it should be able to generate not only the information needed for computation of the outcome indicators (final and intermediate), but also other control variables essential to improve the attribution to the effects on the program. In addition to this, the collection of cost data should not be neglected, which should be brought about by the monitoring of the systems in place, and which will enable *ex-post* cost-benefit or cost-effectiveness analysis.
- *Number of waves of data collection*. For the typical impact evaluation, at least two waves of data collection are necessary for both the treatment and control groups: one at baseline level (prior to intervention) and one as a follow up (subsequent to intervention). When to do the follow-up data collection depends upon several factors, including the period of time it will take before an impact can be observed. When there is sufficient data for two or more rounds, it is time for the analysis to take place. The more follow-up analyses, the easier it will be to measure the medium- and long-term impact of a program. Second, the more sequences of follow-up analyses, the more the statistical power of the sample and the likelihood of detecting the minor effects will be.
- *Timing of data collection*. For most impact evaluations, it is crucial to select the most appropriate time to collect the data. For example, in the case of an impact evaluation that

aims to identify the effect of a program on agricultural productivity, it would be recommended to conduct the survey immediately after the main harvest, in order to reduce measurement error and avoid a recall period. Appropriate timing depends on the specific substance for which the data is collected. Since CDPs require time to reflect any effects (often years), it may be necessary to let some time pass before carrying out an impact evaluation.

Box 6: Five Steps and Ten Tips for Data Collection

Gertler et al. (2011) provide a list of five **steps** that help to organize the data collection process:

1. Engage the help of an agency that specializes in data collection; that is, the institution that implements the program, another government institution, or an independent firm. It is important to ensure that no contamination or bias will be introduced by the agency (e.g., in the event that it is the agency that conducts the program) and, if baseline data is collected, that no program operations have taken place prior to collection.
2. Commission the development of a questionnaire. It is necessary to collect, selectively and consistently, all information required to respond to policy questions, including the intermediate and final impact indicators, control variables, exogenous factors, and other characteristics of the intervention.
3. The agency or firm, responsible for collecting data, will recruit and train field staff, and will pilot the questionnaire (for comprehension and accuracy) before it is finalized, making any necessary adjustments.
4. The agency or firm will conduct the fieldwork, supervise the team (including technical staff, supervisors, field coordinators, support staff, and data entry personnel), and provide training for the data collection team. It will provide a manual of practices and procedures. It is crucial, at this stage, to attend to any non-sampling errors there may be—in particular, Nonresponse and Measurement Errors—so as to avoid any bias in the evaluation.
5. Compute, process, and validate the data collected.

Stecklov and Weinreb (2010) provide a list of **10 tips** that can be taken to reduce measurement error:

1. Follow basic administrative guidelines.
2. Confirm the regional and national "key players," and ensure that consideration is given to ways in which to work with them, so as to reduce the chance of "spoilers."
3. Conduct a pilot test of all questionnaires.
4. Engage a relatively large number of potential interviewers for training and set high standards for reward, including conducting an examination at the end of the training course.
5. Interviewers should be assigned using interpenetrating sampling techniques.
6. Consider all potential errors of non-observation, including sampling, coverage, and non-response.
7. Include questions that can allow *ex-post* identification of different types of errors of measurement.
8. Carefully evaluate whether there are systematic non-response patterns that might affect the interpretation of findings.
9. Design clear guidelines for filling in missing data, preferably using interviewer teams done shortly after each day of data collection.
10. Attempt to compare results to those that might be obtainable from routine statistics or other data sources.

3.3. Possible Difficulties and Advice when Collecting SNA Data

As with any data collection, it is important to increase the response rate. This means that when a CDP is designed, it already includes the program beneficiaries for the evaluation process. Therefore, the evaluation should be prepared in tandem with the policy design so as to ensure that the actors participating in, or benefitting from, the CDP will also participate in the evaluation process by responding to the survey, both before, during, and after CDP implementation.

Evaluators should be aware that obtaining full network data (Table 3i) may prove to be a challenge in terms of non-responsiveness, in which case it is recommended that the policy design include the SNA process for an impact assessment. This will ensure that CDP beneficiaries are well informed, prior to the start of the program, about the data requirements. It is important that as many beneficiaries participate in the survey as possible, excluding the non-beneficiaries.

Ethical concerns could arise relating to the collection of network data. Contrary to conventional methodological approaches, anonymity at the data collection stage relating to CDPs is not possible. The respondents are required to report and provide the names of those within their network (Borgatti and Molina, 2003), some of who may not wish to be named, particularly in terms of sensitive relationships (e.g., bribes for business licenses; transfer of sensitive information about a firm; proposed strategic alliances). These ethical issues should be borne in mind. To overcome them, respondents need to be assured that all relationship information will be maintained in confidence, and that network maps will not include the names of the actors, unless previously agreed with the respondents. In parallel, relationship questions should be formed so as to ensure that responses do not expose sensitive or highly strategic information. Borgatti and Molina (2005) provide an example each of a disclosure agreement and a consent form, which can be tailored to CDP evaluation surveys to avoid any misunderstandings relating to the collection and use of network data.

3.4. Other Issues Related to Data Collection

3.4.1. Pilot Testing and Survey Preparation

Before administering a baseline and follow-up survey, the norm is that the data collection strategy should include a pilot survey, so as to ensure the validity of the questionnaire. This can expose questions or words that could be misinterpreted or misread, and will confirm the relevance of the questionnaire. Questionnaires often include such tools as checklists for cross-referencing for quality assurance. The quality assurance can be confirmed at the time when the questions are checked.

The number of pilots depends significantly on the resources available to finance the data collection. Nevertheless, at least two pilot surveys should be undertaken prior to implementation of the baseline and before every follow-up survey. The size of the pilot may vary from project to project, depending on the sample size of the survey, which, in turn, depends on the survey stratification. The test also serves to identify the logistics required to carry out the survey. Since it will mirror the data collection process in the field, it can indicate how best to organize both the data collection and the data input. In general, the sooner the data input follows data collection, any errors identified in the output of data can be resolved in the field. Conducting a survey generally includes a team of approximately five people; one is assigned as a task leader, responsible for the input of data in the field, while the rest are analysts. Any errors that may be found in the data output can be addressed immediately by sending the analysts back to the field.

Every group of interviewers should be familiar with the area(s) in which their questionnaires will be circulated, as well as which routes to take. Each team member should possess a data collection schedule, outlining the process in its entirety, as well as for that pertaining to each group. Prior to any type of fieldwork, it is essential that the interviewers, as well as the supervisors, have received adequate training, either together during the pilot stage or separately. Training is essential so that the interviewers become familiar with the survey process, and it should include theoretical, as well as practical, exercises in the field with other interviewers and firm owners. The training period depends on the length and difficulty of the questionnaire, and should include supervision in the field, as well as during the administration process. Finally, the procedure for data collection should include a timeline of activities for all those involved in the process. Box 7 provides an example of an impact evaluation timeline.

Box 7: Impact Evaluation Timeline: An Example

The timeline illustrated below describes the main activities, the actors involved, and the dates on which the activities will be conducted for the Farmers' Direct Support Program (Programa de Apoyos Directos al Campo (PROCAMPO)) impact evaluation in Mexico. Overall, three main actors are responsible for relevant activities are the IDB (through its Strategy Development, ,Rural Development, and Environment and Natural Disasters divisions), a technical assistance (TA) firm, responsible for the field work, and the National Statistics and Geography Institute of Mexico (Instituto National de Estadística y Geografía (INEGI)) that will provide the sampling methodology, based on the 2007 Agricultural Census.

Activities	Responsibility	Dates
Baseline administration		
Questionnaire design	IDB	Nov. 2009–Apr. 2010
Pilot (1st round)	TA Firm	Sept.–Oct. 2010
Pilot (2nd round)	TA Firm	Oct.–Nov. 2010
Sampling design and methodology	INEGI	June–Aug. 2010
Training for interviewers and supervisors; preparation of materials for survey administration.	TA Firm	Nov.–Jan. 2011
Baseline administration	TA Firm	Feb.–Mar. 2011
Data input	TA Firm	Apr.–May 2011
Baseline data	TA Firm	May 2011
Baseline analysis	IDB	June–Sept.2011
Follow-up Survey		
Questionnaire design	IDB	Nov. 2011–Apr. 2012
Pilot (1st round)	TA Firm	June 2012–July 2012
Pilot (2nd round)	TA Firm	Aug.–Sept..2012
Training for interviewers and supervisors; preparation of materials for survey administration.	TA Firm	Oct. 2012–Jan. 2013
Follow-up survey administration	TA Firm	Feb–March 2013
Data Entry	Consultant Firm	April–May 2013
Follow-up Data	Consultant Firm	May 2013
Impact Evaluation Analysis	IDB	June–Sept. 2013

3.4.2. The Questionnaire

In order to compile information that is relevant to the analysis, we recommend a properly structured questionnaire that includes information on, at least, the following:

1. Respondent firm's (or other actor's) key characteristics (location, year established, mergers and acquisitions [if any]; size, sector, main local activities, etc., examples of which are illustrated in Box 8).
2. Respondent firm's (or other actor's) participation in the CDP (date it became part of the initiative, types of new activities undertaken, perceived benefits from participation, year in which participation is estimated to end).
3. Respondent firm's (or other actor's) key performance indicators.
4. Formation of, and participation in, local networks. The nature and characteristics of networks and their linkages will depend on the local context and will, thus, have to be determined after an initial evaluation, and then validated in the field during the pilot.

