

Multidimensional assessment of technology and innovation programs: the impact evaluation of INCAGRO-Peru

Sergio L M Salles-Filho, A Flavio Avila,
Juan Ernesto O S Alonso and Fernando A B Colugnati

This article presents a methodological proposal and an empirical validation for the assessment of ST&I programs, particularly for choosing indicators and metrics from a multidimensional perspective. Its key feature is the construction of indicators designed to apprehend the wide array of characteristics displayed by the many different types of innovation program. Thus it combines deductive procedures (decomposition of goals) and inductive procedures (discussion with the players involved) with evaluation results and impacts in specific dimensions linked to the nature of a given program. We illustrate the approach using a World Bank Program called INCAGRO. The proposed method is a contribution to systematizing the definition of indicators and metrics for the assessment of ST&I programs. It helps define ‘what should be evaluated’ in a given program. Many different (unforeseen) results and impacts can emerge simultaneously with the main subject of the program. Thus multidimensional issues and multiple actors should be considered in the impact evaluation process.

METHODOLOGIES FOR assessment or evaluating innovations in agriculture have developed intensely in the last decade (Hall *et al.*, 2003; Adrian *et al.*, 2005; Payraudeay and Van der Werf, 2005; Ny *et al.*, 2006; Gomés and Gonzáles García, 2007; Jassen and Van Ittersum, 2007). The main driver of this search for increasingly

refined methodologies is the need to measure and forecast the impacts of innovation more accurately. The inherent complexity of natural and social systems, together with the uncertainty involved, continues to present challenges, hindering the full achievement of this goal, but the dynamism of the objects evaluated forces methodological proposals to evolve uninterruptedly (Blazy *et al.*, 2009), in terms of both tools and ways of addressing the challenges.

Sergio L M Salles-Filho (contact author) and Juan Ernesto O S Alonso are at the Grupo de Estudos sobre Organização da Pesquisa e da Inovação, Departamento de Política Científica e Tecnológica, Instituto de Geociências, Universidade Estadual de Campinas, São Paulo, Brazil; PO Box: 6152 CEP 13083-970 Campinas/SP, Brazil; Email: sallesfi@ige.unicamp.br; juan.ernesto@gmail.com. A Flavio Avila is at the Empresa Brasileira de Pesquisa Agropecuária, SAIN Parque Estação Biológica, Asa Norte, 70770-910 – Brasília, DF, Brazil; Email: avila@sede.embrapa.br. Fernando A B Colugnati is at the Instituto de Pesquisas em Tecnologia e Inovação, Av. São Luís, 86, cj 192 – 19 andar, Cep 01046 000, São Paulo, Brazil; Email: fernando@ipti.org.br.

The methods used to measure the impact of innovations in agriculture have tended to focus on economic issues (Alston *et al.*, 2001), environmental issues (Payraudeay and Van der Werf, 2005; Ny *et al.*, 2006), land use (Gomés and Gonzáles García, 2007) and evaluation of ‘farm models’ (Jassen and Van Ittersum, 2007). Although all these authors consider the innovation dimension, none of them makes it the central axis of the method. The majority of the authors have evaluated these impacts in an isolated form.

The authors would like to thank INCAGRO personnel for the opportunity to develop the project that originated this article and for their contribution to the methodology and to the data collection.

This article presents a multidimensional method to assess the impact of technology and innovation programs at farm level that considers the appropriation of

knowledge or technology as the central axis of the innovation process. The method was developed by the authors and applied at INCAGRO.¹

INCAGRO is a World Bank-financed program of the Peruvian Government set up with the principal mission of promoting innovation among small farmers by supporting research, training and the introduction of new management and production techniques. It provides co-financing for projects through a competitive bidding process. It lasted initially for four years (2001–4) and was then renewed for a second period (2005–9). In its first phase, INCAGRO invested US\$6.4 million and arranged a further US\$6.6 million in monetary and non-monetary resources, accounting for more than 50% of overall ST&I investment in small farming (‘crowding-in’). In the second phase it has so far adjudicated US\$11.9 million.² INCAGRO was evaluated independently to validate its results and its impact on small farming in Peru.

The following sections outline the conceptual framework on which the multidimensional assessment methodology is based, a selection of findings to show how the methodology operates and its potential coverage, and finally the principal advantages and drawbacks, based in the experience obtained with the INCAGRO evaluation.

Conceptual model

The methodology framework described here for multidimensional assessment of the impacts of innovation and technological development programs is based on the concept of innovation. This covers the assessment of the knowledge or technology appropriation in a broad context of social relations that may or may not show up through the market (Bin and Salles-Filho, 2008).³ Thus, when the assessment is based on the concept of social appropriation, innovation takes place only when the effect on the potential users is observable. This approach should take into account a broader and more complex set of actors and spaces than is usual in traditional approaches to assessment, which are not based on the concept of innovation as social appropriation of the benefits generated. Figure 1 shows the relationship between R&D and innovation. With regard to what Figure 1 shows, the following points are worth highlighting:

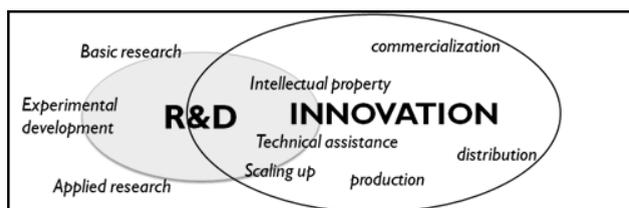


Figure 1. Relationships between research and development and innovation

- R&D and S&T activities have a partial connection to innovation, that is, not all R&D nor all technology is converted into innovation (Freeman, 1998; Rosenberg, 1992);
- Innovation is not always related to R&D and S&T activities, that is, many innovations are not science-based and do not derive from research (Rosenberg, 1982);
- Innovations are produced by different actors, some related to S&T and others not (Nelson, 1988); and
- The interactions between R&D and innovation have multiple causes and are almost always two-way (Kline and Rosenberg, 1986).

Thus innovation is a collective, socially constructed process whose impacts may have multiple dimensions (economic, social, environmental, political, organizational, etc.), and an impact assessment must take into account the multidimensional nature of the effects of the innovation process.

In the specific case of INCAGRO, therefore, multidimensional assessment should identify the results and impacts of agricultural activities throughout the production and commercialization chain, focusing above all on the impacts in the social, environmental and economic dimensions.

