3. Conducting EpIA

This chapter identifies the principles of good practice for conducting epIAs. These principles refer to expectations on the availability of information that conditions transparency and analytical rigor. The presentation is tailored to Stage I-type disaggregated economic rate of return assessments. However, where appropriate, expectations about the availability of information pertinent to multi-dimensional impact assessments (especially of the Stage II type) are also discussed.

The discussion is organized roughly in the chronological order of activities that take place in a Stage I epIA. The selection of the research-derived technology or policy for epIA (in a well-defined institutional setting) is the subject of the first section.

The second section considers how adoption is a necessary condition for the impact of technology-oriented epIAs. The same section then looks at policy-oriented epIAs and discusses how documenting uptake and influence are key ingredients for making a persuasive case for attributing behavioral change to research. This section examines in depth the adoption and diffusion of technologies and the uptake and influence of policies. Reliable documentation on these important topics is a prerequisite for persuasive epIA.

The third section describes the counterfactual scenario, which is the center of attention in the general literature on impact assessment, along with its good practice application in epIAs of agricultural research.

The estimation of benefits and costs and the presentation of results are discussed in the fourth and fifth sections respectively. The fifth section ends with extensive coverage of how epIAs can inform about a program's effect on alleviating poverty, which is the apex impact in Figure 2.1 and one of the MDGs that attracts the most interest from donors (Raitzer and Winkel, 2005).

Selecting the research-related technology or policy for epIA

The selection of research-related technologies for epIA is constrained by funding and by the number of perceived success stories at any point in time. Some research-related technologies that build on previous work may have special project (restricted) funds for epIA. But, in general, funding for epIA is scarce. A typical scenario is that economic rate of return assessments have to be supported by unrestricted core funds, whilst special funding usually needs to be solicited for more costly multi-dimensional impact assessments. A scarcity of funds, particularly for operational expenses, argues for greater scrutiny in the selection of research-derived technologies and policies for epIA.

The previous chapter endorsed the success story approach for general accountability purposes. To qualify as a success story, a research-related technology should be characterized by a level of adoption that has the potential to generate sufficient benefits to at least cover the costs of the research on technology generation and transfer. Therefore, evidence of adoption is the primary consideration in determining the suitability of candidate technologies and policies for impact assessment. At the selection stage, adoption evidence tends to be largely anecdotal and may be contested. Candidate technologies with anecdotal evidence should be screened for their desirability and feasibility for epIA.

The first step in an illustrative protocol for screening candidate technologies on the grounds of their desirability for epIA is described in Figure 3.1. This figure sets out three minimal thresholds that should be passed by the prospective epIA technology before advancing to the next step, which consists of a desirability scorecard.

Component novelty

That a technology should be novel to prospective adopters appears so self-evident as to seem trivial. Although the criterion of

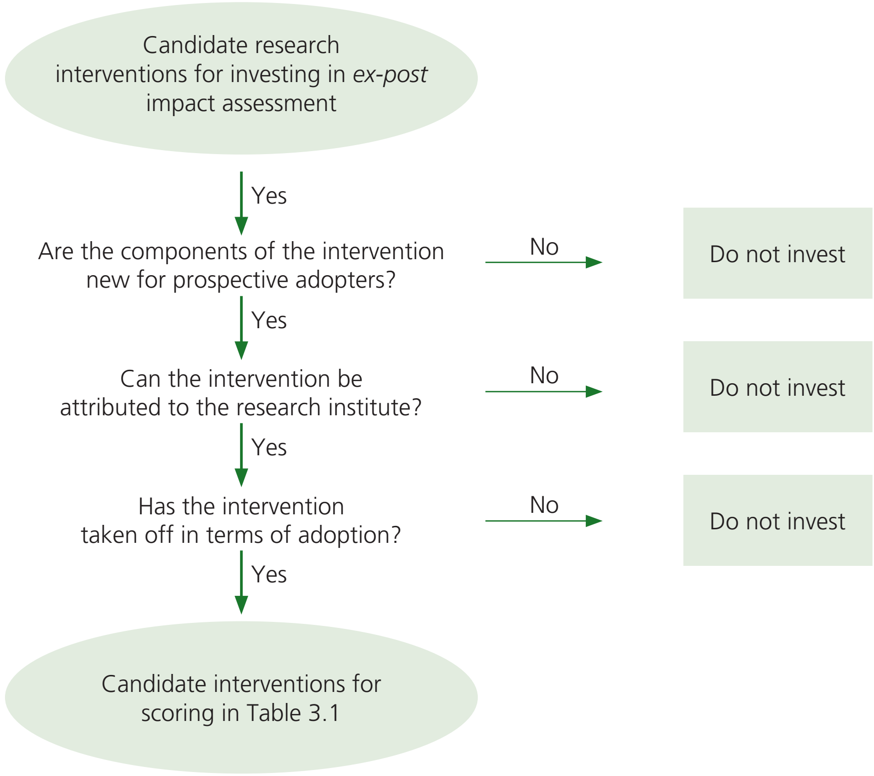

Figure 3.1. Minimum thresholds for selecting interventions for *ex post* impact assessment.

newness seems to result in a very porous sieve, some components of crop management and NRM technologies may not pass the newness criteria. Novelty is a variation on the theme of the counterfactual scenario and focuses on what potential adopters were doing in the 'without the technology' research position. Have users adopted or would they in the near future have adopted some of the components of the candidate technology in the absence of the research? Also, it is relevant to ask: if some components were available prior to the research and adoption did not take place, then why hadn't adoption occurred then? Was the research responsible for influencing the underlying conditions that now improve the likelihood of adoption? Novelty is about transparency and the role of the research in generating the technology.

Emphasis needs to be placed on research results that have been translated into technological components or options. For epIA, the technology should be broken down into its new components. Component definition is especially warranted in 'integrated' crop management, pest management, and resource management technologies that may be very location-specific. Scientists who cite preliminary evidence about adoption should be able to say what components have been adopted by farmers in a few of the main locations. The technology for selection should then be recast in terms of those options so that the word 'integrated' is shorn of its fuzziness.

Policy-oriented research should be held to the same standard of specificity as technology-oriented research. Candidate policies should be characterized in detail, not only by the recommendations of the policy research, but also by the changed policy decisions on which they supposedly had an influence.

Institutional attribution

The second filter for judging an intervention's suitability for epIA is institutional attribution. This refers to the role played by the institution in developing the technology (Figure 3.1). Can a persuasive case be made that the research institute played an important role in making the technology available for adoption? In other words, if the research did not exist, would the potential beneficiaries have accessed the technology? The case for attribution is stronger for technologies or policies that feature longer-term strategic research and weaker for those located at the shorter-term adaptive end of the research-for-development continuum. This is largely because the former require a sustained commitment accompanied by more years of effort compared to the latter. Like novelty, institutional attribution is associated with the counterfactual scenario of what would have happened if the research had not occurred. In principle, epIAs that score well on institutional attribution of research should also score well on the persuasiveness of the counterfactual.

Presumably, the majority of candidate technologies receive a passing score on institutional attribution[6]. But this may not apply to candidate policies where attribution is a more formidable hurdle, because the seeds of policy change may take root and influence conventional wisdom in such a way that assigning attribution to specific lines of inquiry becomes a more difficult undertaking.

6 Two examples of economic rate of return assessments that define and just pass minimal thresholds of international and national institutional attribution are Bofu et al. (1996) and McSween et al. (2006), respectively.

The uptake of research interventions

The third threshold in Figure 3.1 is concerned with whether the candidate technology has 'taken off' in terms of adoption or, in the case of impact assessment on policy-oriented research, on the strength of claims that the research has influenced the adoption decision. As discussed in the previous section, 'older' successful technologies are usually characterized by wider diffusion and broader impact. The potential for learning about their impact is greater than for technologies that have good prospects, but are only in the initial stages of being adopted. The risk when selecting technologies in the early stages of diffusion is the uncertainty of their future benefits (see Box 3.1). But even epIAs of technologies that do not take off have value for learning. For example, the epIAs described in Box 3.1 – of a technology promoted by the CIP – contributed to identifying where and when a more generic version of this technology could 'work' over space and time (Chilver et al., 1999).

Scoring desirability

Candidate interventions that pass the minimal standards in Figure 3.1 can be scored in Table 3.1, which specifies six

Box 3.1. Technology takeoff and the CIP economic rate of return assessments

From the early 1990s to the early 2000s, 14 economic rate of return assessments on diverse center-related technologies were conducted at CIP. Several technologies were 'done deeds' in the sense that practitioners had firm estimates of the maximum area over which they had been adopted. For those in early adoption, diffusion curves had to be projected.

With the benefit of hindsight it can be seen that a few assessments resulted in over-estimates of adoption, and at least one resulted in significant underestimates. Of the 14 technologies, one did not take off. At the time of that epIA, the technology, an innovative horticultural seed system, had been sown very intensively by several hundred farmers on about 200 hectares. Chilver et al. (1997) conservatively estimated that by 2015 this new propagation system would cover about 7,500 hectares. Because of a change in government policy that favored the import of a competing form of seed and because of weaknesses inherent in the technology, the area under the new seed system stagnated and declined. The technology did not take off. With the exception of an integrated pest management technology, which did not spread beyond its pilot recommendation domain, the cases that overestimated adoption and subsequent benefits were confined to this one research area.