Box 8: Examples of Relevant Questions

1. Identification number.
2. Principal product.
3. Size (mall, medium, or large)
4. Year operations began.
5. Total annual sales in any given year.
6. Percentage of sales in any given year (national sales, indirect exports, direct exports).
7. Total annual cost of raw materials, materials inputs, or supplies in any given year.
8. Introduction (if any) of new or significantly improved products or processes for producing or supplying merchandise or services.
9. Lines of credit or loans from financial institutions.
10. Number of permanent, full-time employees/workers at the end of any given year.
11. Total annual cost of labor, including wages, salaries, bonuses, and social security in any given year.
12. Net book value of machinery, vehicles, and equipment in any given year.

The questionnaire should include a section relating specifically to network data, with questions related to actors' relationships with other actors and requesting specific names. This contrasts with other approaches, where questions relate to the relationship between the respondent and various categories or groups of actors (e.g., suppliers, clients, universities). Depending on the sampling strategy (Table 3), it is possible to opt for different relational data collection methods.

In the case of a sampling design for interviewing an entire population of cluster actors (Table 3i), the “roster recall” method is suggested, where the questionnaire includes a complete list (roster) of the cluster actors. It also should seek the existence or importance of a given type of relationship between the interviewee (focal actor) and the rest of actors listed in the roster. The questions can be formed for collecting either dichotomous relational data (existence/non-existence of a link) or valued data, which weight relationships on the basis of their importance, frequency, size, value, etcetera. Relationships can also be mutual by design (e.g., a question about the geographical distance between two actors) or un-directional (e.g., relationships where an asset is transferred from actor *i* to actor *j*). Item 9.1 in Box 9 illustrates examples of questions, based on the roster recall method.

The collection of these types of relational data can be undertaken face to face or through other means, such as online questionnaires, depending on the nature and characteristics of the interviewees. For rural farmers, face-to-face interviews are a more appropriate tool, while ICT managers may be willing and sufficiently capable to respond to an online survey. The roster recall method is the most efficient and recommended way to collect network data relating to an entire population of cluster firms. It minimizes the risk of data loss due to lapses of memory by a respondent, given that each interviewee is supplied with a complete list of other actors within the cluster to consult with prior to tackling the relationship question(s). The same method can be applied when the sample is represented by a sub-population of cluster actors (Table 3ii), with the caveat that, in this case, the roster includes only the actors of the sub-population.

When these sampling options are not available, and the focal firms are selected using ordinary statistics sampling techniques (Table 3 (iii)), respondents should be questioned about their relationships with the rest of the cluster population. If these are too numerous to be included in a roster, or are unknown prior to data collection, the “free recall” method is recommended. The respondent will be free to name cluster actors with this method, without drawing from a predetermined list, and the questions can be developed within a “free choice” or “fixed choice” design. In the free choice design, respondents have no constraint on how many names to provide, which has the advantage of including all links, regardless of relevance, although it runs the risk of producing incorrect responses (Item 9.2 in Box 9). In the fixed choice design (Item 9.3. in Box 9), respondents are requested to provide a maximum number of names to ensure they will produce the most important actors in the network, instead of the less relevant ones. To improve the accuracy of ego-centered networks in both cases, it is suggested that relationship data be compiled to determine whether the actors linked to the ego-network are connected to each other. One way to do this is to first interview the ego, and then use a snowball approach to interview the alters about their relationships. Alternatively, the ego can be questioned in such a way as to report the existence of linkages among the alters (Item 9.4. in Box 9).

Box 9. Examples of Relational Questions

9.1. Roster-Recall

Seeking or giving technical advice (TA)

If TA is urgently required, which local companies included in the roster do you select for their expertise? [Please rate the importance attached to the knowledge linkage, established with each of the firms, according to its expertise and quality, on the basis of the following scale: 0=none; 1=low; 2=medium; 3=high.]

Which of the firms in the roster do you think have benefited from the technical support provided by this company?[Please rate the importance attached to the knowledge linkage, established with each of the firms according to its expertise and quality, on the basis of the following scale: 0=none; 1=low; 2=medium; 3=high.]

Collaboration in marketing

Indicate the firms in the roster with which this company has collaborated in developing a joint marketing initiative within the past two years. [Please rate the importance attached to the collaboration, according to the impact it had on your business activities, based on the following scale: 0=none; 1=low; 2= medium; 3=high.]

Collaboration as a result of the CDP

Indicate the firms in the roster with which this company has collaborated in developing a new product for the domestic market as a direct consequence of the policy initiative [specify which]. [Please rate the frequency of collaboration, according to the following scale: 0=none; 1= a few times a year; 2=monthly; 3=weekly.]

9.2. Free-Recall (Fixed Choice)

Contribution to product upgrading

Name up to five local farmers whom you consider to have contributed to the upgrade in the quality of your crops.

University-industry linkages

Indicate the name of up to ten university researchers with which you have interacted on at least one of the activities, listed below, within the past five years:

- Joint research agreements (research undertaken by both parties)
- Consultancy work (commissioned by industry; not involving original research)
- Informal contacts (TA not based on a market transaction)
- Attendance at conferences with industry and university participation
- Other (specify______________________).

Innovative links as a consequence of the CDP

Name up to five public research institutions with which this firm has established innovative links as a consequence of the CDP [normally a precise definition of "innovative link" is provided.].

9.3 Free-Recall (Free Choice)

Whom you trust (free choice)

Name the entrepreneurs operating within the local cluster that you trust most [normally a precise definition of "trust" is provided].

Whom you trust (name generators)

Name the local entrepreneurs to whom you would lend a piece of machinery for free.
Name the local entrepreneurs to whom you would lend money.
Name the local entrepreneurs whom you would inform first about a new market opportunity.

Whom you trust more after the CDP (free choice)
Name the entrepreneurs operating within the local cluster, whom you trust more as a consequence of the cluster policy initiative [specify which] + [normally a precise definition of "trust" is provided].

9.4. Ego-Network
Question to the 'ego'
Name the firms within the cluster with which this company has collaborated for the solution of local environmental problems within the past three years.

Question to the 'alters' (only about ego's alters – included in a roster)
Indicate the firms in the roster with which this company has collaborated for the solution of local environmental problems within the past three years.

Box 10. SNA Methodologies Applied to CDP Impact Evaluation: The Córdoba Electronics Industry in Argentina

A CDP in the electronics industry of Córdoba was undertaken between 2003 and 2007, and was part of a wider set of CDPs, called Productivity Integration Programs (Programas de Integración Productiva, or PIP), co-funded by the Multilateral Investment Fund (Fondo Multilateral de Inversiones, or FOMIN), the IDB, and local resources. Some of the key objectives of the Córdoba CDP were to strengthen local linkages and cooperation among the private actors and between private actors and local institutions; ease local firms' access to new production technology and organizational innovations; and promote access to new markets.

SNA analysis was used for determining (1) the evolution of the inter-organizational networks within the cluster; (2) the benefits of the CDP, with particular reference to the changes that took place within the inter-organizational networks; and (3) whether these changes positively impacted the performance of firms. The study relied on two waves of data, collected in 2005 and 2012, respectively, drawn from direct interviews with local entrepreneurs and managers. A focus group was also part of the evaluation study.

The study analyzed the characteristics of two types of local inter-organizational networks: the information network (IN), formed to transfer business information, and the collaboration network (CN), based on projects with joint collaboration. The specifics of these networks were initially analyzed by means of descriptive SNA, which successfully measured the network structure and identified the central firms within the network, defined as the dominant players. These central enterprises appeared to be sufficiently important to ensure network connectivity and create the link between treated and untreated firms. Subsequently, by comparing the principal structural measures of the 2005 network and the 2012 network, it was discovered that they became, over time, less dense and more hierarchical, which was consistent with the fact that local firms tend to plan their linkages, learn how to "economize" on the number of relationships they can form, and select only partners from which they believe they can obtain tangible benefits. The study also found a relatively stable pattern of interaction over time, represented by the consolidation of a critical mass of firms—mainly the dominant players and their direct contacts—which are conclusive in maintaining the local inter-organizational network active.

Descriptive SNA analysis was also employed to demonstrate that the CDP had, indeed, led to the strengthening and creation of new technology-transfer ties between the electronics firms in Córdoba and other local/provincial or national institutions (among which are local universities). However, it showed no impact on the promotion of new linkages related to the promotion of export-oriented activities.

Finally, a SAOM analysis was applied on the network changes to identify what factors contributed to the formation of new links between 2005 and 2012. The investigation found that firms that had more intensively participated in some of the activities promoted by the CDP, were more likely to form new information links over the period 2005-12. To differentiate, the activities that promoted new information ties were those that were concrete, such as participating in the local community center for supplies and supplier development (Centro de Abastecimientos Comunitarios y Desarrollo de Proveedores (CACyDP) and strategic planning workshops. Those that were designed to promote networking (i.e., the Affinity Group Workshops and other institutional activities) failed to do so. Due to data limitations, the study was unable to demonstrate a robust relationship between the CDP, network participation, and firm-level performance. More details on this case are available in Giuliani and Matta (2013).