Multidimensional assessment process

The multidimensional impact assessment process comprises four stages:

1. Decomposition and validation of relevant themes and indicators;
2. Production of a data collection tool (specific questionnaires with built-in redundancy to verify causality for different program stakeholders);
3. Design of the sample; and
4. Statistical analysis (univariate and multivariate) and qualitative analysis.

Decomposition identifies the themes to be considered in the multidimensional impact assessment, the most appropriate metrics to be used in gauging the results and impacts for each theme, and the relevant actors who participate in the program, based on the views of all key stakeholders in the program being assessed. This process uses the classic method of scientific analysis, which begins with specific and simplified observation of a given reality to try to understand the general laws that govern it, combined to a dynamic and progressive method for quantification and qualification of results and impacts (Salles-Filho *et al.*, 2007).

The reality observed, the object of the decomposition process, is broken down into a sufficient number of parts to identify general laws that enable the phenomena observed to be explained. Once the hypothesis has been validated by explaining each of the parts, the various explanations are confronted

and concatenated, and the verification of isomorphic structures in different sets of hypotheses and true theses leads to progressive theoretical generalizations. It thereby becomes valid to deduce the explanation of how the reality in question functions and is structured by observing particular behaviors. Thus it is possible to explain phenomena by means of structures that relate the different natures of their parts, decomposing the relevant parts in terms of different dimensions and actors rather than just a single relationship.

The scale of indicators and methods used to build it must adequately capture the phenomena described and be validated by the program stakeholders, who possess knowledge and perceptions that enable the program to be seen in its totality (producers and consumers of innovation). In other words, this means that selected indicators must meet the following conditions:

- Be related to the theme or part of reality it is intended to explain, expressing a clear causal relationship between input and output (pertinence);
- Be redundant (complementarity);
- Be precise or unequivocal; and
- It must be obtainable with an adequate level of quality and acceptable cost (viability).

The classic logical-deductive method works only for closed systems or for those that can be measured with reasonable exactness. In the case of social and other open systems, indeed when social structure modification or transition is present, mathematical and computational formalization is far more complex due to its inherent non-ergodic behavior.⁴ Thus while obtaining magnitudes from observations of reality to measure quantifiable impacts and evaluation routines to implement computable and stable algorithms presupposes the classic method, identifying impacts within the framework of innovation systems requires a dynamic and progressive method because of the complexity of the systems being evaluated. For the multidimensional evaluation method to reconcile these antagonistic requirements, therefore, it is necessary to concatenate in structured contextual level the classic principles, of analysis and of synthesis, and to build models that measure results and impacts from premises, bearing in mind the coexistence of different dimensions of analysis and multiple rationalities identified – that is, the self-perception of the object have being assessed.

The key feature of the decomposition method is the construction of indicators designed to systematize learning and knowledge identified for the various parties and relevant actors in the observed reality, pinpointing the ensemble of perceptions that an innovation promotion program may have. Thus, from the perspective based on the different actors is possible to reach a heuristic understanding of the program's impacts. The method combines deduction with induction (based on the views of stakeholders),

The decomposition method comprises two main elements: decomposition of the formal and informal goals of the program (obtained from documents and interviews), and dialogue with the participants through structured panels

in order to evaluate the results and impacts for specific dimensions of ST&I programs and in this respect it is itself an innovation among assessment or evaluation methods, particularly in light of its ability to focus more sharply on analysis of the object and the participation of stakeholders.

The decomposition method comprises two main elements: decomposition of the formal and informal goals of the program (obtained from documents and interviews), and dialogue with the participants through structured panels. The themes and indicators used for the assessment arise from these two activities. Figure 2 shows a summary of the method. Each of the five steps shown is detailed in the following subsections.

The next stage is designing the data collection tool, that is, the questionnaire. This is derived from the results of the decomposition process (which identifies themes, indicators and metrics) and comprises three steps. The first profiles the ST&I program under assessment and its participants. The second outlines the themes and indicators directly relating to ST&I and prioritized in the decomposition process. The third focuses on the themes and indicators relating to the dimensions in which the technologies and innovative processes created by ST&I were appropriated. The questionnaire is designed to find out first what technologies or new ways of doing certain things have been converted into innovation, and then to identify the impacts on the dimension prioritized (e.g. economic). This means locating the subject around the object analyzed and determining the central axis along which the reality is to be observed, followed by extraction of the subject's perceptions regarding the results and impacts on the various dimensions concerned. Throughout the construction process the aim

- 
1. Analysis of program goals
 2. Decomposition of goals into terms
 3. Transformation of terms into evaluation themes
 4. Identification of indicators to qualify and measure assessment themes
 5. Validation of indicator set by theme

Figure 2. Synthetic schema of decomposition method

is to maximize the quantity and quality of information collected via a relatively small number of structured questions.

Attributing causality

Thinking of assessment, the causality attribution plays one of the main roles, since a reliable interpretation of the actual effect of the program is a key issue for the decision process. In other words, the quantifiable assessment of the pure effect of the program, controlled by other sources of possible effects (taking off these effects), is the main goal. Effectively, this can be achieved only by means of *proxy* methods,

Assessments of ST&I programs generally have presented a lack in capturing the inherent dynamics of the innovation process, particularly in relationships among stakeholders and the multiple effects that might influence the impacts of the program under evaluation. A temptation to cover this lack is proposed by Rogers and Bozeman (2001) under the *knowledge value framework*, consisting of two central concepts: *knowledge value collective* (KVC) and *knowledge value alliance* (KVA). The KVC represents a set of actors connected themselves by the specific use of an information corpus, driven to an application that induces the value attribution for the knowledge provided by that information and its complements. The KVA targets the capacity for creating value to the collective efforts.

These methods are examples of the growing trend to create evaluation methodologies such that the assessment of the value originated by the generation and communication of the knowledge is the focus. It is well known that this is not an easy task, indeed because the causality attribution between the observed/ measured and the program interventions is methodologically complex. This attribution for socially constructed processes, as knowledge construction and appropriation are, brings methodological challenges that are still open. Some recent studies presented in the literature also try a blended approach, like using the change theory and comparison groups defined *ex-post* (Weitzman *et al*, 2002). In fact, few program theory evaluations have tried to address causality attribution (Rogers, 2000).

When using comparison or control groups, the most applicable technique, given a specific indicator, is the comparison of the differences between a baseline observation (t_0), prior the program, and a posterior observation (t_1), from both groups. The difference between these two differences, given that other variables are controlled (e.g. matched groups, propensity scoring, conditional regression), is the effect attributable to the program. So, additionality representing the change between repeated observations is the basis for causality attribution.