Table 3.1. An illustrative checklist for scoring the desirability of economic rate of return assessment (for epIAs).

Characteristic	Description	Valuation (indicative weighting)
Size of benefit relative to total portfolio costs:	Large Medium Small	3 2 1
Prospects for documenting impacts further along the impact pathway:	Bright Dim	2 1
Uniqueness of epIA:	Not many similar types of epIAs found in the literature Not many similar epIAs conducted in the CGIAR Not many similar types of epIAs conducted in the center's program Not a unique epIA in the center's program	4 3 2 1
Programmatic history and epIA:	No epIA conducted in the program EpIA conducted but benefits not sufficient to cover costs of the program EpIA conducted and benefits substantially exceed the cost of the program	3 2 1
Location:	In a high-priority region Not in a high-priority region	2 1
Geographic coverage:	Global assessment (i.e., spread over several continents) Multiple countries (~ 2–5) assessment Multi-locations (regions) within single country assessment Single location within single country assessment	4 3 2 1
Incidence of poverty in the recommendation domain	Higher than the global average for developing countries Lower than the global average for developing countries	2 1

attributes of desirability as an example. The choice of these attributes is quite arbitrary and scoring is only indicative. For example, preference is given to technologies that are characterized by large prospective benefits, have bright prospects for documenting consequences further along the impact pathway, are novel, are located in programs where epIA has not yet taken place, and are sited in geographic poverty traps. For international agricultural research, priority is also given to the location of benefits; for example, across more countries and in priority regions. Larger, multi-country benefits should brighten the prospects for documenting consequences (other than farm-level effects) as these two attributes are likely to be positively correlated and will reinforce the benefits accruing from international public goods.

A desirability scorecard for multi-dimensional impact assessments should be similar to one for economic rate of return assessments, with one exception. Multi-dimensional impact assessments are most applicable to research areas and programs for which one or more economic rate of return assessments have been carried out. The valuation on programmatic history should be the opposite of that illustrated in Table 3.1, since large benefits previously established in Stage I epIAs point to brighter prospects for the documentation of impact in Stage II epIAs.

Feasibility of generating reliable estimates

Investment in the top ranking technologies and policies on desirability grounds needs to be balanced by the feasibility of generating reliable estimates of impact. Few if any of the desirable candidate technologies will have data on adoption, but most should have experimental information on component treatment effects. Those characterized by extensive on-farm research and baseline surveys will be better placed to generate reliable estimates for epIA. The quality of national rural household survey data, which could be used to support an epIA, also varies from country to country. In general, feasibility should be the deciding factor for selecting top ranking technologies that receive roughly the same scores on desirability as per the criteria given in Table 3.1. The feasibility of generating reliable estimates looms much larger for impact studies characterized by intensive field work, which often span several years prior to the initiation of the assessment, such as the studies reported in Adato and Meinzen-Dick (2007)[7].

Institutionalizing the selection process

Selection of topics on which to carry out epIAs should be done routinely every year. The lists of candidate and screened technologies and policies should be dynamic, but the content may change little from year to year because agricultural research is 'slow magic' (Pardey and Beintema, 2001). Over time, younger technologies and policies should make their way onto these lists.

The provision of information on the candidate lists and the results of selection are valuable because they provide information about what has potential and is on the horizon for epIA. For example, in international agricultural research, comparative knowledge of the prospects across centers and Challenge Programs for epIA should strengthen the assessment of impact in the CGIAR as a whole. Such knowledge could point to opportunities for packaging thematic impact assessment across several centers in proposals for donor funding. Knowledge of candidate technologies and policies for epIA is also useful to see emerging gaps in impact assessment where extra effort may be needed to improve the coverage of outcomes further down the impact pathway or in generic research areas. The presence of continuing gaps can also give early warning about themes where impacts have not materialized as expected and will signal the need for corrective action by system management and donors.

Good practice 3.1. The process for selecting candidate technologies and policies for epIA should be open to all scientists in a national or international agricultural research institute. Likewise, the results of the selection should be communicated to all of these scientists.

Good practice 3.2. Candidate technologies and policies for epIA should be screened for minimal performance on the criteria such as component novelty, institutional attribution, and uptake (takeoff).

Good practice 3.3. Screened candidate technologies should be ranked on desirability criteria subject to a feasibility assessment.

Good practice 3.4. The lists of candidate and screened technologies and policies with a brief description of the results of the selection process should be documented annually (or periodically) and become a part of the institutional memory of the international or national research organization.

Key references

Pardey P.G., Alston J.M., Chan-Kang C., Magalhaes E.C., and Vosti S.A. 2006. International and institutional R&D spillovers: Attribution of benefits among sources for Brazil's new crop varieties. • State-of-the-art research on institutional attribution.

Traxler G. and Byerlee D. 1992. Crop Management Research and Extension: The Products and their Impact on Productivity. • Incisive evaluation of component novelty from the perspective of recommendations made to farmers.

Estimating adoption of the technology or policy selected for epIA

The adoption or influence of a technology or policy is a necessary condition for it to have an impact. Compared to estimates on

7 Of the five studies, the trade-off between the expected size of effects and the existence of benchmark data in selection was most noted in the impact assessment of agroforestry fallows (Place et al., 2007). Larger farm sizes and greater soil phosphorus availability meant outcomes on productivity and impacts on household welfare were more marked in Zambia than in western Kenya. The latter was, however, chosen for assessment because of the greater availability of data. Small and often insignificant impacts – because of the very small areas planted by adopters, and phosphorus deficiency – were a repeated finding in their assessment.

other variables in an epIA on agricultural research, those on adoption are usually shrouded in uncertainty. Economic rate of return assessments are predicated on annual estimates of adoption. It is only for very few technologies that annual estimates can be furnished from primary or secondary data without having to resort to projection or backward forecasting. Sensitivity analysis often shows that estimates of the size of net benefits are more sensitive to adoption levels and rates than to those of any other variable (Walker and Crissman, 1996). Typically, a survey is used to elicit information on adoption and this inquiry, which often relies on mixed methods, may account for the bulk of operational expenses in any epIA. The above argues that adoption should not be taken lightly. Most of this chapter and its resulting guidelines focus on the informational requirements of adoption that define good-practice epIA. Special attention should be paid to choosing selected policies for epIA because documenting policy adoption or influence carries its own unique challenges.

Good practice 3.5. Reliable data on adoption in technology-oriented epIAs and sound documentation of causal inference in policy-oriented epIAs are critical for persuasive story-telling about successful R&D.

Key references
Feder G., Just R.E. and Zilberman D. 1985. Adoption of agricultural innovations in developing countries: A survey. • A readable review cited by agricultural economists.

Rogers E.M. 1995. Diffusion of Innovations (fourth edition). • The classic interdisciplinary compendium on the adoption and diffusion of innovations.

Ryan B. and Gross N.C. 1943. The diffusion of hybrid seed corn in two Iowa communities. • The first and most influential adoption study; widely perceived to be the first use of the case study approach in the evaluation literature.

Cross-checking with information on tangible technologies

The adoption of some tangible technological components can be estimated from data on market sales. Examples include hybrid and open-pollinated variety seed (from IARC parents), agricultural machinery, and biological control products. Skillful use of these data on production, sales, and distribution over time may well provide 'harder' estimates of adoption than survey data on user uptake. At a minimum, the former should be used to check the latter on the availability of embodied technological change. This is because adoption estimates derived from surveys and simple extrapolation models should not exceed the stock of a technology available for use at a certain time.

However, care must be taken that output sales and distribution data reflect the use of tangible technologies. Apart from seed, the initial demand for these tangible (usually) public-sector products can be heavily subsidized and institutional. Data on production tells us little about the extent of adoption when the incidence of product abandonment and/or underutilization is significant (Kshirsagar et al., 1984). Suspicions about abandonment point to the need for a survey to document the level of use. If confirmed, such suspicions move the analysis from providing estimates on adoption to enhancing understanding on use or disuse. However, strictly speaking, this is beyond the boundaries of epIA as described here, since it explores the reasons for lack of impact rather than documenting realized impact.

Good practice 3.6. It is recommended to cross-check survey estimates on adoption with sales and distribution information for tangible technologies and with availability calculations for all technologies.

Key reference: Morris M.L. and Heisey P.W. 2003. Estimating the benefits of plant breeding research: Methodological issues and practical challenges. • An insightful discussion of many of the pitfalls in the application of epIA, including the use of non-survey information to inform about the adoption of technologies related to plant breeding.

Multi-purpose field research for epIA: The adoption survey

Field research for an epIA is based on multiple instruments, ranging from rapid rural appraisals to detailed field measurements on technological performance. Adoption surveys are usually a staple element in field research for technology epIAs. Different types of adoption surveys related to research purposes are described in CIMMYT's Economics Program (1993), with impact assessment adoption surveys belonging to the set of these adoption instruments. In contrast, traditional diffusion surveys are oriented towards understanding what determines adoption and the constraints on diffusion (Rogers, 1995). Impact assessment adoption surveys can also supply information in these areas, but their main objective is to serve as a basis for estimating the level of adoption and for contributing information to estimating benefits, as discussed in the next chapter.