4. Determining Causality: The Attribution Problem

4.1. The Attribution Problem

To assess the impact of a CDP, a causal relationship between the policy treatment and the selected outcome variable(s) needs to be identified. The causal effect of a policy is the difference between the value of the outcome variable, after the policy has been applied, and the value of the outcome variable in the absence of the policy. Suppose the policy variable takes value one when firm *i* participates to CDP and zero otherwise – i.e. $C_i = \{0,1\}$. Suppose in addition that if firm *i* participates in a CDP ($C_i=1$), the value of the outcome variable would be Y_{i1}; if it does not participate ($Ci=0$), the value of the outcome variable is Y_{i0}. In technical terms, Y_{i1} and Y_{i0} are known as *potential outcomes*. Thus, the outcome variable of firm *i* can be written as:

$$Y_i = \begin{cases} Y_{i0} & \text{if } C_i = 0 \\ Y_{i1} & \text{if } C_i = 1 \end{cases}$$

or

$$Y_i = Y_{i0} + (Y_{i1}-Y_{i0})\, C_i.$$

The only difference between the two scenarios is that in one, the firm participates in the program, and in the other, it does not. Therefore, the difference in the outcome variables, $Y_{i1} - Y_{i0}$, can be *attributed* to the policy. The difference between the potential outcomes is the *causal effect* of policy *C* on the outcome variable *Y,* and the objective of an impact evaluation is to measure it.

It is important to note that only one of these potential outcomes is observed. If firm *i* participates in the CDP, Y_{i1} is observed and, if it does not, Y_{i0} is observed. Therefore, to estimate $Y_{i1} - Y_{i0}$ for participants, the counterfactual, Y_{i0}, needs to be estimated. This is the first and most difficult challenge to be addressed.[15] A second challenge that needs to be addressed during any evaluation is the fact that it is impossible to estimate the effect of *C* on *Y* for each individual *i* and therefore we estimate the average effect of the policy instead of the individual effect. The rest of the section aims at solving the first challenge; that is, the estimation of the counterfactual.

The first solution that usually comes to mind is to estimate the counterfactual by using the firms that do not participate in the program. However, is the information of the non-participants a good estimate of the counterfactual? Or, in other words, if participants had not

[15] This has been called The Fundamental Problem of Causal Inference (Holland 1986).

participated, would their outcome be the same as the outcome of non-participants? There are several reasons to think that the answer is negative. For example, imagine a situation in which we want to know if a CDP was effective in increasing firms' productivity by comparing the productivity of participants and non-participants after the program is applied. Even if we find that the productivity of participants is higher than the productivity of non-participants, that difference cannot be *attributed* to the program. In fact, if the participants had been already more productive than the non-participants prior to implementation of the program, the difference after the program would reflect the difference before the program and not only the effect of the program. Unfortunately, this is not an extreme example with little empirical relevance; it is generally the rule. Firms that decide to participate in a CDP are more likely to be motivated, entrepreneurial, and are better managed than those that do not participate. In the evaluation literature this is known as *selection bias*.

The second solution that usually comes to mind is to estimate the counterfactual using the information of participants prior to implementation of the program. This method is known as the *before-after* comparison.

Worth mentioning is that the before-after comparison may prove effective when appraising the first phase of a CDP. One can reasonably argue that the level of articulation before the design and implementation of a CDP could be considered an appropriate counterfactual of articulation at the end of a CDP. In this case, a simple comparison of the social network indicators that measure articulation before and after a CDP takes place can give use the average effect of the program:

$$ATET= E(Y_{it}|X_i) – E(Y_{it-1}|X_i)$$

where Y_i is a network indicator for agent *i,* and X_i are observable characteristics of the agents. Reliability depends on whether the other factors affecting changes in articulation levels can be considered negligible, or whether they can be attributed to the preparation phase of a CDP. Thus, in the short run, a before-after approach is likely to be appropriate for attribution purposes. If the before-after comparisons are applied after several years, it will be likely that other things may have changed sufficiently as to affect articulation.

A third challenge in the evaluation of CDP is the presence of *spillovers.* This challenge is particularly important in this type of policy because it aims at promoting spillovers. To include spillovers in the evaluation of CDP it is necessary to recognize that there are two types of beneficiaries and, therefore, two causal relationships of interest. On the one hand, direct beneficiaries are those firms that participate in the cluster program; that is, they accept to actively participate in the cluster development plan. On the other hand, indirect beneficiaries are those firms that do not participate in the program, but have links to participants—for example, they are located in the same municipality.

4.2. Randomized Experiment

Like in other type of policies, the ideal experiment that would answer the causal effect questions is a randomized assignment of the cluster policy.[16] [17] Even when it is not possible to randomize treatment, it is always recommendable (and feasible) to imagine which would be the experiment that would answer the questions of interest. In the case of CDP, the randomization has to be done at two different levels (double randomization) in order to select the locations where the cluster program should be applied and the firms in the selected locations that would receive the policy.[18]

Once the program has been randomly assigned, two comparisons can be done to evaluate the effect of the program. The first comparison provides the direct impact of the program by comparing direct beneficiaries (CDP firms) and similar non-beneficiaries:

$$DATE = E[Y_i|D_i=1,C_i=1] - E[Y_i|D_i=0,C_i=0]$$

[16] With a large enough number of observations, the randomized assignment of the policy ensures that beneficiaries and non-beneficiaries have statistically equivalent averages not only for their observed characteristics but also for their unobserved characteristics before the policy is applied.

[17] Alternatively, the effectiveness of CDPs can also be evaluated using a structural model. A structural model is a collection of stylized mathematical descriptions of behavior and environment that are combined to produce predictions about the way economic agents would behave in different scenarios. Structural models are constructed and used in the following way. First, the evaluator defines the equations describing the behavior of relevant agents—for example, consumers that select goods to maximize their utility for a given budget or firms that maximize profits using their production function. These equations are function of the policy parameters; for example, budget constraints are function of taxes and profits are function of subsidies that firms might receive. Second, these equations are combined to define the market equilibrium. Given that behavioral equations depend on policy parameters, the equilibrium also depends of policy parameters. Finally, the model is used to predict changes in the equilibrium derived from changes in policy. It is worth noting that in structural models the counterfactual is constructed from the underlying theoretical model, and thus can be used for ex-ante evaluation because it does not require post treatment data. These models can also be used to evaluate programs when counterfactual outcomes cannot be constructed empirically (Di Nardo and Lee 2010). It is important to mention that these models need to be validated. By construction they are good to make predictions in the scenario they were constructed. However, the interest is in doing predictions in different scenarios and therefore they need to be tested for predictions in other scenarios.

[18] In this toolkit, we focus in geographic spillovers because it is the most common source of spillovers in cluster programs. However, there are other sources of spillovers. Castillo et al. (2013b) presents the results of an impact evaluation of an innovation program in Argentina, using labor mobility as the source of spillovers.

where C_i takes value 1 if the firm participates within the cluster policy , or 0 otherwise; and *Di* takes value 1 if the location is a treated location, or 0 otherwise, and *DATE* is the Direct Average Treatment Effect (the difference between the average value of the outcome variable for direct beneficiaries and non-beneficiaries. The second comparison is between the indirect beneficiaries and similar non-beneficiaries, which will identify the Indirect Average Treatment Effect:

$$IATE = E[Y_i|D_i=1,C_i=0] - E[Y_i|D_i=0,C_i=0]$$

For the evaluation purpose, as described above, a random assignment of treated municipalities and the cluster policy (double randomization) is the best way to proceed. However, in practice, most cluster policies are not designed using a randomized control trial and, thus, non-experimental methods should be applied.

4.3. Non-Experimental Methods

The procedure to identify the impact of a CDP without random assignment of the policy is the same as in the case of random assignment; that is, to compare direct and indirect beneficiaries with non-beneficiaries. However, in this case, the simple comparison of averages gives biased results because of the selection problem mentioned above and econometric techniques are needed. The impact evaluation literature provides us with a set of methods that can be used to reduce and ideally eliminate the selection bias.

This section includes a brief discussion on each of the non-experimental econometric methods that can be used relating to the performance of firms in the impact evaluation of a CDP:

1. Propensity Score Matching
2. Difference-in-Differences
3. Instrumental Variables
4. Regression Discontinuity

4.3.1. Propensity Score Matching

As we mentioned above, the selection problem appears because beneficiaries (direct and indirect) are different than non-beneficiaries even before the policy is applied. The propensity score matching method defines a control group –non-beneficiaries similar to beneficiaries— using information on observable characteristics of beneficiaries and non-beneficiaries. The procedure involves two steps: first, estimating the probability of participating in the program—that is, estimating the "propensity score"—and second, comparing beneficiaries and non-beneficiaries with similar probability of participating in the program –finding non-beneficiaries with similar probability of participating in the program is "matching on the propensity score." The main idea of the method is that by matching firms using the propensity score, the observable characteristics used to estimate that probability will be balanced between beneficiaries and non-beneficiaries (Rosenbaum and Rubin, 1983).

The assumption necessary for identifying the effect of the policy using this method is that there is a set of covariates, observable to the evaluator, such that after controlling for these covariates, the potential outcomes are independent of the treatment status (receive or not receive the program). This assumption is known as Conditional Independence Assumption (CIA), Unconfoundedness or Selection-on-observables. It means that after controlling for the covariates (observable characteristics), for the evaluation, the treatment is as good as random.

The evaluator needs to take several decisions when applying the propensity score matching procedure. It is necessary to choose the probability model used to estimate the propensity score. This involves the model and the variables included in the model. Then it is necessary to define the matching algorithm. Next, depending on the matching algorithm, it is necessary to define other parameters. For example, in the case of nearest neighbor matching it is necessary to define if the procedure is with replacement or not and how many neighbors to consider. For a complete discussion on the technical decisions that have to be taken see Heinrich et al. (2010) and the references therein.