However, because of all the difficulties in quantifying causality for this type of evaluation, it is fundamental to introduce a factor for causality

verification, as is proposed herein. The nature of the multidimensional assessment proposed, with many dimensions, and sometimes different types of stakeholders and/or actors, makes even harder the definition of adequate control or comparison groups. The *proxy* chosen, and subject of this article, is an alternative methodology that makes use of the additionality and a redundant causality identifier. Actually this redundancy relies on the structure of questions about the indicators, trying first to identify if there is any change, if so what was the direction of that change, how much changed and, finally, what part of this change may be considered (in the respondent judgement) attributable to the program under evaluation. Note that perceived redundancy is the subjective causality of the program's effect on change observed in an indicator. This mean assumes, first, the stakeholders or actors used their best information, from their set of knowledge and experience, to measure their perception of attribution of an impact's effect in changes observed in indicator. Second, the number of stakeholders or actors involved in assessment impact is large enough to achieve statistical significance in subjective causality. What follows is a deeper description of this structure, and its mathematical formalization, aiming to present an expression to estimate the impact.

The first question in the sequence aims to find out whether there has been a change in the indicator being evaluated due to the ST&I program and if so in which direction. We will denote the response to this answer as:

$$g_{ij} = [-1, 0, 1].$$

The second asks for a description of the magnitude of the change observed, or

$$(|a_{ijt1} - a_{ijt0}| = |\Delta(a_{ij})|)^5.$$

This can be calculated objectively, asking the measures in both moments, or subjectively asking the change in terms of percentage ranges or semantic scales, like Likert Scales. The choice of either of these approaches depends on the nature of the indicator, the moments approach being more appropriate for quantities (income, area, number of employees) and the scales or ranges more suitable for opinions or judgements.

The third sets out to identify the causality attributed by the actor to the program (α_{ij}), in percentages, for the change observed ($\Delta(a_{ij})$). The change observed is the redundant verifier of causality, as stated before. Based on a simple representation for each question, the impact on the indicator j observed by an individual i is given as follows:

$$g_{ij} | \Delta(a_{ij}) | \alpha_{ij} = I_{ij} \tag{1}$$

So, the average impact given by an indicator over all n respondents, I_j is given by:

$$I_j^+ = \frac{1}{n^+} \sum_{i=1}^n \delta_{ij}^+ \Delta | \alpha_{ij} | \alpha_{ij} \quad (2)$$

It is important to note that change is measured as a module in order to differentiate positive from negative impacts, allowing stratification in the analysis. Thus additions of both individuals and themes or other characteristics can differentiate these effects and point to the need for specific analysis of each indicator.

The stratified approach for the positive changes can be given by:

$$I_j = \frac{1}{n} \sum_{i=1}^n I_{ij} \quad (3)$$

where n_+ is the total number of respondents that indicated positive changes, and if $g_{ij} = +1$ and 0 otherwise. The same applies for the negative part, changing positive to negative signs.

Economic surplus approach

In the evaluations of the economic impact of agricultural innovations generated by INCAGRO, a variant of the concept of economic surplus was used to calculate benefits; a different hypothesis from those used by most of the other studies carried out on the basis of this method (Alston *et al.*, 2001) was adopted as to the elasticity of the offer and the demand. This hypothesis, which was initially adopted by Tosterud *et al.* (1973) and later by Kislev and Hoffman (1978), presents two variations regarding the elasticity of the offer, depending on the type of impact posed by the technological innovation:

- Increase in production (yields or expansion of the area) — perfectly elastic curve of demand (D) and a vertical curve of offer (S); and
- Reduction in costs — horizontal curve of offer and vertical demand. This variant of the economic surplus is also used at Embrapa, in Brazil (Avila *et al.*, 2005).

The economic surplus of INCAGRO was estimated in terms of the additional net income generated by the agricultural innovations induced by the program. This additional income was measured at the producer level, meaning how the producers benefited from INCAGRO, estimating their net economic benefit by comparing the income situation before and after the INCAGRO support. The surveys applied to a sample of the producers allowed the authors construct a flux of economic benefits for nine years (baseline – 2007, five years before and a three-year projection). In case of the costs, the authors used all the costs of the 44 INCAGRO projects during the 2002–09 period, including the

counterpart of the producers based on the budget of the projects when they were submitted to INCAGRO. The projects included in the sample used in B/C analysis were responsible for 45.47% of the investment of INCAGRO in the program as a whole.

The cost–benefit analysis was developed using the net economic benefit flux generated by INCAGRO, meaning the additional income that, according to the producers, was due to INCAGRO.

To analyze the profitability of the INCAGRO investments three indicators were used:

- The internal rate of return (IRR);
- The benefit–cost ratio (B/C); and
- The net present value (NPV).

A discount rate of 6% was used for the NPV analysis.

Finally, the costs of the program were deflated to 2008 values, based on the inflation indexes estimated by the National Institute of Statistics (INEI).

Complementary methodological issues

A specific questionnaire should be designed in accordance with the importance of the themes and dimensions for each type of relevant actor. The contents will depend on the role played in the different dimensions of the program by the actor in question. Each questionnaire should be suitable in style and language, so that the questions addressed to the actor concerned appropriately contextualize his or her perceptions. All questionnaires must be validated by the stakeholders and shareholders.

The final design of the questionnaire will depend on the data collection strategy. A variety of different means can be used to collect data, including both face-to-face interviews and electronic surveys or polls. The choice of strategy will depend on the resources available for data collection and the actor's familiarity with the media and tools used in this activity.

Sample design is determined by the criterion of minimizing standard error and maximizing the representativity of the various contexts, dimensions, actors and scenarios involved in the ST&I program being evaluated, that could configure strata or sampling clusters, depending on the design. It is also important to take into account the trade-off between feasibility and statistical significance.

Once the data have been collected, they are filtered for consistency and practical significance. The valid data are then used to build a database for analysis by theme, dimension and actor. Univariate or multivariate analysis is chosen according to the nature of the indicator and/or hypotheses, to ensure that the causal measures are as representative as possible.

Application and results

This section presents an example of multidimensional evaluation as applied to INCAGRO. Some of the findings are discussed for the sake of illustration.