Focusing firstly on estimating the extent of adoption, the analysis of adoption survey data is expected to result in:

- Identifying the technological components of success stories that were adopted and those that were not
- Estimating adoption for the most widely accepted components for at least one point in time
- Revising the description of the recommendation domain (the area in which the technology is promoted)
- Specifying the peak or ceiling level of adoption if the selected technology is at an early stage of adoption.

If the selected technology is at an early stage of adoption at the time of the fieldwork, adoption surveys should also address the threat of disadoption. Such threats may be internal, because aspects of the technology can 'break down' and stop working over time. Alternatively, they may be external, for example changes in policies or relative scarcities in factors of production that condition the desirability of the technology to users or the advent of new technologies (Neill and Lee, 2001). Disadoption in the late adoption stage of a selected technology is natural, whilst very late disadoption – beyond 20 years from initial acceptance – is unlikely to affect the results of epIAs unduly (Boys et al., 2007).

Depending on the geographic coverage of an epIA, an impact assessment adoption survey is often conducted to estimate benefits. The survey should seek to answer the following questions:

- What are some of the characteristics of 'typical' adopters?
- What technology did the new practices and varieties replace?
- What are the strengths and weaknesses of the 'new' technology vis-à-vis the 'old' technology?

Hence, information in the survey identifies the adopters, the technology that was replaced, and the relative attributes of the new technology that point to trade-offs in sources of benefits. Aside from generating the raw material for estimates on adoption and contributing information on benefit estimation, impact assessment adoption surveys should also be able to uncover human interest stories that have potential use in raising public awareness.

Analysis of the adoption data may reveal a fundamental flaw in the selected technology or a level of acceptance that is significantly below that of the anecdotal evidence factored into the selection process (Kelley et al., 1990). Another unfavorable outcome for accountability would be if demand for the technology was highly subsidized or institutionally motivated. If the analysis suggests that the prospective success story is not viable at the time of the adoption survey, the impact practitioner may not want to expend the extra effort needed to conduct a thorough epIA, which may be premature and characterized by so much uncertainty that its validity will be questioned. Insignificant results are rarely published and, if reported, they are usually found only in the grey literature. The conventional practice of not engaging in epIAs for negative rate of return technologies means that these technologies are visible only by their absence.

Research on adoption of the selected technology should not be counted as 'a failure' if it identifies non-adoption, so long as it demonstrates lessons to be learned. For that potential to be realized, adoption research needs to be reported and presented to concerned scientists and research managers. It should be made clear whether non-adoption occurred primarily because researchers mis-specified the technology need, or whether there were constraints beyond their control. In the former case, lessons should be drawn by the institution for future research planning, perhaps by putting greater emphasis on participatory research with farmers. In the latter case, collaborative research and/or advocacy could be used to try to alleviate these limitations and to anticipate them in future strategic planning and *ex ante* impact assessment. The conduct of these types of epIAs should at least result in lessons being learned.

Good practice 3.7. Information on the adoption component for a good-practice epIA includes 1) describing the size and heterogeneity of the re-specified recommendation domain, 2) a point estimate of component adoption when the fieldwork was conducted, 3) the values of parameters used to estimate or project the level of adoption, 4) any distinguishing characteristics of adopters and non-adopters – especially their poverty status, 5) perceived strengths and weaknesses of the technology, and 6) the threats that could lead to early disadoption.

Key references

Bellon M.R. 2001. Participatory Research Methods for Technology Evaluation: A Manual for Scientists Working with Farmers. • Excellent reference on qualitative techniques applicable to the conduct of multi-purpose adoption surveys.

CIMMYT Economics Program. 1993. The Adoption of Agricultural Technology: A Guide for Survey Design. • An informative guide describing multiple types of adoption surveys including those directed towards assessing impact.

Diagne A. 2006. Diffusion and adoption of NERICA rice varieties in Côte D'Ivoire. • State-of-the-art research on projecting diffusion from data on early adoption.

Griliches Z. 1957. Hybrid corn: An exploration in the economics of technological change. • Path-breaking article on the use of logistic curves to project rates of adoption over time during the first ever epIA.

The adoption of outputs from policy-oriented research[8]

The most challenging task in carrying out epIAs of policy-oriented research (POR) is to establish the degree to which a newly adopted policy can be attributed to specific well-defined research. For policy-oriented epIAs, the discussion of adoption centers around skillful illumination of the nexus between the policy research in question and any decision-making that led to the policy change in question[9].

Measuring the economic impact of a POR output requires investigating the extent of dissemination and uptake, the stimulation and or contribution to policy influence, evidence of policy change, and then effective implementation prior to the use of models and economic analysis to assess the consequences of all this (Box 3.2). A key step is attributing changes in these variables to the selected policy. The attribution must consider both the direct effects of the POR over and above that of other research and information, and the many other influences on policy-makers. It is here, in attributing the cause(s) of change, that the counterfactual scenario is critical. This is because, unlike assessing the impact of crop genetic improvement, where technologies resulting from research may be embodied in a product whose adoption can be tracked, with POR we are dealing with the much less tangible aspects of information or disembodied change. The output of POR is but one input into the policy development process, and separating its contribution from that of other inputs and then valuing it is much more of a challenge than for epIAs of embodied technological change.

Uptake. Empirically, there are two steps that can be usefully investigated by epIA practitioners. Firstly, evidence that policy-makers pay attention and are familiar with POR outputs; and secondly, evidence that after paying attention, the research influences their views and hence their decisions. To generate influence on policy-makers, POR outputs must reach them by being

8 The discussion in this subsection draws heavily on Gardner (2007).

9 POR is defined here as referring to any research (economic, social, biological, or physical) that aims to inform and ultimately influence choices made by governments or other institutions whose decisions are embodied in laws, regulations, and investments that generate benefits and costs for people whose interests are affected by those governments or institutions.

Box 3.2. Identifying the factors governing the impact of the CGIAR's policy-orientated research

The Science Council SPIA recently investigated the factors and issues governing the impact of the CGIAR's POR. The first phase of this initiative reviewed the literature on different types of POR carried out by the CGIAR to evaluate its direct and indirect outcomes along the impact pathway (SPIA, 2006a; Raitzer and Ryan, 2008). The review identified only 24 such epIAs, about half of which were conducted by the International Food Policy Research Institute (IFPRI). Most of these epIAs went only as far as assessing the uptake/diffusion or the influence of the research, with no attention given to measuring the actual impacts. Only three of the studies attempted to estimate an economic value of the research (Babu, 2000; Ryan, 2002; Ryan and Meng, 2004). The review's conclusions were:

- An inadequate range and number of POR impact assessments were carried out, with far too few having gone beyond mere 'diffusion' to critically examining 'influence' and 'impact'
- Significantly higher degrees of difficulty (methodologically) were inherent in documenting the impact of POR, which implies the need to push the methodological frontiers
- It is desirable and feasible to execute a more credible set of studies to document influence and impact of this type of research further along the impact pathway.

SPIA addressed the second and third concerns in a second phase of this initiative looking at case studies on seven CGIAR centers. These studies are emphasizing methods and practices that rigorously establish influence and generate credible estimates of impacts – impacts that contribute to achieving the CGIAR's main goals of poverty alleviation, food security, and sustainable NRM. Previous work had shown that analysts must confront at least the following eight key issues to assess the impact of POR:

- Scale. Will the evaluation focus on the institutional, programmatic, thematic, or project level?
- Time lags and discontinuities. Since the policy process is not linear, how will the evaluation acknowledge discontinuities in the use and influence of information and lags in the generation of benefits from resulting policy change?
- Demand side versus supply side. Ideally, evaluators should start at the point of demand for information and work backwards from actual policy changes to the research, although simply tracking how policy-makers have used the research may be more feasible.
- Surprise. How did the addition of new information to policy-makers' perspectives affect their perceptions about the desirability of policy change?
- Attribution. Many actors participate in the policymaking process and use various sources of information to inform their decisions. Therefore, attributing impact to any one source is difficult, although donors will continue to press research institutions to identify the extent to which their work contributes to bringing about policy change.
- Choice of indicators. Evaluators must determine the variables of interest, i.e., the output, the outcome or influence, the policy response, and the impacts.
- Sampling. Even in evaluations using case studies, evaluators must decide whether to make a random or purposive sampling. Each approach has its pros and cons.
- *Ex ante* and *ex post* assessments. Researchers can develop logical frameworks to gauge the level of success of research in achieving its objectives. They can also document a project's outputs, outcomes/influences, and policy responses. This promotes internal learning and enhances institutional effectiveness. Independent evaluations are still, however, needed to ensure credibility.

From a review of the seven case studies, IFPRI identified the nine most important factors for POR to generate influence and impact (Ryan and Garrett, 2003):

- The production of high-quality, independent research
- The timely availability of relevant research information
- The long-term and in-country presence of researchers
- A policy environment conducive to research results
- The presentation of empirical data and simple analysis
- The likely trade-off between the immediate production of results and impact and the long-term (a) building of in-country capacity to undertake their own research and (b) impairment of research quality
- The strategic choice of partners and the identification of 'policy champions' who may effectively advocate for policy change
- Building a consensus for change among stakeholders
- Learning from cross-country experiences to improve ways of conducting research and influencing policy change.