The principal advantage of this method is that it can be applied to a cross-section of firms, which is important because, in many cases, there is only information for one period after the program has been implemented. The main disadvantage, however, is that the CIA is too strong. It indicates that the evaluator observes all the information that drives participation in the program. However, this is not generally the case, and therefore it is a meaningful limitation. As mentioned previously, it is likely that only the best firms decide to participate in the program. If ability or the motivation of the entrepreneur is one of the determinants of firm's participation, it is not possible to control for self-selection using PSM, since the evaluator does not observe ability nor motivation. This is a problem that it is always present. Nevertheless, some researchers, who use PSM, argue that they can minimize the problem because they have a large set of observables that will allow them to control for the selection. As will become clear in the following section, however, most of the available datasets in Latin America contain few variables. As a result, there are several unobserved (for the evaluator) variables that can explain participation in a program. Therefore, this method cannot control for self-selection and provides biased estimates of the effect of CDP.

4.3.2. Difference-in-differences

As mentioned above, the conditional independence assumption is not met when there are unobservable factors driving the selection of firms into the cluster programs. However, if pretreatment data are available and the unobservable factors driving selection are time invariant, the conditional independence assumption can be relaxed. In this case, the effect of unobservables can be cancelled out by taking the difference in outcomes before and after the program. With two periods—pre and post treatment—the implementation is similar to matching except that outcome is measured in changes; that is, if t and t' are pre and post treatment and Yi is the original outcome, the new outcome variable is $\Delta Yi = Yit' - Yit$.

When more time periods are available the difference-in-difference (DD) estimator is the parameter δ in the following regression:

$$Y_{it} = \delta\, C_{it} + \gamma\, X_{it} + \mu_t + c_i + v_{it}$$

where Y_{it} is the outcome variable and C_{it} is a dummy variable that takes value 1 after the firm is a beneficiary of the program. Note that two equations need to be estimated: one for direct beneficiaries and the other for indirect beneficiaries. μ_t is measuring all the unobserved time-varying factors that affect all the firms in the same way (in practice, these factors are captured by a set of year dummies). c_i captures all the unobserved time-invariant firm-specific characteristics that affect the decision of participating in the program or affect the value of the outcome variables.

This method is better than PSM in controlling for selection. While PSM only control for the bias associated to observable characteristics, DD controls for the bias associated to observable and unobservable time invariant characteristics. The basic assumption of this method is that, in the absence of the program, the growth of the outcome variable of beneficiaries would be equal to the growth of the outcome variable of non-beneficiaries.

When there are large differences between beneficiaries (direct and indirect) and non-beneficiaries, it is difficult to assume that, without the program, the outcome variable of these firms would have the same growth. In this case, it is possible to apply PSM to find non-beneficiary firms, which, before the program was implemented, were similar to the beneficiaries.

When there are data for the years prior to implementation, it is possible to use PSM to find the non-beneficiaries with the same ex-ante trend in outcome variables than the beneficiaries. When comparing beneficiaries and non-beneficiaries that have the same characteristics prior to program implementation, including the trend in the outcome variables, it is easier to assume the equality of trends in the absence of the program. It is possible to argue that if, prior to policy application, beneficiaries and non-beneficiaries were growing at the same rate, in the absence of the program, they would also growth at the same rate.

The combination of PSM and DD involves three steps: (i) estimating the propensity score before the treatment takes place, (ii) defining a common support of firms through matching, and (iii) running a fixed-effect model on the common support. Heinrich et al. (2010) provide guidelines for the application of this method.

This method works particularly well when beneficiaries enter the program in the same year, or when it is possible to evaluate a cohort of beneficiary firms. When firms enter the program, sequentially, and their participation in the program depends on past outcomes, it is necessary to consider the inclusion of lagged dependent variables to control for the self-selection. (see Angrist and Pischke, 2009: Chapter 5).

4.3.3. Instrumental Variables

When a CDP includes some degree of self-selection, and there is a concern that unobservable differences between beneficiaries and non-beneficiaries might lead to biased estimates of the impact, this method is a powerful alternative to establish the impact of the program.

For example, say, a cluster program aims to increase the competitiveness of firms that self-select into a credit program for financing the implementation of a phytosanitary certification that would allow them to access international markets. In this case, it is expected that some characteristics—such as the ability and motivation of entrepreneurs—that determine firms' participation in the program might also affect their competitiveness. As is mentioned above, the comparison between beneficiaries and non-beneficiaries will include the impact of the program, as well as the intrinsic characteristics of participating firms. In this case, it is not correct to use a regression to estimate the causal effect of program participation (Ci) on the variable of interest (Yi) –i.e. $Yi = \beta\ Ci + ui$. This is because the unobserved characteristics, captured by the error term (ui), would be correlated with both competitiveness (Yi) and program participation (Ci). To solve this problem, it is necessary to have an instrument (Zi) or set of instruments. The instrument needs to satisfy two conditions: it has to predict program participation, $cov(Zi,Ci) \neq 0$; that is, it has to be relevant, and it cannot to be correlated with the outcome variable, except through program participation (i.e., it has to be exogenous), $Zi \perp ui$ or $cov(Zi,ui)=0$.

Although the method of instrumental variables (IV) is a powerful way in which to evaluate the impact of a CDP, finding an instrument after the program has been implemented is not always an easy task. One approach to ensure an instrument is available is to implement the program with a random encouragement design. In this case, some firms are randomly encouraged to participate through different mechanisms. For example, the program can randomly provide information to some firms to reduce the search cost of credit and, therefore, encourage firms to take the credit. In this scenario, it is reasonable to believe that firms that receive valuable information are more likely to participate in the program than those firms that do not. Given that incentives are randomly distributed across firms, there is no reason to believe that the encouragement mechanism is correlated to the outcome variable, making it a reasonable instrument.

If not embedded in the design of the program, it is quite difficult to identify an instrument that can guarantee both relevance and exogeneity *ex-post*. In the previous example, it is difficult to think of a variable that might affect the participation in the program, but does not affect competitiveness of the firms.

When an instrument is available, the IV approach is one of the best methods to identity the effect of a program. However, it also has some limitations. In particular, it can only estimate local average treatment effects (LATE), which means that its results are relevant only for those firms whose behavior is affected by the instrument (Imbens and Angrist, 1994). In the previous example, this implies that the results are valid only for the firms that participate in the program because they received the information about the program, and would not have participated without the information. In Impact Evaluation literature, these firms are known as *compliers*. The results are not valid for those firms that received the information and would have participated, even without the information (i.e., the results are not valid for *always-takers*). It is important to note that if the program was designed so that the treated and non-treated groups were representative of a certain population, the IV estimates might not have external validity for the whole population.[19]

4.3.4. Regression Discontinuity

Regression discontinuity (RD) is a powerful approach to identify the effect of a CDP on firm performance. The approach is based on the idea that in a highly rule-based environment, some rules are arbitrary and, therefore, provide for natural experiments. In this framework, LATE can be estimated at the discontinuity that determines which firms are assigned to treatment (receive the program) and to control (not receive the program). Intuitively, the treated units just above the cut-off point are very similar to the control units just below the cut-off point. We then compare outcomes for units just above and below the cut-off point. RD designs come in two forms: fuzzy and sharp. The sharp design can be seen as a selection-on-observables description, while the fuzzy design leads to an instrumental-variables-type setup (Angrist and Pischke, 2009: Chapter 6).

In the sharp RD, treatment is a deterministic and discontinuous function of a covariate; for example, the cluster policy can be applied for a particular administrative district. In this case, there are no untreated firms in the treated district and there are no treated firms in the untreated district. If a policy without spillovers is applied to firms in a particular administrative district, it is relatively simple to find good controls. As long as the administrative district is arbitrary, firms in the adjacent district could serve as controls because they are not treated, and are potentially equal to those treated. However, in the case of cluster programs, if firms in the adjacent district are part of the cluster, they cannot be used as controls because they are the indirect beneficiaries that receive the spillovers.

The fuzzy RD differs from the sharp discontinuity in that the covariate does not perfectly determine treatment and control but, rather, influences the probability of treatment. In this case, the covariate affecting participation in the program can be used as an IV to predict treatment, and the model can be estimated using two-stage least squares. Given that this type of regression discontinuity can be seen as a particular case of instrumental variables, its advantages and limitations are the same as the advantages and limitations of the IV approach.

[19] Also, we have to take into account the problem of weak instruments (see, for example, Bound et al. 1995). An instrument is weak when its correlation with the treatment is low. If the instrument is weak, it can generate bias and increase the standards errors of the IV estimation.

Summarizing, random assignment is less likely to occur in a CDP, and econometric techniques are needed to control the selection of firms into the program. Weighting pros and cons of each technique, DD (both as a combination of PSM and fixed-effect or the model with lagged outcome variable as additional control) appears to be a strong technique to evaluate the impact of the second stage of a CDP on direct and indirect beneficiaries. Although the IV and RD techniques are preferred, they are more demanding in terms of the design of the policy.

4.4. Considering Other Crucial Aspects of the Impact Evaluation Design

A well-designed evaluation provides additional information that is crucial for a complete understanding of the effects of the program, and enables the gathering of relevant policy recommendations that would help to reformulate the program or better design future ones. In this subsection, some aspects that should be always considered in the design of an impact evaluation are discussed.

4.4.1. Timing of the Effects

In general, it takes time for certain programs to demonstrate the desired effect on the efforts of firms and, in turn, their effect on economic performance. The materialization of concrete outcomes requires a period of gestation after a program has been implemented. The time lags may differ, according to the selected performance indicators. For instance, it may take more time before the program increase profits or productivity. More generally, the impact of different programs may display very different patterns over time. An intervention may generate a one-shot increase in the outcome, or may have a strong impact that fades out progressively over time; the impact of a program may only appear after a certain period, or may even generate an initial decrease in the outcome that is, later, increased in subsequent years.