Application of the decomposition method took place at a workshop held in Lima in April 2008 at the headquarters of INCAGRO. Program directors, technicians and managers of the various decentralized units in the regions of Peru took part. A workshop specifically designed to enable INCAGRO shareholders to define the overall objective based on the directives established for the program by the government and World Bank had produced the following:

Evaluate the impacts of the program using solid and reliable evidence that enables the dimensions and significance of the changes to which INCAGRO has contributed through its interventions to be identified and measured.

The goal of the evaluation exercise was to identify the program's impacts by producing information of sufficient quality for three concomitant purposes: learning, planning, and accountability. To determine themes and indicators, the multidimensional effects of the activities promoted by INCAGRO on Peruvian agriculture were discussed and prioritized. Themes were determined on the basis of a synthesis between the objectives of the program, the goals of the evaluation exercise, and prioritization of the guidelines laid down by the agencies responsible for executing and financing the program.⁶

The first step in determining indicators is to divide the themes into sub-themes to facilitate identification and organization of the issues that are truly important for program evaluation. For example, if introducing innovations is one of the program objectives and strengthening the culture of innovation is a goal of the financing agency, it can be determined that a relevant theme is changing the propensity to innovate among program beneficiaries (agricultural producers and extensionists, in the case of INCAGRO). The second step is to decide what is truly relevant to evaluating the propensity to innovate. In the present case, this led to defining the following

sub-themes: interest in adopting new technologies, and willingness to pay for new technologies (these are examples of a larger set of sub-themes).⁷

The indicators are constructed on the basis of the sub-themes. They may be quantitative, qualitative or a mixture of both, and they should be designed to capture a large set of changes derived from R&D activities in a more systematic manner than in the linear innovation model. From this perspective, indicators are identified based on:

1. Input variables (financial resources, human resources);
2. Output variables (scientific production, direct and indirect technological results, technology transfer);
3. Structural changes (formation of human capital, organizational change, networking); and
4. Causality variables (extent of program's contribution) (Furtado *et al*, 2006; Roessner, 2000).

The definition of indicators is an exhaustive task and should not lose sight of or confuse the program and evaluation objectives. The next step, based on all the previous information and an analysis of program management via in-depth interviews with program managers and technicians, is to define the number of actors and the structure of the questionnaires.

Note that there are three different moves. In the first instance we made a motion synthesis, on analyzing the set of themes and objectives to be evaluated. Second, in the desegregation movement, leading indicators were constructed to be evaluated, from the result obtained in the first movement. Although it might appear that we have reached the same point of departure, because the goals were built from the themes and sub-themes, and then indicators were constructed that represented the goals, in fact the result of the second movement is a third movement, where we identified the best way to measure the indicators using the knowledge of the stakeholders and experts in the subject assessed.

The questionnaires were then produced and distributed. In the present case distribution was by email, as determined by the evaluators and top management in defining the evaluation exercise, and the tool used to design the questionnaires was Adobe Acrobat 9, plus LiveCycle Designer.

INCAGRO's indicators and questionnaires were validated by technical staff and senior management. During field validation, however, the evaluators noticed that most actors, with the sole exception of a specific type of actor, did not have internet access, and in many cases were not even familiar with computers. As a result, they printed the questionnaires, and validated the indicators and questionnaire structure personally, knowing in advance that the questions should be designed for a face-to-face survey in cases where internet access does not exist. The final wording of the questions was validated by program technicians, who cleared up semantic problems and

The first step in determining indicators is to divide the themes into sub-themes to facilitate identification and organization of the issues that are truly important for program evaluation

made sure the language was suitable for the actors to be interviewed. Test results were systematized and the final drafts of the questionnaires were produced. Table 1 shows themes and questionnaire structure by type of actor.

It is important to note that within the indicators for a set of themes the respondent should answer questions only on those for which he or she can identify program impacts. For this reason, for example, researchers were not asked about the quality of training (they do not work in that part of the program).

The sample design used was a two-stage design with unequal sampling probabilities. The primary sampling unit was the project. There was a total of 216 projects located in different regions across Peru, and strata definition took into account four sampling criteria defined jointly with INCAGRO's technicians and senior managers:

1. Program phase (the program has had two phases since its inception);
2. Type of service (extension, training, or research);
3. Natural region (coast, jungle, or highlands); and
4. Degree of success (project achieved goal on time, project underperformed from management standpoint, project performed better than expected).

All told, 49 projects distributed in 39 strata were selected.⁸ The number of strata should have been 54 (three regions, two phases, three types of service and three degrees of success), but there were no projects in some strata and in the case of training services, so it was decided not to differentiate between regions (i.e. to ignore the 'regional division' criterion). Projects were drawn using unequal sampling probabilities proportional to the number of projects in each stratum. Then, in the second stage, individuals were drawn from each project with unequal probability, proportional to the size of the project within its stratum in terms of the number of people (researchers or producers) and projects. The sample of individuals

was calculated for a margin of error of plus or minus 3 percentage points at 95% confidence.

Some sampling restrictions were applied because of operational concerns. In the case of researchers, it was decided to interview all 78 of the researchers responsible for the 49 projects in question. Because the researchers had internet access and were computer-literate, questionnaires were emailed to them. The response rate in this case was 86%. However, for the other actors, it was decided jointly with the INCAGRO team to limit the sample size to roughly 900, a feasible sample size for the project structure. Based on this limit and considering the distribution of the selected projects, the final sample size was 904, with the following breakdown: 739 producers, 44 managers, 31 extension trainers and 90 extension workers. The response rate was 100%.

The data were organized into banks by type of actor for the purposes of analyzing consistency and practical significance. In all cases almost all the data passed the consistency test and no problems were observed with practical significance. INCAGRO is to be congratulated on its efficient data collection.

To illustrate the organization of the findings and how they can be compared by theme and type of actor, examples were prepared showing selected results for the themes 'Technological development', 'Propensity to innovate', and 'Market for services'.

Table 2 presents the results for some of the indicators used to capture the changes observed in propensity to innovate due to the program. Equation (1) was used to calculate impact *I* by type of actor.

A comparison of the responses from the various types of actors on this theme shows that INCAGRO's impact was about 30%, which means that the influence of the program on the propensity to adopt new technology attributed by the different actors, the α_{ij} , exceeded, on average, 50%.