Sources: SPIA (2006a) and Ryan and Garrett (2003) disseminated to them and the outputs need to gain their attention in such a way that they take it up – they 'use' it. The dissemination of research findings happens through conferences, policy briefs, and other activities aimed at informing policymakers. Effective dissemination is crucial. Quantitative data on how much dissemination happens can be gained by estimating the amount of resources devoted to it.

In terms of gaining the attention of policymakers, it is necessary, but not sufficient, that policy-makers know about the research. The research output then has to enter the policymaker's frame of mind or 'beliefs'. Although measuring the extent to which this occurs is difficult, Bayesian methods can be applied to quantify the use of evidence in transforming prior beliefs into posterior (subsequent) beliefs (Schimmelpfennig and Norton, 2003; Norton and Alwang, 2004). Such beliefs will be expressed as subjective probabilities about a state of affairs. Using this Bayesian approach leads to questions focusing on how policy-makers' posterior beliefs differ from their prior beliefs, with the intervening information that probably led to the changes being that provided by the research under investigation.

A good example of an attempt to obtain evidence for documenting uptake is Landry et al. (2003) on the use of university research[10]. This type of approach has relevance for the CGIAR in suggesting the types of questions that policy-makers need to be asked in order to document the uptake section of the impact pathway. However, it has less relevance as a rigorous tool for statistically testing uptake and influence.

Influence. The impact pathway of POR is at its most complex in the conduit between the research output, as taken up by its immediate 'clients', and its influence among a broader set of users. It is essential to be clear about what exactly is being influenced. The influence of ultimate concern to the CGIAR and its donors is on policies articulated, approved, and subsequently implemented by the decisions of legislators and regulators. These are political actions, of course, and there exists a huge literature aimed at understanding political action. As Sabatier (1991) has pointed out, it becomes easier to deal with the 'influence component' of POR impact assessment as the impact pathway is clarified and more certain knowledge of the connections along the pathway is obtained. Hence, the clearer and more convincing the presentation of the impact pathway, the more credible the case for decision-makers having indeed been influenced.

Ideas that can be helpful in POR impact assessment studies come from a variety of sources, including the results of surveys or interviews with policy-makers and their staff, and reports from advisors or others who seek to understand the policy process, together with other competing claims of influences on policy change. Eliciting convincing evidence on these matters is by no means a trivial exercise and it is here that there is scope for methodological improvement.

There is substantial scholarly literature on the development of statistical methods that improve on the crude correlations used in many studies. Mohr (1995) is an excellent example of this. He provides a good discussion of many aspects of the impact pathway, although his main focus is on developing statistical (mainly cross-sectional and time series regression) models that enable analysts to test hypotheses about the causes of policy change. Of most interest is the relationship between variables that measure POR output on the right-hand side of the equation in a regression analysis as one factor (among other qualitative factors) against a measure of influence explained on the left-hand side.

The main obstacle to applying statistical methods for estimating influence is that statistical methods obtain their evidence of influence by comparing a number of different outcomes, each associated with a set of explanatory variables. For many case studies, including the SPIA POR impact

10 Uptake was scored based on the answers to key questions put to policy-makers, such as whether they had seen reports of the research, understood the results, and had been influenced by them. Thus, they incorporated measures of both uptake and influence. Using multiple regression, they explained 'uptake score' as a function of characteristics of the research at issue and the university from whence it came.

assessment projects described in Box 3.2, only one or a few influence outcomes are considered, with only one or a few values given for the explanatory variables. With these case studies, the influence was scored as either 1 or 0 and most likely 1 since these projects focused on successful POR.

In this and other POR impact assessment, the best that can be done is to rely on interviews with policy-makers to provide information that would be generated more objectively by statistical studies relating measures of output to a measure of influence. Thus, instead of observing cases where policy-makers read or do not read POR outputs and then associating the reading (or not reading) with their subsequent votes for or against policy change, policy-makers are asked whether they have read the POR output and, if they have, whether it influenced their vote. This puts a great deal of importance on policy-makers knowing in some detail what they did and why they did it, and makes the substantial assumption that they will answer frankly. This makes the approach taken to policy-makers – the questions asked, the way they are asked, and when they are asked – crucially important. In many instances it is policy-makers' advisers who actually read POR outputs and use them as they deem appropriate in influencing their advisees, who have the power to vote in change. This adds a layer of further complexity to the elicitation of causal influences.

In summary, in contrast to the relative ease of identifying the impact pathways for the adoption of technologies, it is a much more complex and difficult undertaking to identify how POR outputs are taken up and influence policy change. This highlights the need for further research by economists, political scientists, and others on this subject and for methodological advances in elicitation techniques and qualitative data analysis.

Good practice 3.8. An assessment of the strength or weakness of threats to the *ex post* validity of the causal linkage between POR and the selected policy is just as important to an epIA of a selected policy as the analysis of adoption is to an epIA of a selected technology.

Good practice 3.9. Clearer and more comprehensive identification of the impact pathway (relevant to policy) adds credibility to impact analysis, especially in making the case for how research influences the development of policy.

Key references

Pardey P. and Smith V. (Eds) 2004. What's Economics Worth? • The best overview of assessing the impact of POR.

Standing Panel of Impact Assessment (SPIA). 2006a. State of the Art in Impact Assessment of Policy-oriented Research in the CGIAR: A Scoping Study Report. • A recent synthesis of epIAs of policy research in the CGIAR.

Arriving at a persuasive counterfactual

Elucidating the counterfactual scenario is tantamount to describing the 'without research' picture, i.e., what would have happened if the research had not been conducted and the resulting technology had not existed. The construction of an appropriate counterfactual scenario lies at the heart of assessing the impact of social programs and policy research. In technology-oriented epIA, the challenge of constructing a convincing counterfactual is, in most cases, less daunting.

Adoption surveys usually supply information on one version of the counterfactual – the technology replaced by the new technology. It is hoped that the replaced technology will be the same or similar to the control treatment used in experimentation or, at least, the control treatment will provide the basis for constructing the counterfactual.

The control treatment will usually reflect the *status quo*, although an epIA should not assume that the status quo is going to remain the same throughout the life of the selected technology (Morris and Heisey, 2003). For example, assuming that a local variety of a crop is the counterfactual makes sense in a region where there is negligible varietal change. But, in a region of dynamic varietal change, the assumption that the farmers' variety is the counterfactual will overestimate the impact of varietal change. This is because it is likely that more productive replacement varieties would have been adopted anyway, aside from the impact of the research. In this case, empirical evidence on mean varietal age, as based on the date of cultivar release, helps in identifying the counterfactual. The source of counterfactual varietal change could be the same institute that carried out the research under investigation, another public sector institute, or the private sector. Foreseeing what is likely to happen begs the issue of fungibility; in other words, without the research, when would an equivalent substitute have become available from any source, whether from spill-over effects, from learning by doing, or from another research body?

In some epIAs of technologies, the counterfactual can occupy center stage in the evaluation. This is the case for the biological control of exotic pests, where cataclysmic changes in pest incidence can usher in a counterfactual that was never envisaged.

The challenge of constructing an appropriate counterfactual scenario for policy research is not as difficult as finding a reasonable 'without research' scenario for the biological control of exotic pests; but it requires more thinking than the specification of a counterfactual for most technology-oriented epIAs. The expected duration of the without research scenario can be highly uncertain in policy-oriented epIAs.

The counterfactual is almost always a candidate for sensitivity analysis in the presentation of results. Such an analysis should demonstrate the sensitivity of results to changes in key variables. In epIAs of NRM technologies aimed at addressing resource depletion or conservation, models can be used to construct scenarios for multiple counterfactuals. The sensitivity of the results to these alternative counterfactuals is potentially one of the more informative aspects of epIAs that focus on resource management technologies. In this thematic area, the reliability of the counterfactual is synonymous with the value of the predictions of the models.

Summing up: counterfactuals should attempt to mimic the true 'next best' options. To do so, they should always be explicit and, as much as possible, be derived empirically. Laxmi et al. (2005), Lindner (2006), and Hossain et al. (2007) are three examples of relatively recent epIAs that have done a good job in specifying reasonable counterfactual scenarios.

Good practice 3.10. All epIAs of agricultural research should explicitly identify a counterfactual scenario to help explain how the outcomes of interest (impact indicators) would have changed if the research had not been done.

Good practice 3.11. In technology-oriented epIAs, the control treatment in field experimentation and an identified replaced technology in the adoption survey provide the starting point for identifying an appropriate counterfactual. The counterfactual should be based on empirical evidence.

Good practice 3.12. The challenge of specifying an appropriate counterfactual depends mostly on the type of research intervention and the context in which it is adopted. Sensitivity analysis is a desirable element of counterfactual analysis.

Key reference: Zeddies J., Schaab R.P., Neuenschwander P., and Herren H.R. 2001. Economics of biological control of cassava mealybug in Africa. • Presents four realistic 'without research' scenarios to cassava mealybug control that enable readers to draw their own conclusions.