As a result, appropriate consideration of the timing of the effects is crucial in an impact evaluation of a CDP, and failing to account for these issues may lead to misleading conclusions and policy recommendations. A clear distinction should be made between short-, medium- and long-term effects to properly evaluate the costs and benefits of a public program. For instance, considering only a short period of time after an intervention may result in an underestimation of the impact, if the effects take several years to appear. On the other hand, evaluations focusing only on later periods may conclude with an underestimation of the costs, if an adjustment process occurs within the first years.

4.4.2. Intensity of Treatment and Dosage Effects

While Impact Evaluation literature usually includes analyses of the binary case of participation against non-participation in a certain program, in practice, units may generally differ not only on their binary treatment status (participant versus non participant), but also on treatment intensity. For instance, firms may receive different amounts of public subsidy. This fact raises important issues to be considered, when designing an evaluation of a cluster program: the question is not only whether participants perform better than non-participants, but also how different intensities of treatment may affect performance and whether it is possible to find an "optimal level" for the intervention (e.g., the amount of financing that maximizes the effect on firm performance).

4.4.3. Multiple Treatments

In contexts where "multiple treatments" are available, the evaluator may be interested not only in the individual effects of each one, but also on potential interactions between them. In fact, it is not obvious that the effect of multiple cluster programs will be additive; instead, it may be the case that the combination of different interventions has multiplicative effects or, on the contrary, one treatment cancels out the effect of the other. Therefore, research into the joint effect of different types of interventions may be crucial for the design of the programs. Castillo et al. (2013a) provide an example of the evaluation of different types of innovation support on the employment and wages of firms.

4.4.4. Heterogeneity of Impact

In most relevant contexts, it might be hard to assume that CDPs have the same impact on all the beneficiaries. CDPs can have a differential effect for different groups of firms. For instance, they can have a stronger effect on less productive firms or on smaller firms. Restricting the analysis to the average impact for the whole population (or treated population) may give an incomplete or, at least, imprecise assessment of the impact of a program. Therefore, it is necessary to account for the possibility of impact heterogeneity (Heckman et al., 1997).

The analysis of the heterogeneity is important for modifying a program or improving the design of future programs. For instance, important lessons about the targeting of the program can be obtained through this analysis. If, for example, the program is not effective for some firms, it has to be modified (re-designed or excluded) for this group.

4.5. Understanding the Policy-Coordination-Performance Relationship: How to Integrate Social Network Analysis Methods with Econometrics

So far, this publication has focused on two questions: What is the impact of coordination on the program, and what is the impact of the program on firm performance? To answer these questions, two reduced form equations are proposed:

$$y_1 = \alpha_1 + \delta_1 P + u \qquad (1)$$

$$y_2 = \pi_0 + \pi_1 P + v. \qquad (2)$$

where y_1 are measures of coordination and y_2 are measures of firm performance. Given that coordination affects firm performance (even without a program in place), one could identify, separately, the effect of the program on coordination and the effect of the coordination and program on firm performance. In this case, the following system of equations needs to be estimated:

$$y_1 = \alpha_1 + \delta_1 P + u_1 \qquad (3a)$$

$$y_2 = \alpha_2 + \gamma y_1 + \delta_2 P + u_2. \qquad (3b)$$

Solving the system for y_2, the following reduced form for the performance measures occurs:

$$y_2 = (\alpha_2 + \gamma \alpha_1) + (\gamma \alpha_1 + \delta_2) P + (\gamma u_1 + u_2)$$
$$= \pi_0 + \pi_1 P + v. \qquad (4)$$

Note that this equation is equation (2). To identify the effects of interest, the values of π_1, γ and π_2 need to be known. The reduced forms (3a) and (4) provide estimates of α_1, δ_1, π_0, and π_1. To recover the information g π_2, there are two equations ($\pi_0 = \alpha_2 + \gamma \alpha_1$, $\pi_1 = \gamma \alpha_1 + \delta_2$) with three unknowns ($\alpha_2$, γ, δ_2). Therefore, the system is not identified.

If there is some variable that affects coordination and does not affect firm performance directly (i.e., that affects firm performance only through coordination), then the system is

$$y_1 = \alpha_1 + \delta_1 P + \beta x + u_1 \qquad (5a)$$

$$y_2 = \alpha_2 + \gamma y_1 + \delta_2 P + u_2 \qquad (5b)$$

The reduced form for y_2, in this case, is equal to

$$y_2 = (\alpha_2 + \gamma \alpha_1) + (\gamma \alpha_1 + \delta_2) P + \gamma \beta x + (\gamma u_1 + u_2)$$
$$= \pi_0 + \pi_1 P + \pi_2 x + v. \qquad (6)$$

Now there are three equations ($\pi_0 = \alpha_2 + \gamma \alpha_1$, $\pi_1 = \gamma \alpha_1 + \delta_2$, $\pi_2 = \gamma \beta$) and three unknowns. Therefore all the effects of interest can be identified. Solving these equations, the value of the parameters are $\gamma = \pi_2/\beta$ (the effect of coordination on performance) and $\delta_2 = \pi_1 - \pi_2/\beta\ \delta_1$ (the effect of policy that is not through coordination on performance).

To identify the effect of coordination on performance, evaluators would need to measure coordination through one or more of the network centrality indicators discussed in Section 2 (Table 1). These indicators would then be used as independent variables in the regression, measuring impact evaluation. In this way, it is possible to determine not only whether coordination influences firm performance, but also what kind of coordination, measured through different centrality indicators, is more likely to produce beneficial effects at the firm level.

5. Planning, Budgeting and Other Practical Considerations: Designing the Evaluation Plan and Strategy

Some practical elements of impact evaluation have already been addressed in this document (for instance, selecting the right methodology and gathering the data). The aim of this section is to review the key activities required to support the preparation and the implementation of a successful impact evaluation strategy. For this purpose, reference will be made to a fundamental tool for its success: the **evaluation plan**.

Although impact evaluations are usually conceived as an *ex-post* activity (i.e., something to be done at program completion), in practice, the conceptualization and design of a successful evaluation should follow a program from inception to completion. For this reason, the evaluation plan and strategy should be key components in the preparation of a CDP, and an evaluation specialist should always be included in the team responsible for the program's design and implementation.

Evaluation plans can have different formats. Here, we briefly discuss the outline and basic contents of what is considered an appropriate plan, on the basis of the standards adopted by the IDB and other multilateral organizations.

5.1. Main Evaluation Question(s) and Existing Knowledge

This section includes a brief discussion of the theory of change at the basis of the CDP to be evaluated. Section 2 provided an example of what the theory of change is behind the implementation of a typical CPD. Although this theory can, generally, be applied to many CDPs, an in-depth discussion with policymakers and other relevant stakeholders will definitely help the evaluation specialist to identify the specificity of the intervention model adopted by each CDP. Section 2 also defined the specific evaluation questions to be answered in the impact evaluation, including some examples. An evaluation should be put in the perspective of its contribution to existing knowledge of the effectiveness of the CDP, an extremely useful exercise because it helps researchers and evaluators to identify what evaluation questions may need more attention in order to close important knowledge gaps.

5.2. Key Outcome Indicators

An impact evaluation is about measurement. A detailed work on what indicators will be used in the analysis is critical and, often, research questions are too broadly defined because they are not clearly related to the measures of the expected outcomes that are SMART. For instance, a CDP objective may be to improve firm productivity. This may seem to be a rather specific objective, but measures of productivity vary significantly, depending on the specific production factor considered (it could be labor productivity, land productivity, and/or total factor productivity, or TFP) and on the specific assumption used to compute this measure (TFP, for instance, can be estimated in different ways). Although the choice of the specific indicators is often considered mostly a matter of data availability, it actually has more to do with the specific questions that the impact evaluation needs to seek answers to: for instance, if a CDP is expected to have a significant impact on the overall capacity of the beneficiary firms to manage their production, a measure of TFP will certainly be more accurate (specific, achievable, and relevant) than a measure of labor productivity.

A properly prepared evaluation plan should, therefore, show the consistency between research questions and indicators. In addition, for each indicator, the formula used for its computation should be clearly defined, and the frequency with which it will be calculated over the duration of the intervention should be clearly defined.

5.3. Evaluation Methodology[20]

In selecting the methodology to be used for an empirical analysis, it should always be remembered that the object of the analysis determines the methodology, not the other way around. So, in the case of an impact evaluation, the specific characteristics of the program determine which method can be used. Section 3 has extensively discussed what the main advantages and limitations of the methods are that can be used for the estimation of the effects of a CDP. In general terms, the main factor that differentiates these methodologies is how a counterfactual can be defined or, more practically, how a control group can be identified. The following will suggest a list of project characteristics that, without doubt, will affect how a comparison group can be identified.

---

[20] This section mainly refers to experimental and quasi-experimental approaches. Similar considerations for approaches using SNA techniques are developed in Sections 2 and 3. It builds upon Chapter 10 of Gertler et al. (2011) and Crespi et al. (2011).

a) **Targeting**. Equitable, transparent, and accountable targeting standards are a necessary precondition for proper identification of a comparison group. These rules can vary from program to program, but usually they define: (i) who is eligible for the program benefits (for instance, all the furniture producers with less than 10 employees, located in a certain municipality); (ii) the prioritization criterion (for instance, priority will be given to firms that apply first or to firms that submit the best business plan); (iii) the administrative mechanism to be used (for instance, one call for proposal a year or an "open window" approach); (iv) how many individuals in the eligible population are expected to be supported over a certain period (for instance, 500 firms over a five-year period); and the specific timing of the selection (for instance, 100 firms a year).