Table 3 shows the results for some of the indicators used to capture the changes observed in capabilities and competencies for technological

Table 1. Themes by type of actor

Theme	Actor	Producers	Managers	Extension Trainers ^a	Extension Workers ^a	Researchers
Regional innovation systems			x	x		x
Technological development		x	x	x	x	x
Propensity to innovate		x	x		x	
Quality of training					x	
Competitiveness		x	x	x	x	x
Social development		x	x	x	x	x
Environmental development		x	x	x	x	x
Market for services		x	x	x		x

Note: ^a The group of extensionists was divided into two: extension workers trained by the program and extensionists who trained producers, i.e. extension trainers

Table 2. Changes in propensity to innovate

Indicators of propensity to adopt	Managers			Producers			Extension workers		
	P(<i>g</i>)	$\Delta(a)$	<i>I</i>	P(<i>g</i>)	$\Delta(a)$	<i>I</i>	P(<i>g</i>)	$\Delta(a)$	<i>I</i>
Propensity to adopt new technology and knowledge	100	56.7	37.8	93.1	58.9	35.1	100	57	36.5
Propensity to seek capabilities and entities that promote innovation in agriculture	NA	NA	NA	81.6	58.4	34.3	92.1	54	31.9
Willingness to pay to acquire knowledge and capabilities in activities that promote innovation in agriculture and the rural environment	84	49.7	30.1	77.1	51	29.8	96.6	55	34.7

Source: adapted from the original presented in the INCAGRO Final Impact Evaluation Report

Key: P(*g*) = respondents on changes as a percentage of the total sample; $\Delta(a)$ = observed change; *I* = impact of INCAGRO; NA = not applicable

development. Equation (1) was used to calculate impact *I* by type of actor.

Although for most actors the results in Table 3 are similar to those shown in Table 2 as far as the influence of the program is concerned, the change observed by producers falls by half and parameter *I*, which represents the redundant verifier of causality, falls to less than a quarter; hence, the importance of comparing findings from a common perspective in terms of indicators and actors. For example, program managers will have been alerted by these findings to the difficulties experienced by producers in identifying changes in their capabilities and knowledge with regard to activities or processes other than production proper. Thus managers and technicians will have been able to draw the conclusion that this aspect needs to be reinforced in future program operations, enabling producers to acquire new capabilities and knowledge that strengthen the culture of innovation in small farming in Peru.

Table 4 shows the results for some of the indicators used to capture the changes observed in the market for technological services. Table 4 is interesting for two reasons: first, because it presents both demand and supply in the market for technological services; second, because it shows how positive findings are divided from negative findings.

For positive but not negative changes, most actors' perceptions corresponded on the demand and supply side, although for producers the influence of the program was the same on both sides. It is worth noting the importance of parameter α . If the impacts of the program had been aggregated without differentiation, they would have appeared much smaller than they actually were, for example, in terms of the quantity and cost of services, and ultimately the conclusion might have been that the program had no impact at all. Separation of the meaning of changes made clear the relative importance of the observed impacts. Thus in order to explain the values represented by the indicators under analysis it is necessary to revise open-ended questions, if any (open-ended questions ask the respondent to provide practical evidence of impacts), and also revise the practical significance of changes.

Lastly, some multivariate findings are presented as an aid to understanding how integrated analysis of the data is processed. Multiple correspondence analysis (MCA) and cluster analysis⁹ were used to identify clusters that could determine different profiles of producers in the sample, looking for differences in their adoption of innovations. First, the MCA was applied using the variables services received from the program, age, gender, schooling, land tenure, natural region, and participation in other INCAGRO

Table 3. Changes in capabilities and competencies for technological development and innovation

Indicators of change in capabilities and competencies	Managers			Producers			Extension trainers			Extension workers		
	P(<i>g</i>)	$\Delta(a)$	<i>I</i>	P(<i>g</i>)	$\Delta(a)$	<i>I</i>	P(<i>g</i>)	$\Delta(a)$	<i>I</i>	P(<i>g</i>)	$\Delta(a)$	<i>I</i>
Development of new production practices	95.5	58.5	40.1	80.9	32.8	11.2	96.8	58.7	39.4	100	62	38.4
Development of new management practices	88.6	48.8	32.4	70.9	23.2	5.76	96.8	54	39.6	87.6	55	36.3
Development of new forms of association and/or cooperation	81.8	57.4	38.6	73.7	28.6	6.63	83.9	56.9	40.3	67.4	59	34.8

Source: adapted from the original presented in the INCAGRO Final Impact Evaluation Report

Key: P(*g*) = respondents on changes as a percentage of the total sample; $\Delta(a)$ = observed change; *I* = change attributed to INCAGRO

Table 4. Changes in the market for services

Indicators of changes in services	Type of change	Managers			Producers			Extension trainers			Researchers		
		P(/g)	Δ(a)	I	P(/g)	Δ(a)	I	P(/g)	Δ(a)	I	P(/g)	Δ(a)	I
Quantity of services	Increase	91.9	50.4	29.1	95.3	51.8	29.4	96.8	48	28.3	83.6	52.6	32
	Decrease	5.4	60.0	18.0	4.5	62.8	24.7	–	–	–	–	–	–
Diversification of services	Increase	97.2	48.3	26.9	98.3	51.6	28.5	87.1	52	31.7	–	–	–
	Decrease	2.8	10.0	2.0	1.5	44.3	17.7	–	–	–	–	–	–
Quality of services	Increase	79.5	50.1	27.5	42.4	55.6	32.5	96.8	51	29.6	85.1	59.4	37
	Decrease	–	–	–	0.6	33.3	11.1	–	–	–	–	–	–
Cost of services	Increase	54.5	38.3	19.7	95.9	52.0	29.7	80.6	44	26.4	–	–	–
	Decrease	–	–	–	4.1	43.9	22.3	–	–	–	–	–	–

Source: adapted from the original presented in the INCAGRO Final Impact Evaluation Report

Key: P(/g) = respondents on changes as a percentage of the total sample; Δ(a) = observed change; I = change attributed to INCAGRO

projects. Over the coordinates of the first two dimensions generated by MCA — which take into account the similarities among the individuals, locating them as close as they are similar — were then applied a *k*-means clustering algorithm resulting in a dendrogram that evidenced three clusters. Figure 3 shows the main characteristics of the three clusters found in the producer sub-sample.¹⁰

Further, in terms of economics, C1 seems to be the wealthiest one, presenting major crop areas and income in the baseline, against the other two clusters, and also having better education and previous contact with R&D activities, mainly because of previous participation in INCAGRO.

Impacts of projects financed by INCAGRO were present in all the clusters found. For the sake of illustration, some of the findings are briefly described below, focusing on breadth and intensity of impact, level of schooling, land tenure, and type of service in which the co-financed project participated in INCAGRO for each of the clusters shown in Figure 3.