Estimating benefits and costs

Generating reliable estimates of adoption and net benefit per unit adopted is synonymous with good-practice epIA. The results of epIAs are usually insensitive to changes in assumptions on costs relative to comparable changes in the rate and level of adoption and the magnitude of net benefits per unit adopted (Walker and Crissman, 1996). If the selected technology or policy is well-defined, then research costs can also be known with more certainty than the adoption and net benefit levels. For these reasons, this section describes the estimation of benefits in some detail.

As with adoption, estimating the benefits of the effects of technologies generated by agricultural research deviates somewhat from the conventional good-practice impact assessment of government interventions in the health and education sectors. This section highlights these departures.

Evaluating sources of information for estimating benefits

Although baseline survey information is useful for good-practice epIA, its existence is not a necessary condition for making reliable estimates of benefits for economic rate of return assessments. Before-and-after comparisons are notoriously sensitive to annual changes in production conditions, particularly in rainfed agriculture. Fluctuations across production years usually swamp the effects of technological change in all but the most assured rainfed and irrigated production environments. If before-and-after comparisons are used as a basis for estimating benefits, epIA practitioners have to argue persuasively that the 'before' year was just as good or even better for production than the 'after' year.

In contrast, establishing a baseline is usually viewed as a prerequisite for evaluating government interventions and social programs (Center for Global Development, 2006). It is easy to see how, in evaluating impact in the health and education sectors, variability across years could be substantially smaller than inter-household variability between participants and non-participants. A baseline is substantially more desirable for Stage II multi-dimensional impact assessments than Stage I economic rate of return assessments. The demand for information in the former is considerably greater than that in the latter.

Pragmatically, reliable baselines cannot be established in all recommendation domains, particularly for IARCs that generate international public goods. But, when funds are available, a baseline survey is a worthwhile investment that generates returns for future impact assessment, especially when technological change is dynamic.

If funds are scarce, then targeting 'minimal informational' baselines for benchmark sites, where one or more well-characterized technologies is expected to generate impact, will be potentially beneficial for future epIAs. Also, as the multi-dimensional impact assessments in Adato and Meinzen-Dick (2007) implicitly suggest, there are dividends for concentrating fieldwork at the same sites over time.

For NRM technologies, before-and-after comparisons are particularly valuable when a technology results in an increase in or prevents the erosion of the resource base. For example, a land management technology that loosely translates to 'ridge tillage' has been adopted on tens of thousands of hectares in West Africa's semi-arid tropics (Institut D'Economie Rural et al., 2006). The adoption of this technology was accompanied by an increase in tree and shrub vegetation, indicating that the technology has had a positive effect on groundwater recharge and has enhanced the availability of drinking water. The availability of before-and-after hydrological data is extremely useful for establishing the size of this effect, which may or may not be accurately identified in a with-and-without comparison.

Evaluations that combine both before-and-after and with-and-without comparisons will usually give a more rigorous estimate of benefits than either carried out alone. In practice, relatively few assessment studies in agriculture have carried out such double difference assessments based on before-and-after and with-and-without survey data. One of the few examples is Walker and Kshirsagar (1985).

With-and-without comparisons are the norm for estimating benefits in technology-oriented epIAs. Advances in statistical techniques such as propensity scoring have helped to reduce the variability between with-and-without groups (Baker, 2000; Godtland et al., 2003). Although they are standard practice, with-and-without comparisons do have their problems. One of the fuzziest areas is not being able to attribute farm-level effects to component technologies even with well-executed, multivariate analysis. A large residual effect between the with and the without groups may remain unexplained (Dalton et al., 2005). That effect should not be credited to the technology selected for the epIA.

Productivity comparisons based exclusively on survey data are a risky choice for benefit estimation in epIA in developing country agriculture. As much as possible, the estimation of benefits should be based on experimental data, particularly on-farm research data, and on the field measurement of productivity effects. Even adjusting experimental station data to farmer conditions will often provide a sounder basis for estimating benefits than using single-interview survey data. Results from survey data should be regarded as a source of triangulation for estimating benefits from experimental data, particularly for on-farm trial and test data.

Economists and other social scientists tend to undervalue the worth of experimental data and detailed field measurement of productivity effects in benefit estimation for epIA. A lack of experimental data on the selected technology can reflect problems in the institutional attribution of the technology. An absence of such data suggests that agricultural research may have had little influence on the technology, and that this may be more the result of another factor such as a training exercise.

By the same token, the availability of experimental evidence does not necessarily mean that it should always be used. Crop and resource management technologies may feature heavy participation of farmers in adapting technologies to site-specific conditions (Douthwaite et al., 2005). Specific examples of technologies fitting this description are minimal tillage and the selection of species for vegetative barriers in soil conservation (Kerr, 2002; Winters et al., 1998). The role of agricultural research in this case is to initiate the process of adaptation. If agricultural researchers conduct detailed experiments on the 'wrong' treatments – the ones that farmers do not adopt – then experimental evidence is not going to translate into the reliable estimation of benefits. In this case, the carrying out of true *ex post* experimentation on the treatment effects of the main components should be considered if such consequences cannot be documented from analyzing the survey data.

The ideal experiments for benefit estimation in epIA are those that feature plus and minus components in their design. The package approach to technology transfer is at the opposite end of the spectrum to this ideal. The focus in epIA is on the components of the selected technologies. Results from experimental or demonstration packages are rarely a viable source of information for reliable epIA as this would involve breaking down packages into their component parts and gathering information on their contributions and interactions. In that regard, adapting a farmer-field school approach to research is inimical for reliably estimating benefits in epIAs. Such an approach may be fine for extension, but the emphasis on one main field plot with only two treatments during the growing season combined with a lack of directive scientific input into technology validation makes the identification of component effects a daunting task. In contrast, the mother–baby format of experimentation could be conducive to reliable benefit estimation for epIA and is increasingly used to measure the size of component effects in an on-farm setting (Joshi et al., 2007). The increasing popularity of the mother–baby approach is attributed to its simplicity and its capacity to generate transparent treatment effects for both researchers and farmers.

One of the gaps in the literature is a key reference on evaluating data sources, particularly experimental data, for estimating benefits in epIAs of specific research-generated agricultural technologies. The conventional experimental approach may prove viable for estimating the benefits of agricultural policies linked to government interventions and programs. Indeed, much is known about the validity of experimental (e.g., randomized control trials) and quasi-experimental methods in the impact evaluation of social projects and programs (Shadish et al., 2002).

Such knowledge should be directly transferable to assessment of agricultural policies that have a project or programmatic content. Recent research at the IFPRI employed experimental methods to evaluate government interventions to good effect (Skoufias and McClafferty, 2003; Ryan and Meng, 2004; Maluccio et al., 2005).

However, for research-attributed policies, where influence is at the national level, and for research in general that is aimed at generating international public goods, conventional experimental approaches tend to be substantially less applicable for evaluating impact.

Good practice 3.13. Experimental data on the treatment effects of adopted technology are preferred to survey data for estimating benefits in economic rate of return assessments. With-and-without comparisons are preferred to before-and-after comparisons when survey data are the basis for estimation, but the ideal is to have both sources of data available so that a double difference can be estimated. In good-practice epIAs, both experimental and survey information should be used to increase the reliability of benefit estimation.

Good practice 3.14. Baseline data are useful in economic rate of return assessments of technology-oriented research, but their availability is not a necessary condition for good-practice epIAs. In contrast, baseline data are highly desirable for elaborating Stage II multi-dimensional impact assessments and to enable both before-and-after and with-and-without comparisons to be made.

Good practice 3.15. For NRM technologies, selective baseline data on resources that are conserved or augmented are potentially very valuable for estimating benefits in epIAs.

Key references

De Datta S.K., Gomez K.A., Herdt R.W., and Barker R. 1978. A Handbook on the Methodology for an Integrated Experiment-survey on Rice Yield Constraints. •A pioneering methodology that responded to the perceived problem of identifying the causes of low yields among adopters. This method is very compatible with benefit estimation in a Stage I epIA.

Feder G. and Slade R. 1986. The impact of agricultural extension: The training and visit system in India. • Perhaps the best known natural with-and-without comparison in agricultural development.

Ryan J.G. and Meng X. 2004. The Contribution of IFPRI Research and the Impact of the Food for Education Program in Bangladesh on Schooling Outcomes and Earnings. • Apt example of using a double difference approach combined with propensity score matching.

Shadish W.R., Cook T.D., and Campbell D.T. 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. • A must read for impact practitioners on research design in the 'Campbellian' tradition.

Snapp L. 2002. Quantifying farmer evaluation of technologies: The mother and baby trial design. • Arguably, the single most important methods innovation in on-farm research in the recent past. This method generates results that are user friendly for economic rate of return assessments.

Tripp R. and Wooley J. 1989. Planning Research for Farmers' Fields: Identification of Factors for Experimentation. • Good practical advice for prioritizing treatments in agricultural experimentation from an anthropologist and an agronomist.