    It is quite clear how these targeting standards can affect the selection of a comparison group. For instance, when the target population is defined in such a way that the eligible population is much greater than the expected beneficiaries—and no strong prioritization criterion has been defined—then a randomized evaluation is the most feasible methodological option.

    The targeting standards should be designed consistently with the program's overall rationale and, therefore, are usually "a given" for the evaluation specialist.[21] Often, some standards can be equally valid from a programmatic point of view, in which case the specialist may propose the adoption of an alternative that would allow the implementation of the most rigorous evaluation method.

b) **Size of the intervention**. Econometric methods require a minimum number of observations to detect a certain level of effects. This minimum sample size can be easily calculated, following the procedure described in Section 3 herein. It is clear that the extent of the program strongly affects the possibility to apply any of the proposed methodologies. In some cases, a CDP could be simply too small for its effects to be estimated (or, at least, some of its effects, given that the sample size varies according to the specific outcome variable used in the estimation). In the case of small programs, the

[21] However, sometimes they may be defined on the basis of some institutional routine that has not really a clear connection with the project design. For instance, an organization may feel more comfortable in working with an "open-window" scheme, because it has always managed policy instruments, for which the choice among alternative investment projects has never been a relevant factor.

choice of methodologies should favor those that include multiple observations of the outcome variable over time. In fact, a reduced level of n (number of individuals/firms) can be "empowered" with a higher number of T (times these individuals/firms, which are observed).

c) **Budget and data availability**. Funding is always an issue, especially when it relates to allocating resources to an impact evaluation. This is probably due to the fact that such resources are not perceived as directly advantageous to the end beneficiaries of public programs. In actuality, they are advantageous, as they provide for a better understanding of what works and what does not work—probably one of the best investments in public policy.

In reality, budget considerations usually have very little to do with the selection of impact evaluation methodology. Perhaps because of some existing biases, policymakers often assume that there is a positive correlation between cost and the rigor of the methodology applied, which is incorrect. There is no reason why a randomized control trial (RCT) should cost more than a non-experimental methodology (in actual fact, an RCT may even cost less, given that the data processing is, in theory, much simpler than in most quasi-experimental approaches).

A budget can (and should) influence the choice of the impact evaluation methodology only in relation to data availability. As will be discussed in greater detail, a significant part of the costs of an impact evaluation are related to the need to work with micro data and, in many cases, to generate these data through ad-hoc data collections. However, these costs can be avoided, or dramatically reduced, when the evaluation can rely on existing data in which case, the choice of the evaluation methodology can be also driven by the opportunity to take advantage of existing (and available) data. A note of caution, however, is required: "data-driven" evaluations are not advocated, which would require a focus on studies for which data are already available or, even worse, would limit the evaluation questions to those that can be answered with pre-existing information. What is simply suggested is that, prior to considering a collection of new data, a thorough review of the existing data and an evaluation strategy be made, based on that data.

5.4. Sampling strategy and data collection

By this stage, the evaluation plan should have already introduced the basic criteria for the identification of the population (both treatment and control groups) that will be the object of the study. A more detailed discussion on the following topics should be included: (i) the unit of analysis and the strategy for its determination; (ii) the size of the sample to be used in the study; (iii) the distribution of the sample between treated and comparison groups; and (iv) a strategy for data collection, all of which were discussed in Section 3.

5.5. Resources

To be comprehensive, the plan should clearly determine the resources required for an evaluation. The following three factors are key elements: (a) to form a team and establish the responsibilities and tasks of each member; (b) to create a budget and work plan; (c) to identify source(s) of funding.

a) *Evaluation team: responsibilities and tasks*. Ideally, the team should be composed of a combination of external evaluators and operational experts (those who were involved in the implementation of the program). The external evaluators will (i) exercise more independence, since they have a significant stake in the success of the evaluation (rather than on the program's success) and have a desire to achieve a high degree of objectivity and credibility; and (ii) be more focused, as they will concentrate exclusively on the evaluation exercise, rather than program implementation. The operational experts will provide (i) a clear understanding of program objectives and the execution of mechanisms; (ii) easy and timely access to program data; (iii) smooth dialogue with the authorities; and (iv) stronger buy-in of the evaluation results.

   The plan should include the qualifications and technical expertise required for assignment. Depending on the program being evaluated and the resources available, the team should, collectively, include the following:

   - Design capability for evaluations, including choice of evaluation method; strong statistical background;
   - negotiating skills when liaising with key stakeholders regarding evaluation design;

- operational expertise for data collection[22], including relating to questionnaire and sampling plan development;
- Systems development expertise to promote an independent study with integrity;
- Statistical analysis management for estimate impacts;
- Presentation experience to report the findings to a wide audience, including academics and policymakers.

It is important to keep in mind that the operation of the program and the implementation of the evaluation are interlinked and, therefore, should not be dealt with separately. The IDB and similar institutions consider the following as best practice to ensure good linkage between both:[23]

- The evaluation team should have one or more key qualified people as part of the operations team (both within the IDB and with the respective government or implementing agency), who will participate in the design of the evaluation and be accountable for its implementation. These members can play a critical role in ensuring that the program implementation actually follows the agreed framework. Should there be deviations from the initial design, they can work with the evaluation experts to resolve the issue, thus minimizing the impact on the integrity of the process.
- The evaluation team should include independent technical experts who, in addition to providing advice on such elements as methodologies, measures, and sampling, do not have a vested interest in the project, and can guarantee independence and credibility. They can include evaluation experts from within the IDB, external consultants (including academics), or outside firms (technical assistance firms, think tanks, research institutes, universities, etc.).

Budget and work plan. The evaluation plan should include a careful estimate of the resources needed to fund the exercise. For this purpose, the design should contain a work plan that describes each team member's responsibility and at what point an activity should

[22] If the project team plans to procure external individuals or firms to help conduct the impact evaluation, it can be advantageous to consider the evaluation and data collection as two separate activities that can potentially be executed by two different individuals or agencies. For this model to be successful, the two entities will need to work cohesively, with some coordination by the team leader.
[23] We are indebted to Sebastian Martinez for these suggestions.

take place, as well as a budget for each activity, in order to accurately determine funding requirements, mobilize resources, and ensure appropriate levels of funding.[24] Table 4 presents an example of a budget.

As Table 4 illustrates, an important driver of evaluation costs is the funding needed for the data collection. In a recent review of the World Bank's impact evaluations, it was concluded that over half of the resources were devoted to data collection activities, alone. This cost depends on various factors, but two key ones that should be considered during the early stage of design are the size of the sample and the number of rounds of data collection that is necessary. The Living Standards Measurement Study (LSMS) Manual provides cost estimates for collecting data through household surveys, based on country experience. Chapter 10 in Gertler et al. (2011) provides costs associated with various impact evaluations that were conducted by The World Bank.

Source of funding. There are many potential sources of funding for an impact evaluation exercise, the main one being the government that is implementing the program. While governments often view evaluation resources as competing with resources meant for the programs, themselves, it is essential for a government to recognize that impact appraisals are critical to improving a program's objectives in terms of timely interventions and improvement in the allocation of resources across interventions. Thus, budget allocations should be made not only on the basis of cost, but also—as for any investment—on the basis of the benefits. In many cases, while the cost of an impact evaluation exercise may appear relatively high in absolute terms, the investment in producing evidence on what works and what does not work could present tremendous returns on the decisions of whether to disband a program, modify it, or scale up public programs.

At the IDB, potential funding for evaluations include the resources from (i) the program, itself (IDB and/or counterpart funding), (ii) technical cooperation (TC), (iii) operational inputs (OI), (iv) knowledge and capacity building projects (KCPs), and (iv) departments/divisions. Sources outside the IDB include organizations, such as the International Initiative for Impact Evaluation (3ie), Global Development Network (GDN) and U.S. National Institutes for Health (NIH).

[24] Gorgens, M., Kusek, J. Z. (2009).

Table 4. Budget and Work Plan for an Impact Evaluation

Activities	Year 1				Year 2				Year 3				Responsible	Cost (US$)	Funding Source
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4			
I. Impact Evaluation															
Personnel															
Monitoring and Evaluation (M&E) specialist at executing agency															
Principal researcher															
Research assistant															
Statistics specialist															
Field coordinator															
Travel															
International and/or domestic air travel															
Ground transportation															
Per diem															
Data collection															
Design of questionnaire															
Pilot testing															
Training															
Travel and per diem															
Survey materials, equipment															
Printed questionnaire															
Field personnel															
Interviewers															
Supervisors															
Transport (vehicles and fuel)															
Drivers															

Data entry																
	Data cleansing and data entry															
Analysis and diffusion of data																
	Workshops															
	Documents, reports															
Other																
	Office space															
	Communications															
	Software															
Impact evaluation report																
	Final Report															

5.6. Diffusion of Results

A component of coordinating evaluation activities includes the timing, content, and delivery of the appraisal report. A report with baseline data will generally cover the success of the data collection and whether a control group representing a reasonable counterfactual to the comparison group was done within the months following data collection. A report on any post-treatment data collection following a survey should also be included. Finally, the report should describe all the rounds of data collection, based on the indicators, each of which should be included in the evaluation plan.

The final results of the evaluation should be disseminated, using appropriate methods of communication. Experts may tend to steer towards journals through which scientific articles can be effective in not only disseminating the knowledge gained from an evaluation among renowned scholars and policymakers, but also in guaranteeing the credibility of the results.