In general, for producers in C1 the impacts were generally more intense but less broad than for C2 and C3 (i.e. a smaller number of C1 producers were

affected with greater intensity). The greater intensity of impacts on C1 was probably due to the presence of more producers from the coastal and jungle regions who had previously participated in projects co-financed by INCAGRO. Given the characteristics of these producers, they may have had a slightly more pronounced background of participation in innovation systems than producers in other clusters. Another factor worth taking into account is that most of the projects involving adaptive research were in C1, suggesting that these projects presented more intense impacts than extension projects.

For producers in C2 the impacts were generally broader than for C1 but somewhat less so than for C3, and less intense than for C1 and C3. Despite their greater institutional stability (title to land), they displayed less propensity to innovate than C1 or C3. This was probably due to the fact that they were older and had a very low level of schooling (mostly incomplete primary), making them less open even to changes with a broad impact. Support from INCAGRO was most frequent in extension services, possibly because this cluster participated least in innovation systems.

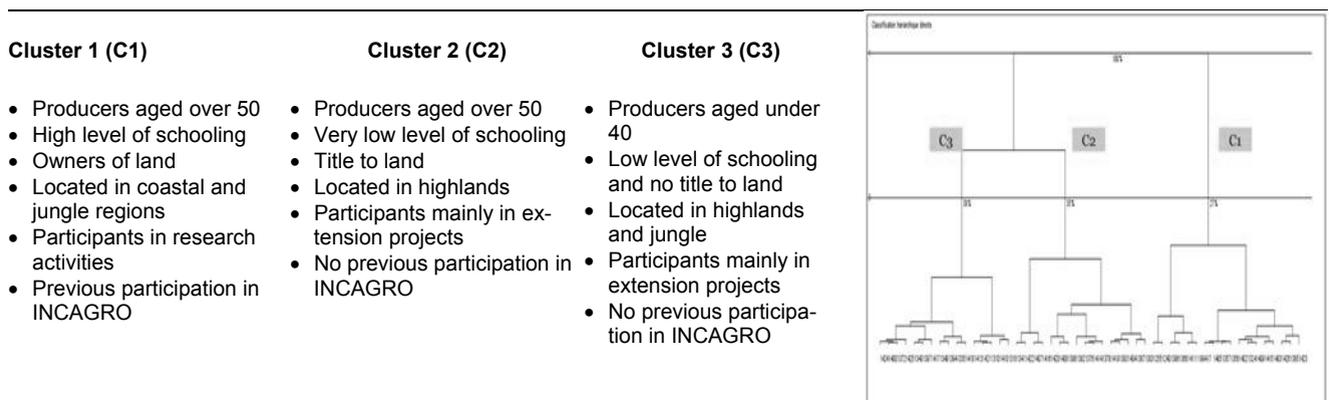


Figure 3. Clusters located in producer sub-sample and resulted dendrogram of cluster algorithm

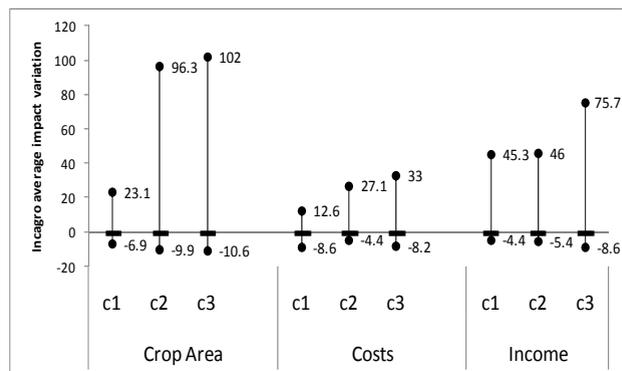
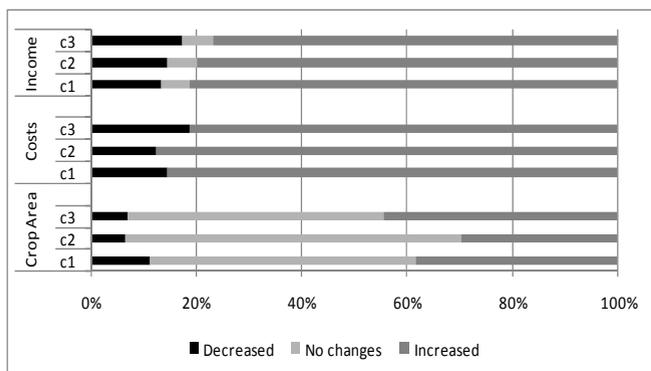


Figure 4. (a) Percentage for each response category of g, and (b) estimated variations attributed to INCAGRO for 'crop area', 'production costs' and 'income'

For producers in C3 the impacts were the broadest, in the sense that more producers were affected by changes, albeit with less intensity than for C1. This profile represents producers with very low levels of participation in innovation systems and institutional instability due to the conditions of land tenure (no title). Generally speaking the impacts were more intense than for C2, probably because C3 producers were younger and had higher levels of schooling than C2. The impacts mostly came from extension activities and less frequently from research.

To detail these differences among the clusters of producers, Figures 4 and 5 present the impact results from two specific dimensions: 'Economic' and 'Propensity to innovation'. For the economic dimension, variation in 'Crop area', 'Production costs' and 'Incomes' are presented, and for innovation, already presented in Table 2, variation in 'Interest for association with Researchers', 'Willing to pay for innovation' and 'Interest for association with other producers'. These are good examples because the first dimension is composed of objective indicators, whereas the second one is subjective perceptions, behaviors. Besides the fact that these two dimensions showed more intense differences, the diversity of the indicators' nature can show how this methodology allows unifying the analysis attributing the same rationale of causality regardless of the type of indicator under analysis.

Effects in the economic dimension did not follow the general pattern cited above, where C1 presented more intense variations, but less broad than the others clusters. 'Crop area' was the indicator with less indication of any variation, having around 50% or responses in the 'No changes' category in all three clusters, but showing a major increasing variation in C3. However, with the monetary indicators 'Costs' and 'Income', almost 80% of respondents answered for an increasing variation, with minor differences among clusters.