Demonstrating that net benefits are sufficient to scale up adoption

Internal validity is a foremost concern for impact practitioners (Shadish et al., 2002). The drawing of inferences about the cause and effect relationship between a technology and its impact should start by estimating the technology's profitability. For epIAs of technology-oriented research, the practitioner should be able to provide evidence that the technology can generate sufficient net benefits to stimulate wide-scale adoption. In epIAs of POR, adoption will usually depend on political factors whose documentation, although it may be interesting, may not have direct implications for the cost–benefit analysis of the research.

Demonstrating that the net benefits derived from technological change will be sufficient to justify scaling up adoption is usually done through a partial budgeting exercise that gives an estimated marginal economic rate of return for adopting the technology. In epIAs carried out by economists, the magnitude of benefits is somewhat difficult to discern as they are usually expressed as a percentage cost reduction per unit of output. For the larger and wider audiences of epIAs, benefits expressed as net benefits per unit adopted are easier to understand than percent cost reductions. For many farm-level productivity effects, the unit of adoption is land. Founding the partial budget on the change in net benefits per unit of land is a transparent way to convey the adoption decision for most technologies. In land-abundant agroecologies, net benefits per unit of labor may be more illuminating if applying the technology significantly saves relatively scarce seasonal labor.

Aggregate benefits are often estimated using a simple economic model for estimating the change in economic surplus (Akino and Hayami, 1975). If the selected technology or policy generates very large impacts in society – so much so that the prices of inputs are affected and producers sharply reduce supply because of rapidly falling output prices – then estimating the increases in producer surplus as the aggregate benefits to society is considerably more complicated than focusing on net benefits to adopters. Assessing the secondary effects may also be important when looking at macro NRM research projects, which may have economy-wide effects and large-scale impacts on several variables of interest.

In this case, a general equilibrium analysis at the level of the industry, country, or even the world may be warranted. Because the skills available in research institutes for conducting more inclusive economic analyses are usually scarce, and because most non-Green Revolution success stories do not result in such large-scale impacts, a staged approach seems best suited to the context of most research institutes and types of technological change.

If the results of a Stage I epIA suggest truly large economy-wide effects, then investing in a Stage II epIA that focuses on the general equilibrium effects of the selected technology is justified and called for. The ACIAR guidelines contain some excellent advice based on a decision tree approach, suggesting which one of six alternative economic models to choose for such impact assessments (Gordon and Davis, 2007).

Good practice 3.16. An emphasis on direct beneficiaries and on reliable experimental data will diminish the threats to validity in attributing net benefits per unit of adoption to the selected technology in Stage I epIA.

Good practice 3.17. The net profitability of the technology to adopters per unit of the most limiting input should be demonstrated as a basis for calculating net benefits.

Good practice 3.18. Documenting the potential for large economy-wide development impacts in a Stage I epIA underscores the need for investing in a Stage II epIA to exploit this exciting opportunity for showing the multi-dimensional impacts.

Key references

Marra M.C., Pardey P.G., and Alston J.M. 2002. The Payoffs to Agricultural Biotechnology: An Assessment of the Evidence. • This thorough study highlights the use of net benefit estimation.

Perrin R.K., Winkelmann D.L., Moscardi E.R., and Anderson J.R. 1976. From Agronomic Data to Farmer Recommendations: A Manual for Economic Evaluation. • A classic treatment of calculating net benefits from experimental data using a partial budgeting format.

Attributing net benefits

Successful research interventions can rarely be attributed to just one agent or organization, as almost always several institutional actors will have made a contribution. The actors range from advanced research institutes and universities to public-sector extension agencies and non-governmental organizations (NGOs). The main issue here centers on how to apportion benefits when more than one research institute has been involved in generating and adapting a technology. Moreover, inter-institutional collaboration across research institutes is increasing and is likely to grow still further (Pardey et al., 2007).

Having a multiplicity of actors should not be a cause for despair when trying to attribute net benefits. Embodied agricultural technologies are frequently distinctive, tangible, and sufficiently 'hard' for it to be recognized which institutions contributed what to their development. For varietal change, sensitivity analysis based on rules of apportionment has been used in both economic rate of return (Pardey et al., 2006) and multi-dimensional impact assessments (Fan et al., 2007).

For non-varietal technology-oriented research, a full description of who did what and when should be sufficient to let readers decide on the merits of the argument for attribution. The standard practice for dealing with questions of attribution is adequate as it considers project benefits to be a joint output of the institutions involved, thus charging the project-related costs of each and every institution to the project. EpIAs should contain enough information on attribution to make a persuasive case for an institution having exceeded a minimum threshold level of involvement in developing the technology.

A more challenging issue is how far back in time to attribute a selected technology or policy to past research; i.e., when should cost–benefit analyses start? For example, research on successful varieties usually 'starts' when the crosses are made. For the adaptation and testing of material from a national perspective, the research will have started when the material was first imported into a country or when the program first started importing similar characteristic material. Some epIAs of plant breeding programs – similar to the work of Pardey et al. (2006) – have imputed a part of the benefits to previous research that was instrumental in developing the present-day material (Byerlee and Traxler, 1995). These epIAs do not move the date the research started, which would have large consequences for most epIAs, but rather reduce the allocation of benefits by clear rules of apportionment. In the cost–benefit meta-analysis of epIAs in the CGIAR (Raitzer and Kelley, 2008), studies that apportioned benefits in this manner were viewed more positively on attribution than those that did not. The informational requirements for apportionment are large and the assumptions underlying the rules are subtle. The establishment of certain rules may be possible using pedigree information on varietal change, but the structuring of comparable credible scenarios for resource management technologies and for selected policies could quickly become very arbitrary. Success in agricultural research is difficult enough without having to quantify institutional ownership.

Good practice 3.19. Benefits are considered to be a joint output of all relevant participating institutions, and the project-related costs of each participating research and extension institution should be charged to the selected technology or policy. The attribution of benefits to each institution is not essential in good-practice epIA, but is often sought by investors.

Addressing other benefit-related concerns

Specifying the appropriate counterfactual scenario does not exhaust the thorny issues related to estimating benefits. To conclude this discussion on benefit estimation, this subsection examines the four issues of the duration of benefits, economic pricing, environmental benefits (see Box 3.3), and the use of national survey data on household income or consumption expenditure for calculating benefits.

EpIAs of technology-oriented research will start with the initiation of the research and usually end when adoption of the technology has reached, or is projected to have reached, its maximum. Disadoption in the late phase of diffusion (if it comes after many years) is rarely modeled (Boys et al., 2007). Because of discounting, results are usually not sensitive to the outcomes in later years in the life of selected technologies, but there are exceptions. Doubling the length of the period of appraisal for an agroforestry technology in southern Africa was accompanied by a five-fold increase in the economic rate of return on investment (Ajayi et al., 2006).

In good-practice Stage I epIA, the analysis should be conducted from the perspective of users and the costs they face when taking the decision whether or not to adopt. If the profitability of the selected technology is influenced directly by distortions at the level of the agricultural sector, then the existing financial scenario should be complemented by a scenario that uses economic valuation, usually based on international prices. Economists also typically

Box 3.3 Documenting the impact of NRM research in the CGIAR

The IARCs that make up the CGIAR have conducted research on NRM and associated policies since the 1970s. A dearth of documented credible evidence that this research has contributed to mission-level impacts on poverty, food security, and the environment on a wide scale was underscored in a World Bank (2003) review of the CGIAR. The CGIAR's Science Council responded to this challenge by commissioning two key initiatives through its SPIA: one focused on the impact assessment of NRM research (SPIA, 2006b; Waibel and Zilberman, 2007) and the other on the impact assessment of policy research (SPIA 2006a; Raitzer and Ryan, 2008). The latter is described in Box 3.2.

The initiative selected seven center case studies of NRM research impacts (see note below). While the NRM research epIAs were diverse in the research products they generated and methodologies applied, a common element amongst them all was the primary output of the research: the enhanced productivity and sustainability of renewable and non-renewable resources and the mitigation of negative environmental side-effects. Because of the different methodological and data constraints faced, the studies varied in being able to clearly and credibly define, and then quantitatively assess, impacts. For example, in the two macro NRM policy-oriented cases from CIFOR and the International Water Management Institute (IWMI), it was possible only to identify the impact pathways and establish influence. The cases could identify and speculate about, but not quantify, impacts. The five farm-level NRM technology epIAs from CIMMYT, the International Center for Tropical Agriculture (CIAT), the World Agroforestry Centre (ICRAF), the International Center for Research in the Dry Areas (ICARDA), and WorldFish were more amenable to assessment. They utilized well-developed adoption models and economic surplus techniques to quantify benefits, costs, and economic rate of return parameters. However, positive spill-over environmental and other non-market benefits had not, for the most part, been estimated.

These studies raised some key issues highlighting major challenges for the CGIAR in documenting NRM research impacts, although not all of these are unique to NRM. Some of the key ones relate to the need for advances or improvements in the conceptual analysis of technological impact pathways, appropriate indicators of environmental impacts from research that can be effectively monitored over space and time, survey design and data collection, attribution, and communication.

Notwithstanding the complexity of NRM, the case can still be made for a more systematic treatment of measuring impacts – or their proxy indicators – in quantitative if not in value terms, as occurred in the CIFOR case study (Spillsbury, 2007). The key is defining, prior to project implementation, the critical impact indicators that show changes in key variables related to (1) resource productivity, such as soil fertility and water availability, as well as inherent environmental knowledge, and (2) social value indicators, such as less pesticide use, that the research is targeting and hoping to influence.