Scientific articles should not be considered the only avenue of communication to describe the outcome of an impact evaluation. Other methods of communication should be used, such as: (i) executive summaries for management, high-level civil servants, and ministries; (ii) press releases; and (iii) videos and presentations for civil society organizations and the general public

6. Conclusions

The idea of promoting the formation and development of clusters has become, increasingly, a global concept, with a proliferation of programs having been implemented. They are based on the assumption that firm-level performance will benefit from agglomeration and, in turn, agglomeration will facilitate coordination, particularly in instances where the failure of coordination can significantly hamper the development of industry in certain localities. CDPs have been shown to create incentive and to promote public support to facilitate the interaction and coordination among all agents (private-private and public-private).

Most CDPs include two stages: one targets the coordination of the cluster's actors to prioritize investment decisions, where industry and location specific externalities are accounted for; and one implements the actions defined among the priorities. CDPs can include the co-financing of public infrastructure and club goods that can become a catalyst for new investment projects.

Do CDPs work? Will they be effective in producing results? Do they have an impact on enterprise development? These guidelines offer an insight into the methods that can help answer these fundamental questions through solid quantitative evidence. Answers in any event, will be rather difficult to arrive at, since the process of cluster development and the effort to support it are very complex and multidimensional, and involve decisions on an individual and collective basis. Results will depend on the level of coordination that is achieved and on the actions undertaken as a result of improved coordination and strategy-setting of the relevant actors. A rigorous appraisal of the effectiveness of these programs will indicate, at least, whether coordination among cluster actors do, in fact, change throughout the project, how this takes place, and whether CDPs are the actual cause of the observed results.

The techniques of Social Network Analysis (SNA) can be employed to assess the evolution of coordination among cluster actors, with the requirement that network indicators are observed before and after the implementation of the CDPs. Whilst this particular analysis can assist in monitoring and assessing the process of coordination and its changes throughout the program, other qualitative and contextual information can also assist in interpreting the data and, thus, increase the reliability of results. However, in order to properly assess the impact of CDPs, their causality needs to be explored further by the application of additional quantitative methods. In fact, the effects cannot be attributed to the program, itself, unless a proper counterfactual is built in, such as what would have happened to the beneficiaries in the absence of the program. By definition, this particular counterfactual cannot be observed, but the application of experimental and quasi-experimental techniques can help construct control groups of non-beneficiaries to approximate the counterfactual—already explained at length in the text. Furthermore, a detailed observation of cases and specific interviews can help regarding the interpretation of results derived from the application of both SNA and experimental and quasi-experimental techniques.

To summarize, clusters and cluster development programs are diverse and multi-dimensional phenomena that require a variety of instruments to be fully understood and assessed. The quantitative tools that have been proposed in this publication should be considered as complementary and not as alternatives, with each applied as a means to strengthen the explanatory capacity of the other. Each tool requires specific and challenging data analysis that can be achieved with careful resource planning and having the appropriate team skill set. The overarching objective is to build new and solid evidence on the effectiveness of CPDs and their respective policies. This is crucial in order to justify the funding of these programs with public resources, ensure that recipients are accountable, and, more importantly, to learn how to improve program effectiveness and contribute to improving livelihoods through enterprise and socio-economic development.

References

Alvarez, R., G. Crespi, and C. Cuevas. 2012. "Public Programs, Innovation, and Firm Performance in Chile." Technical Note IDB-TN-375, Inter-American Development Bank, Washington DC.

Andersson, T., E.W Hansson, S. Schwaag Serger., and Sörvik, J. 2004. "The Cluster Policies Whitebook." IKED, Sweden, 2004.

Angelucci, M. and G. De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Injections Affect Ineligibles' Consumption?" *American Economic Review*, 99(1), 486–508.

Angelucci, M. and V. Di Maro. 2010. "Program Evaluation and Spillovers Effects." Impact-Evaluation Guidelines, Technical Notes No. IDB-TN-136, Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank.

Angrist, J.D. and J-S. Pischke. 2008. "Mostly Harmless Econometrics: An Empiricist's Companion." Princeton, NJ: Princeton University Press.

Arráiz, I, M. Meléndez, and R. Stucchi .2012. "Partial Credit Guarantees and Firm Performance: Evidence from the Colombian National Guarantee Fund." Inter-American Development Bank, OVE Working Papers 0212, Washington, DC.

Arráiz, I., F. Henriquez, and R. Stucchi. 2012. "Supplier Development Programs and Firm Performance: Evidence from Chile." *Small Business Economics*.

Bamberger, M. 2010. "Reconstructing Baseline Data for Impact Evaluation and Results Measurement." PREM Notes No. 4, Special Series on The Nuts & Bolts of M&E Systems, The World Bank.

Barabasi, A-L., Albert, R. 1999. Emergence of Scaling in Real Networks. *Science* 286: 509–12.

Bartelsman, E.J. and M. Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38: 569–94.

Becattini G. 1989. "Sectors and/or Districts: Some Remarks on the Conceptual Foundations of Industrial Economics." In E Goodman, Bamford, J., eds., *Small Firms and Industrial Districts in Italy*. Routledge.

Benavente, J., G. Crespi, and A. Maffioli. 2007. "Public Support to Firm-Level Innovation: An Evaluation of the FONTEC Program." Working Paper OVE/WP/05/07, Office of Evaluation and Oversight, Inter-American Development Bank, Washington, DC.

Borgatti S.P, Molina J-L. 2003. "Ethical and Strategic Issues in Organizational Social Network Analysis.".*Journal of Applied Behavioural Science*, 39 (3): 337–49.

Borgatti, S.P. and Molina, J.L. 2005. "Toward Ethical Guidelines for Network Research in Organizations." Social Networks. 27(2): 107–17

Bound, J., D. Jaeger, and R. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90 (430): 443–50.

Cai, H., Y. Chen, H. Fang, and L-A. Zhou. 2010. "Microinsurance, Trust and Economic Development: Evidence from a Randomized Natural Field Experiment." National Bureau of Economic Research.

Castillo, V., A. Maffioli, S. Rojo, and R. Stucchi. 2013a. "The Effect of Innovation Policy on SMEs' Employment and Wages." *Small Business Economics*, Forthcoming.

Castillo, V., A. Maffioli, S. Rojo, and R. Stucchi. 2013b. "Estimating Spillover Effects of Technology Development Funds: The Case of FONTAR." Unpublished.

Ciccone, A., and R. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review*, 86(1): 54–70.

Chudnovsky, D., A. López, M. Rossi and D. Ubfal. 2006. "Evaluating A Program of Public Funding of Private Innovation Activities. An Econometric Study of FONTAR in Argentina." Working Paper /OVE/1606, Office of Evaluation and Oversight, Inter-American Development Bank, Washington DC.

Cole, H.L., L.E. Ohanian, A. Riascos, and J.A. Schmitz, Jr. 2005. "Latin America in the Rearview Mirror." *Journal of Monetary Economics*, 52: 69–107.

Crepon, B., E. Duguet, and J. Mairesse. 1998. "Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level." NBER Working Papers 6696, NBER.

Crespi, G.A., A. Maffioli, and M. Melendez. 2011. "Public Support to Innovation: The Colombian COLCIENCIAS' Experience." Technical Note IDB-TN-264, Inter-American Development Bank, Washington, DC.

Crespi, G.A., A. Maffioli, P. Mohnen, and G. Vázquez. 2011. "Evaluating the Impact of Science, Technology and Innovation Programs: A Methodological Toolkit." Inter-American Development Bank, Washington, DC.

Crespi, G.A., G. Solís, and E. Tacsir. 2011. "Evaluación del impacto de corto plazo de SENACYT en la innovación de las empresas panameñas." IDB Publications IDB-TN-263, Inter-American Development Bank, Washington, DC.

Das, S., M.J. Roberts, and J.R. Tybout. 2007. "Market Entry Costs, Producer Heterogeneity, and Export Dynamics." *Econometrica*, 75(3): 837–73.

De Negri, J.A., M. Borges Lemos, and F. De Negri. 2006. "Impact of P&D Incentive Program on the Performance and Technological Efforts of Brazilian Industrial Firms." Working Paper OVE/WP/14/06, Office of Evaluation and Oversight, Inter-American Development Bank.

Dini, M., Ferraro, C., Gasaly, C. 2007. "Pymes y articulación productiva. Resultados y lecciones a partir de experiencias en América Latina." CEPAL, División de Desarrollo Productivo y Empresarial, *Serie Desarrollo Productivo*, número 180, Santiago de Chile.

Ellison, G. and E.L. Glaeser. 1997. "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach." *The Journal of Political Economy*, 105 (5): 889–927.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia." *Journal of Development Economics*, 75(2): 333–71.

Europe INNOVA. 2008. “The Concept of Cluster and Cluster Policies and Their Role for Competitiveness and Innovation: Main Statistical Results and Lessons Learned.” Commission Staff Working Document, SEC (2008) 2637.

Feldman, M.P. and D.B. Audretsch. 1999. “Innovation in Cities: Science-based Diversity, Specialization and Localized Competition.” *European Economic Review*, 43, 409–29.

Figal, L., A. Maffioli, C.M. Rodriguez, G. Vázquez, J.A. De Negri. 2012. “Assessing the Impact of Cluster Policies: The Case of the Arranjos Productivos Locais in Brazil.” SPD Working Papers 1203, Inter-American Development Bank, Washington, DC.

Foster, L., J.C. Haltiwanger and C.J. Krizan. 2001. “Aggregate Productivity Growth: Lessons from Microeconomic Evidence.” In E.R. Dean, M.J. Harper, and C.R. Hulten, eds., *New Development in Productivity Analysis*. Chicago: University of Chicago Press.