Looking at the impacts (Figure 4b) the differences mainly between C1 and the others are remarkable. In 'Crop area', C1 presented an impact attributed to INCAGRO of 29%, while in C2 and C3 this impact was 96% and 102% respectively. As said before, C1 is the wealthier group, having bigger crop areas in the baseline, so that one would expect that a minor impact when compared to the other two clusters. Regarding 'Production costs', differences are less prominent, but still significant, following the same pattern of area, where C2 and C3 presented similar results in opposition to C1. Again, the fact that these two clusters were incorporating new technologies and methods, while C1 was already using them, might explain this difference. In the 'Income' indicator, the pattern changed a little bit. C1 and C2 have the same result (around 45%) of increase, while

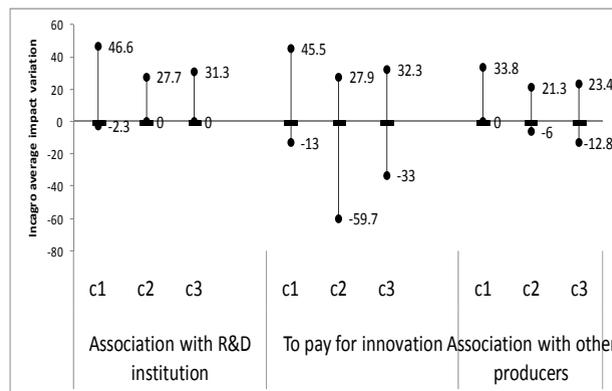
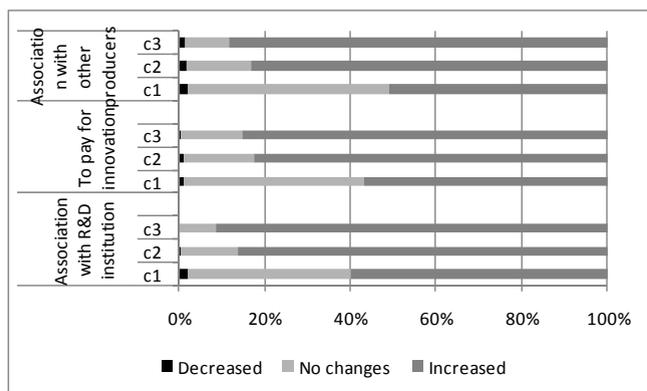


Figure 5. (a) Percentage for each response category of g, and (b) estimated variations attributed to INCAGRO for 'interest in association with R&D institution', 'willing to pay for innovation' and 'interest in association with other producers'

C3 show a 75% increase in 'Income', a great change in terms of economics impact. It is noticeable that negative variations are quite similar among the three clusters, and always have the smallest proportion of responses, and also that nobody responded to 'No changes' in the 'Costs' indicator.

Regarding 'Propensity to innovation', few respondents reported a negative variation; however, there is a noticeable pattern along the three clusters in terms of reporting 'No changes', and consequently in the positive variation. In C1, 40–45% reported no changes, while in C2 this level is 20% and around 15% for C3, such that this gradient for the positive change is inverted. As said before for the general results, in C1 the impacts were generally more intense but less broad than for C2 and C3. The propensity for innovation increased more in the group that already had had contact with this kind of institution, through previous participation in INCAGRO. Negative impacts shown in Figure 5, despite the high intensity, might be ignored since too few respondents pointed to this type of variation.

Finally, Table 5 presents the results of the three indicators from the cost–benefit analysis (IRR, B/C and NPV) by the INCAGRO program as a whole and by Peruvian regions benefited by INCAGRO projects. The results show that the INCAGRO program was profitable, compensating for the funds allocated by Peruvian society, relative to other investments. The agricultural innovations induced by INCAGRO generated an economic surplus that largely compensated the investments at national level (23%) and by natural region (15–28.5%).

Conclusions

The conclusions to this article are divided into two parts: the first relates to the evaluation method, and the second to the significance of the findings for policy to foster innovation in small farming in Peru.

With regard to the method, the approach used in this case proved highly suitable to the object of the evaluation exercise. INCAGRO was planned to use a methodology that could quantify causality, taking into consideration the inherent difficulty and high cost of finding 'counterfactual' control groups for the diversity of situations found in the object to be evaluated.

Table 5. Profitability of INCAGRO investments by Peruvian natural region

Indicators	Costa	Selva	Sierra	INCAGRO
IRR – 15 years	28.5%	25.2%	15.0%	23.0%
NPV (new base of 2008, a 6% discount rate)	3,985,469	11,069,224	4,806,023	9,272,610
B/C ratio	2.60	2.58	1.42	1.29

Attributing causality in the evaluation of socially constructed processes including knowledge production and appropriation poses methodological challenges that have not yet been resolved. But this method is a framework for assessing objects such as ST&I when the program to be evaluated involves social groups or organizations and, in these cases, to measure concentric sets of indicators that as an ensemble (and via statistical cross-tabulation of variables) can be used to confirm previously defined hypotheses regarding specific phenomena. Thus causality verifiers must be used in situations where it is hard to use experimental methods.

Measuring additionality from a baseline (the difference between an indicator measured at time t_1 and the same indicator measured at time t_0) is the most common method for impact assessment. When control group behavior is highly imprecise, it is fundamental to introduce some causality verifier such as the one used in this evaluation, represented by the factor I .

In any event, the methodology described in this article proved capable of pinpointing positive impacts of INCAGRO and of distinguishing among them according to the various producer profiles. The findings were highly consistent with what has generally been described in the literature. For example, the most significant impacts occurred for producers with incipient participation in innovation systems, relatively higher levels of schooling, and location in regions with better conditions for access to knowledge and production techniques.

In sum, the evaluation study identified three types of situation in Peruvian small farming to be considered by INCAGRO:

1. Producers who require supplementary measures to strengthen their participation in innovation systems;
2. Producers with a certain propensity to innovate but lacking a minimum of technical and managerial knowledge to enter an innovation system definitively; and
3. Producers who need access not only to knowledge but also to capital, especially land, in order to join an innovation system.

All in all, the methodology proved adequate to generate relevant information for use by program

The methodology described in this article proved capable of pinpointing positive impacts of INCAGRO and of distinguishing among them according to the various producer profiles

decision-makers and was recognized and validated by the stakeholders.