The SPIA initiative concluded that there is an urgent need for further conceptual and empirical analyses, including paying more serious attention to measuring the benefits and value of information generated through NRM research, and in particular its specific contribution to achieving the CGIAR's goals. This becomes even more essential when 'integrated NRM' research is being evaluated. This type of research is conceptually more inclusive, comprehensive, and process-oriented than conventional and more focused component NRM research; but the concept is new enough that lag times have not passed for impacts to become measurable. Even greater challenges are expected in documenting impacts in integrated NRM.

Note: This study defined NRM as research on land, water, and biodiversity resources management, focused on generating knowledge that results in technology options to sustainably enhance the productivity and stability of ecosystem resources.

correct for overvalued exchange rates and other fiscal and monetary policy aspects that can lead to distorted economies (Ahmed et al., 1995). Correcting for distortionary macroeconomic policy is a good practice, but it requires a higher level of expertise in economics and poses challenging problems in multi-country and programmatic epIAs.

Adjusting for overvalued currencies and policy distortions may not be necessary if the country implemented structural reforms in the 1980s and 1990s that led to sizeable reductions in parallel market premiums for foreign currency. Also, the average net taxation of agriculture fell sharply between 1980 and 1984, and again between 2000 and 2004 (World Bank, 2007). Nonetheless, the taxation of agriculture is still high in many agriculture-based countries, meaning that impact practitioners have to ensure that policy distortions do not lead to distorted results in their epIAs.

A variation on the theme of correcting for policy distortion centers on the need to consider environmental effects that are not market-related or priced. When assessing the impact of agricultural research, good practice suggests that the additional environmental benefits must be taken into account because these benefits may not be market-valued. This may be especially relevant for NRM research.

An assessment of the non-market or non-monetary benefits and costs of agricultural research requires the application of a wide array of valuation techniques, including contingent valuation, hedonic pricing, and travel cost methods (Waibel and Zilberman, 2007). Few of these were applied in the seven case studies of NRM impact assessment because of the cost- and method-related constraints noted in Box 3.3. While the effectiveness of these techniques is sometimes debatable and they may be expensive, they should be used judiciously for evaluating research projects that have environmental impacts.

One of the major conclusions of the SPIA's NRM research impact study was that the development of effective indicators of benefits – preferably those amenable to economic valuation – is a priority in developing a more comprehensive assessment framework of the value of NRM research. If estimating the economic value of these environmental benefits is not possible, then a rigorous description (quantification) of the physical environmental changes is basic and essential[11]. These remarks on the desirability of environmental benefit valuation apply with equal strength to assessing the impact of crop genetic improvement, which also has the potential to generate sizable environmental effects.

Supporting data from national rural household surveys are often an underutilized resource in epIA. Although teasing out effects is never an easy exercise, national or regional data on household income can be extremely valuable in analyzing consequences along the impact pathway. National survey data can also be exploited for estimating benefits, particularly in policy-related epIAs (see Box 3.4).

Good practice 3.20. When net benefits of the selected technology are conditioned by distortionary sectoral policies, impact practitioners should also assess and report on a scenario in terms of economic prices.

Good practice 3.21. Documenting sizable non-monetary benefits is a priority for all thematic epIAs, and not just for those focused on NRM research. At a minimum, such effects should be described in physical terms if they cannot be valued.

Good practice 3.22. Data from national surveys of rural households often represent an underutilized resource that could contribute to more informative epIAs. Practitioners should become more familiar with such datasets in regions where their research institutes operate and should exploit the increasing availability of panel household data.

11 NRM technologies that increase productivity and reduce negative environmental effects may also have secondary effects through macroeconomic multipliers, and may impact on employment levels as well as health. Few NRM research impact assessment studies have adequately addressed these aspects. A notable exception is Dey et al. (2007), in which indicators of nutritional and health benefits beyond the direct impacts were developed and effects estimated.

Box 3.4. Exploiting national survey data to estimate benefits

Intuitive knowledge strongly suggests that more education equals better income over a lifetime. However, proving this is not often easy. An epIA from Bangladesh, reported in Ryan and Meng (2004), is an excellent example of how a project's benefits can be estimated using national survey data and conventional methods. The epIA of the long-term effects of a Food for Education (FFE) program in Bangladesh, which IFPRI helped conceptualize and evaluate as a catalyst for its scaling up, focused on the effect of education on potential earnings. Using data from a survey of household income and expenditure in Bangladesh carried out in 2000 by the Bangladesh Bureau of Statistics, external impact assessors derived an equation relating an individual's hourly earnings to their attendance at school and the duration of their schooling. The assessors estimated potential earnings for males and females and for rural and urban employment separately.

The FFE program rewarded poorer families – female headed household, landless or nearly landless people, and day labor-based low-income workers – with food for the regular attendance of their children at school. Economic analyses by the assessors showed that boys taking part in the program could expect to increase their lifetime earnings by 11–18% and girls by 33–35% because of the estimated effect of the program on the probability of them attending school and the duration of their schooling. The schooling outcome estimates were based on propensity score matching and difference-in–difference methods from special purpose household sample surveys in FFE and non-FFE districts. In each case, the higher earnings figures were based on the students obtaining urban rather than rural employment. These increases would mean an economic return to investment in the FFE program of between 18 and 26%. IFPRI's involvement in the program also resulted in excellent economic returns to investment that were conservatively estimated at 64–96%.

Source: Ryan and Meng (2004)

Estimating costs

Assembling cost data on all the institutions that have contributed to a research intervention can be tedious, although it is not a particularly difficult exercise. But deciding when to stop including institutions can be a very subjective decision. Conservatism argues for an inclusive approach in deciding which research-, development-, and extension-mandated institutions to include on the cost side of the ledger.

Research costs should also reflect the administrative costs of the center in which they are located. One common procedure for charging these costs is to assume that they are proportional to the scientist years involved in researching and developing the selected technology or policy relative to the total number of scientists in the center.

Technology transfer costs can exceed research costs by several orders of magnitude. The results of smaller success stories can be sensitive to assumptions on how these costs are calculated. Again, a common practice in assigning these costs to a selected technology is to divide the time spent by extensionists on transferring the technology by their total working time in the specified government program or NGO and then assigning these people's costs proportionately.

The estimated costs of transferring technologies will lack the precision of estimated research costs. If the technology depends heavily on the public sector or NGOs for its transfer, assumptions about costing the transfer could be a candidate for sensitivity analysis whilst presenting the results. The assumption that extension is a sunk cost, and that its opportunity cost approaches zero, is not consistent with the spirit of erring on the side of conservatism as recommended in the third section of Chapter 2 (Methodological issues).

Good practice 3.23. Practitioners should err on the side of inclusiveness by including as many costs as possible in costing institutional contributions to research and extension that are related to the research intervention which is being assessed.

Good practice 3.24. If public-sector extension and/or NGOs play a large role in the transfer of the technology, then assumptions on calculating the extension costs are candidates for sensitivity analysis when presenting the results of smaller success stories.

Key reference: Manalo A.J. and Ramon G.P. 2007. The cost of product development of Bt corn event MON810 in the Philippines.
• An interesting application of the careful costing of R&D that set the stage for a future epIA.

Presenting results

If an economic rate of return assessment contains sufficient information for another practitioner to replicate the net benefit stream of the selected technology, then the results will have been technically well presented. (Such results could be faulty but at least they will be technically informative because readers can document the outcomes of the main assumptions used in the analysis.) Results are usually presented in a spreadsheet with information on:

- The dates of the start of the research and the end of the period of analysis
- The size of the recommendation domain
- Annual estimates of adoption
- Any parameter values conditioning those estimates
- The estimated net benefit per unit of adoption (this may vary over time)
- The cost components of research and extension over time
- The stream of net benefits that is the result of subtracting total costs from total benefits.

Additional information should be provided on how past costs and benefits have been deflated and on assumptions about future inflation and expectations, if any, on projected prices vis-à-vis projected costs.

Estimates of financial profitability are calculated from the benefit, cost, and net benefit streams. Of the conventional parameters of financial analysis, net present values are preferred in comparative applications of cost benefit analysis (Boardman et al., 2001). Donors' interest in the size of benefits also argues for presenting information on net present value (Raitzer and Winkel, 2005). The only problem with estimates of net present value is that they are hard to understand. Both internal rates of return and benefit–cost ratios are easier to communicate. Estimates of internal rates of return have their problems, but they convey useful information on the relative profitability of projects independent of size. Benefit–cost ratios should be used sparingly because they are easy to manipulate by including or excluding cost and benefit components (Boardman et al., 2001).

For calculating net present value, real social discount rates have declined over time in both developed and developing countries for evaluating public-sector projects. Increasingly, a real discount rate in the neighborhood of 5% is used to evaluate the results of agricultural research in developing country agriculture. A lower real rate of social discount reinforces the profitability of strategic research in generating international public goods with longer benefit horizons. Compared to applied and adaptive research, strategic research is characterized by benefits that are realized later and that are also expected to be larger.