Foster, L., J. Haltiwanger and C. Syverson. 2008. “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” *American Economic Review*, 98(1), 394-425.

Fujita, M., P. Krugman and A.J. Venables.2001. “The Spatial Economy: Cities, Regions, and International Trade.” Boston: The MIT Press.

Gertler, P.J., S. Martinez, P. Premand, L.B. Rawlings, C.M.J. Vermeersch. 2011. “Impact Evaluation in Practice.” The World Bank, Washington DC.

Giuliani E. 2005. “Cluster Absorptive Capacity: Why Some Clusters Forge Ahead and Others Lag Behind?” *European Urban and Regional Studies*, 12 (3): 269–88.

Giuliani E, Bell M. 2005. "The Micro-determinants of Meso-level Learning and Innovation: Evidence from a Chilean Wine Cluster." *Research Policy*, 34(1): 47–68.

Giuliani G. 2013. “Network Dynamics in Regional Clusters: The Perspective of an Emerging Economy.” *Research Policy*, in press.

Giuliani E., Matta A. 2013. “Impact Evaluation with Social Network Analysis Methods, Program for Supply Chain Development in the Province of Cordoba, Argentina.“ (ATN/ME-8112-AR). Unpublished.

Giuliani E., Pietrobelli C., 2011. "Social Network Analysis Methodologies for the Evaluation of Cluster Development Programs." Technical Note IDB-TN-317, Inter-American Development Bank, Washington DC.

Glaeser, E., H. Kallal, J. Scheinkman and A. Shleifer. 1992. “Growth in Cities.” *Journal of Political Economy*, 100 (6): 1126–52.

Granovetter, M. 1985. “Economic Action and Social Structure: The Problem of Embeddedness.” *American Journal of Sociology* 91: 481–510.

Greunz, L. 2004. “Industrial Structure and Innovation: Evidence from European Regions.” *Journal of Evolutionary Economics*, 14: 563–92.

Guerrieri P., Iammarino S., Pietrobelli C. 2001. “The Global Challenge to Industrial Districts: Small and Medium Sized Enterprises in Italy and Taiwan.” Cheltenham, U.K Edward ELGAR.

Guerrieri P. and Pietrobelli C. 2006 "Old and New Forms of Clustering and Production Networks in Changing Technological Regimes: Contrasting Evidence from Taiwan and Italy." *Science, Technology and Society* 11(1): 9–38.

Guillén, M.F. 2000. "Business Groups in Emerging Economies: A Resource-Based View." *The Academy of Management Journal* 43(3): 362–80.

Hanneman, R.A. and M. Riddle. 2005. "Introduction to Social Network Methods." Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/

Hanson, G.H. 2001. "Scale Economies and the Geographic Concentration of Industry." *Journal of Economic Geography*, 1(3): 255–76.

Heckman, J.J., J. Smith, and N. Clements. 1997. "Making the Most out of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts." *Review of Economic Studies* 64 (4): 487–535.

Heinrich, C. A. Maffioli, and G. Vázquez. 2010. "A Primer for Applying Propensity-Score Matching." Technical Note IDB-TN-161, Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank, Washington DC.

Henderson, V., A. Kuncoro, and M. Turner. 1995. "Industrial Development in Cities," *The Journal of Political Economy*, 103 (5): 1067–90.

Hoff, K. 2000. "Beyond Rosenstein-Rodan: The Modern Theory of Coordination Problems in Development." in B. Pleskovic and N. Stern (Eds.) *Proceedings of the Annual World Bank Conference on Development Economics, 2000*, The World Bank.

Holland, P. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association,* 81: 945–60.

Hulten, C. 2001. "Total Factor Productivity: A Short Biography." In E.R. Dean, M.J. Harper, and C.R Hulten, eds., *New Developments in Productivity Analysis*. Chicago: University of Chicago Press.

IDB. 2010. "Development Effectiveness Overview 2008–2009." IDB Publications, Inter-American Development Bank, Washington D.C.

Imbens, G.W. and J.D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62 (2): 467–75.

Imbens, G.W. and J.M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47 (1): 5–86.

Khanna, T., Rivkin, J.W. 2001. Estimating the Performance Effects of Business Groups in Emerging Markets. *Strategic Management Journal* 22: 45–74.

Krugman, P. 1992. "Geography and Trade." (Gaston Eyskens Lecture Series), The MIT Press.

Krugman, P. 1994. "Competitiveness: A Dangerous Obsession." *Foreign Affairs*, 73(2): 28–44.

Lewis, W. 2005. *The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability*. Chicago: University of Chicago Press.

Lucas, R. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22 (1): 3–42.

Lundvall B-Å. 1988. "Innovation as an Interactive Process: From User-Producer Interaction to National Systems of Innovation," In G. Dosi, C. Freeman, R. Nelson, G. Silverberg, L.L.G. Soete, eds., *Technical Change and Economic Theory*. London: Pinter.

Maffioli, A. 2005. "The Formation of Network and Public Intervention: Theory and Evidence from the Chilean Experience." ISLA Working Paper 23, Università Commercial Luigi Bocconi, Milan.

Markusen, A. 1996. Sticky "Places in Slippery Space: A Typology of Industrial Districts." *Economic Geography*, 72(3): 293–313.

Marshall, A. 1920. *Principles of Economics*. 8th Edition, Macmillan & Co., Ltd., London.

Martin, R., P. Sunley. 2003. Deconstructing Clusters: Chaotic Concept or Policy Panacea? *Journal of Economic Geography* 3(1): 5–35.

Martin, P., T. Mayer, and F. Mayneris. 2008. "Public Support to Clusters: A Firm Level Study of French 'Local Productive Systems'." CEPR Discussion Paper No. 7102, Centre for Economic Policy Research.

McDermott, G.A., R.A. Corredoira, G. Kruse. 2009. "Public-Private Institutions as Catalysts of Upgrading in Emerging Market Societies." *Academy of Management Journal*: 52 (6): 1270–96.

Melitz, M.J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Melitz, M.J., and G.I.P. Ottaviano. 2005. "Market Size, Trade, and Productivity." *Review of Economic Studies*, 75(1): 295–316.

Mesquita, L.F., S.G. Lazzarini. 2008. "Horizontal and Vertical Relationships in Developing Economies: Implications for SMEs' Access to Global Markets." *Academy of Management Journal* 51(2): 359–80.

Miguel, E. and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.

Nickell, S. 1996. "Competition and Corporate Performance." *Journal of Political Economy*, 104(4): 724–46.

Pavcnik, N. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies*, 69(1): 245–76.

Porter, M.E. 1990. *Competitive Advantage of Nations*. New York: Free Press.

Pietrobelli C. and C. Stevenson. 2011, "Cluster Development Programs in Latin America and the Caribbean: Lessons from the Experience of the Inter-American Development Bank." IDB Discussion Paper IDP-DP-188, Inter-American Development Bank, Washington, DC.

Pietrobelli C., Casaburi G., Maffioli A. Forthcoming, "Solving Coordination Failures through Cluster Development Programs." In IDB, *Development in the Americas 2014*, Washington DC: Inter-American Development Bank.

Powell, W.W., K.W. Koput, L. Smith-Doerr. 1996. "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology." *Administrative Science Quarterly* 41(1): 116-145.

Rodriguez-Clare, A. 2005. "Coordination Failures, Clusters and Microeconomic Interventions." Working Paper 544, Inter-American Development Bank, Washington DC.

Rodriguez-Clare, A. 2007. "Clusters and Comparative Advantage: Implications for Industrial Policy." *Journal of Development Economics*, 82: 43–57.

Romer, P.M. 1986. "Increasing Returns and Long-Run Growth." *The Journal of Political Economy*, 94 (5): 1002–37.

Rosenbaum, P.R. and D.B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70 (1): 41–55.

Rosenstein-Rodan, P.N 1943. "Problems of Industrialisation of Eastern and South-Eastern Europe." *The Economic Journal*, 53 (210/211): 202–11.

Rosenthal, S.S. and W.C. Strange. 2003. "Geography, Industrial Organization, and Agglomeration." Center for Policy Research, Paper 107.

Schmitz H. 1995. "Collective Efficiency: Growth Path for Small-Scale Industry." *Journal of Development Studies,* 31(4): 529–66.

Snijders, T.A.B. 2001. "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology*: 31(1): 361–95.

Snijders, T.A.B. 2005. "Models for Longitudinal Network Data." In P. Carrington, J. Scott, S. Wasserman, eds., *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.

Snijders, T.A.B., G. van de Bunt, C.E.G. Steiglich. 2010. "Introduction to Actor-based Models for Network Dynamics." *Social Networks*, 32: 44–60.

Stecklov, G. and A. Weinreb. 2010. "Improving the Quality of Data and Impact-Evaluation Studies in Developing Countries." Technical Note IDB-TN-123, Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank, Washington DC.

Syverson, C. 2011. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6): 1181–1222.

Syverson, C. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2), 326–65.

Tybout, J. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38: 11–44.

Van Biesebroeck, J. 2005. "Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms." *Journal of International Economics*, 67 (2): 373–91.

Van Biesebroeck, J. 2007. "Robustness of Productivity Estimates." *The Journal of Industrial Economics*, 55(3): 529–69.

Volpe M., C. and J. Carballo. 2007. "Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and Extensive Margins of Exports." Presented at the 9th Annual Conference of the European Trade Study Group, Athens (September 15, 2007), Inter-American Development Bank, Washington DC.

Wasserman S, Faust, K. 1994. *Social Network Analysis. Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge: Cambridge University Press.