Notes

1. The results in this article were extracted from the Final Report presented on 23 April 2009, by the Board of INCAGRO in Lima.
2. The amounts cited were obtained from <<http://www.incagro.gov.pe>>, last accessed 25 July 2009.
3. In a social context, innovation does not exist before appropriation but is merely a new technology or way of performing a given procedure. Innovation proper materializes when the new technology or know-how is appropriated by the user in a social context generally or in the market in particular. Thus when based on the concept of social appropriation used here, innovation takes place only when the potential users effect a positive selection.
4. Non-ergodic refers to processes in which probabilities change randomly from start to finish (Davidson, 1996).
5. The expression $\Delta(a_i)$ can also be represented by a quotient. This is especially useful for indicators of magnitudes that differ in units of measurement, as is the case with agricultural yields, for example. The problem is circumvented by converting the module into an increase or decrease in line with the stated direction.
6. Generation of themes involved the agencies responsible for execution and financing as well as program managers. Thus the field researcher had to produce a synthesis of the various interests concerned and structure the questionnaire so as to prioritize the items accordingly.
7. A sub-theme can emerge from discussion and does not have to be expressly included among the program objectives or ST&I agency guidelines. For this reason it is important to introduce during this step all the sub-themes that occur directly or indirectly, depending on their nature, and to avoid excluding any of them from the evaluation, given the importance of their contribution to the structure of ST&I systems.
8. This is why the number of strata was eventually reduced to 34.
9. Clusters were defined in terms of services received from the program, age, gender, schooling, land tenure, natural region, and participation in other INCAGRO projects.
10. It is important to note that managers and technicians not only validated the clusters found but also confirmed the hypotheses developed on the basis of the evidence presented.

References

- Adrian, A M, S H Norwood and P L Mask 2005. Producers' perceptions and attitudes toward precision agriculture technologies. *Computers and Electronics in Agriculture*, **48**, 256–271.
- Alston, M J, C Chan-Kang, M C Marra, P G Pardey and T J Wyatt 2001. A Meta-Analysis of Rates of Return to Agricultural R&D: Ex Pede Herculeum Evaluation and Priority Setting. Washington: IFPRI. 148pp. (Research Report, 113).
- Avila, A F D, M Magalhães, G Vedovotto, L J M Irias and G S Rodrigues 2005. Avaliação dos impactos das tecnologias geradas pela Embrapa. *Revista de Política Agrícola*, Brasília, **15**(4), October/November/December, 86–101.
- Bin, A and S L M Salles-Filho 2008. Science, technology and innovation management: specificities and conceptual premises. Presented at the 12th International Schumpeter Society Conference. Rio de Janeiro, 2–4 July.
- Blazy, J-M, M Dorel, F Salomon, H Ozier-Lafontane, J Wery and P Tixier 2009. Model-based assesment of technological innovation in banana cropping systems contextualized by farm types in Guadeloupe. *European Journal of Agronomy*, 10–19. Challenges and opportunities (No. **87**, pp. 47–56). San Francisco, CA: Jossey-Bass.
- Davidson, P 1996. Reality and economic theory. *Journal of Post Keynesian Economics*, **18**(4), Summer.
- Freeman, C 1998. Introduction. In *Technical Change and Economic Theory*, eds G Dosi, C Freeman, R Nelson and G Silverberg. London: Pinter.
- Furtado, A T, A Bin, M B Bonacelli, S R Paulino, M A Miglino and P F D Castro 2006. Evaluation of the Results and Impacts of a Social-Oriented Technological Program in Brazil: the Case of the Prosab (Sanitation Program). Annual Conference Technology Transfer Society Next Generation Innovation: New Approaches and Policy Designs, Georgia Institute of Technology, Atlanta, USA, 27–29 September.
- Goméz Sal, A and A González García 2007. A comprehensive assessment of multifunctional agricultural land-use system in Spain using a multi-dimensional evaluative model. *Agriculture, Ecosystems and Environment*, **120**, 82–91.
- Hall, A, V R Sulaiman, N Clark and B Yogana 2003. From measuring impact to learning institutional lesson: an innovation system perspective on improving the management of international agriculture research. *Agricultural Systems*, **78**, 213–241.
- Jassen, S and M K van Ittersum 2007. Assessing farm innovations and responses to policies: a review of bio-economic farm models. *Agricultural Systems*, **94**, 622–636.
- Kislev, Y and M Hoffman 1978. Research and productivity in wheat in Israel. *Development Studies*, **4**, 166–181.
- Kline, S and N Rosenberg 1986. An overview of innovation. In *The Positive Sum Strategy*, eds R Landau and N Rosenberg. Washington DC: National Academy Press.
- Nelson, R R 1998 Institutions supporting technical change in the United States. In *Technical Change and Economic Theory*, eds G Dosi, C Freeman, R Nelson and G Silverberg. London: Pinter.
- Ny, H, J P Macdonald, G Broman, R Yamamoto and K-H Robert 2006. Sustainability constraints as system boundaries: an approach to making live-cycle management strategic. *Journal of Industrial Ecology*, **10**(1–2), 61–77.
- Payraudeau, S and H M van der Werf 2005. Environmental impact assessment for farming region: a review of methods. *Agriculture, Ecosystems and Environment*, 1–19.
- Roessner, D 2000. Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation*, **9**(2), August, 125–132.
- Rogers, J and B Bozeman 2001. Knowledge value alliances: an alternative to R&D project evaluation. *Science, Technology and Human Values*, **26**(1), 23–55.
- Rogers, P 2000. Causal models in program theory evaluation. In *Program Theory in Evaluation: Challenges and Opportunities, New Directions in Program Evaluation*, eds P Rogers, T Hacsí, A Petrosino and T Huebner, **87**, 47–55. San Francisco, CA: Jossey-Bass.
- Rosenberg, N 1982. *Inside Black Box: Technology and Economics*. Cambridge University Press.
- Rosenberg, 1992 ?????
- Salles-Filho, S L M, M B M Bonacelli, M Zackiewicz, P F D Castro and A Bin 2007. Desenvolvimento e Aplicação de Metodologia de Avaliação de Programas de Fomento a C,T&I: o Método de Decomposição. In: XII Seminario Latino-Iberoamericano de Gestión Tecnológica. Buenos Aires: ALTEC.
- Tosterud, R J, J C Gilson, A E Hannah and B R Stefansson 1973. Benefit cost evaluation of research relating to the development of Selkirk wheat and target rapeseed. In *Symposium on Agricultural Research, Proceedings*, 1, s.l. University of Manitoba, **1**, 149–199 (occasional service).
- Weitzman, B C, D Silver and K Dillman 2002. Integrating a comparison group design into a theory of change evaluation: the case of the urban health initiative. *American Journal of Evaluation*, **23**, 371–385.