Using higher discount rates typical of those recommended in Gittinger (1982) 25 years ago penalizes strategic research because it discounts later benefits more heavily. The use of lower real discount rates is good news for agricultural research that generates national and international public goods, particularly NRM research programs that are expected to result in longer-term environmental effects.

A sensitivity analysis based on changes in the major assumptions that underpin an impact study will complete the information on the spreadsheet of a good-practice economic rate of return assessment. Sensitivity analysis usually involves changes in one or more variables that are characterized by uncertainty. If the argument for the counterfactual scenario is not as strong as desired, then sensitivity analysis can take the form of a sooner-rather-than-later scenario where the main benefit of the research was to accelerate the adoption of a technology or policy that would have been adopted later anyway (Fuglie et al., 1999; Ryan, 2002).

The priorities for presenting other results have been discussed earlier in this document. Farmers' perceptions of the strengths and weaknesses of a technology should be summarized from the perspective of lessons for technology design. Any documented negative effects of a technology should be highlighted and the question: are there losers as well as winners? should be asked. Likewise, the strengths and weaknesses of the epIA should be discussed.

The match between expectations and empirical evidence is another fertile ground for discussion, including surprises and serendipity. Locating an epIA in the context of other epIAs in the same technological genre is another area for discussion with potential implications for future priorities and the allocation of research resources (Ortiz et al., 1996). Any priorities for the further assessment of consequences down the impact pathway should be presented and discussed to inform the desirability and technical feasibility of carrying out future research. If the selected technology is in the early adoption phase, then the prospects for scaling out to other recommendation domains should be speculated upon. Threats to external validity should also be evaluated and 'removable' constraints communicated to agencies that have the power to increase adoption.

Good practice 3.25. A well-presented epIA will be characterized by other practitioners being able to replicate the net benefit stream from the information contained in the impact study.

Good practice 3.26. Although spreadsheet-related information is essential to the presentation of results, the value of good-practice epIA stems from the textured interpretation of these and earlier results of the empirical research in the context of the related literature.

Assessing poverty impacts

Poverty has many dimensions and, in the jargon of evaluation (Shadish et al., 2002; Bamberger et al., 2006), its measurement is susceptible to threats to construct validity (i.e., measurements can lack credibility) because there are no natural units of measurement. Indeed, the effects of agricultural research on reducing poverty can be multi-faceted, with interventions derived from such research being more effective at addressing some aspects of poverty than others. Improving the performance of agricultural research in poverty alleviation is the subject of increasing interest, as evidenced by conceptual and empirical reviews (Kerr and Koavalli, 1999; Hazell and Haddad, 2001; Meizen-Dick et al., 2004), poverty-related priority-setting exercises (Kelley et al., 1995; Byerlee, 2000; Alwang and Siegel, 2003), and methods-oriented discussions on how to improve the documentation of poverty-reducing effects on impact assessments (Walker, 2000; Pearce, 2002).

Because agricultural research is primarily an instrument for increasing productivity and thereby contributing to economic growth, its impact on poverty is usually perceived to be indirect. But sufficient evidence (cited in Annex B) has now accumulated to shed light on the capacity of and conditions for agricultural research to be an engine for reducing poverty. The results from the literature lead to the following expectations for practitioners:

- The adoption of a technology that results in the growth of agricultural productivity in developing countries is very likely to be accompanied by a reduction in absolute poverty
- The adopted technology may increase, decrease, or have no effect on inequality
- The prospects are brighter for agricultural research to contribute to poverty alleviation in the agricultural-based countries than in transforming and urbanized countries (World Bank, 2007).

For impact practitioners, the good news about the ample scope of agricultural research to contribute to reducing poverty is tempered by the growing realization of the difficulty of estimating the size of that contribution, especially when the focus is on a single technology or even a cluster of innovations (as is usually the case in the disaggregated impact studies that are the focus of this guidance document). The rest of this section discusses some of these challenges and presents strategies for assessing impact on absolute poverty (as measured by income, consumption, and inequality indicators), on food security, and on other indicators of poverty.

Challenges and prospects for documenting poverty impacts

A new agricultural technology or policy can affect poverty at the household level through changes in household income, the quantity and quality of the food consumed, and the relative and absolute prices of commodities produced versus those consumed. The nature and magnitude of the impact on poverty of agricultural research at the household level thus depends on the characteristics of both the technology and the adopters. Predictions about the potential for agricultural research to contribute to alleviating poverty at an aggregate level (defined by geographic boundaries) depends not only on the characteristics of the technology and its adopters, but also on the technology's recommendation domain, the actual level of adoption, and the characteristics of non-adopters. Small adopting areas, diverse sources of agricultural and non-agricultural household income, and small farm size all conspire to defy the easy measurement of the impact on poverty of a technology in absolute terms (Hallman et al., 2007; Place et al., 2007).

When carrying out disaggregated impact assessments to quantify the absolute impact on household income and consumption expenditure, the practitioner faces two main problems: (1) interpreting the often seemingly modest impacts on absolute poverty, and (2) finding suitable national-level data on consumption expenditure and/or income poverty to use as a basis for evaluation. A result that says the adoption of a technology is accompanied by a 1% fall in the head count poverty index or a .01 decline in the poverty gap or squared poverty gap does not sound like a sizable impact. Such 'modest' impact usually elicits the often-made observation that agricultural research is one of many instruments needed to combat poverty. But a change of such a magnitude is actually very large because absolute poverty indices are not easily moved by the adoption of specific technology and policy changes in any sector. The expectation that specific technological change is going to bootstrap or help a marked proportion of beneficiaries across a poverty line is unrealistic unless the bulk of beneficiaries were located just below the poverty line prior to the technology being introduced.

The greater availability of rural household income and consumption expenditure datasets paves the way for more applications that quantify the impact of specific technology or policy change on absolute poverty. Thus far (with one exception), applications in this vein have been restricted to priority-setting exercises that have drawn on national household survey data to describe the impact of technological change on income poverty (Alwang and Siegal, 2003; Walker et al., 2006). The exception is Moyo et al. (2007), who probed the absolute poverty prospects for a well-defined technology in an *ex ante* assessment based on household income information. In the future, epIA-related assessments patterned after Moyo et al. (2007) will be desirable for the way they ground estimates of absolute poverty impact in a national or regional context. Prospects for such evaluations are bright, if the commodity is a staple food crop and if the impact practitioner focuses on an indicator such as the squared poverty gap that is sensitive to improvements in income for households living below the poverty line[12].

It is also desirable to use consumption expenditure datasets to investigate more thoroughly the absolute poverty consequences to net sellers and net buyers of food from specific technological and policy changes. Greater familiarity with and access to household income and consumption expenditure datasets by practitioners is a first step in making such applications a reality.

12 The conventional use of absolute poverty indices implies that the impact practitioner does not have to deal explicitly with the difficult question of weights for distributional cost–benefit analysis (Harberger, 1978). The implied weights of these indices are still somewhat arbitrary, but their usage is commonly accepted in applied welfare economics.

Strategies for increasing the poverty content of epIAs

There is a growing demand from investors for more information on the number of poor people who benefit from technological and policy change. In this subsection we discuss how this demand can be satisfied with micro or disaggregated economic and multi-dimensional assessments of impact.

Focus on documented large-scale success stories. Disaggregated multi-dimensional impact assessments are better positioned to provide relative information on the impact of certain benefits on poverty in a specific context. In agriculture-based economies, the number of poor households benefiting directly as technology adopters, and indirectly through reduced prices for net food buyers and increased employment for agricultural laborers, should be strongly associated with the net present values estimated in economic rate of return assessments. The above-mentioned priority-setting exercises, which are based on representative national household data, show that the poverty reduction potential of a commodity is highly correlated with its value of production. In other words, the value of production is a good predictor of the poverty reduction potential across crop and livestock species. Drawing on this stylized finding, more widely adopted research interventions that spread benefits across many people should therefore be characterized by substantially better prospects to alleviate absolute poverty than more localized interventions that concentrate benefits in the hands of only a few people. The former is conducive to large net present values and potential for poverty alleviation; the latter results in small net present values and has limited scope for poverty reduction.

Conducting multi-dimensional impact assessments on ‘small’ technologies that are not fully adopted by users or that do not contribute significantly to household incomes can be a risky business (Place et al., 2007). Thus the strategy should be to identify multi-dimensional impacts (especially poverty-related ones) from large-scale technological or policy successes where impact practitioners are fairly confident about the positive net economic benefits to adopters (and consumers), but uncertain about the economic and social impacts on non-adopters and about environmental impacts.

Improving the power of disaggregate economic rate of return assessments to inform about poverty. Making economic rate of return assessments more informative about impacts on poverty in order to begin to approach multi-dimensional impact assessments could probably be achieved at a relatively small cost in terms of increased fieldwork. The economic analysis of farm-level productivity change shows that rural poverty-related benefits are composed of three main sources: (1) direct benefits to technology adopters, (2) reduced prices to net buyers of food, and (3) increased employment and wages for agricultural workers (Datt and Ravallion, 1998). Although economic rate of return assessments are better positioned to comment on direct benefits to adopters, indirect benefits to consumers, and the distribution of benefits between producers and consumers, other sources of benefits should not be regarded as off limits in such studies. This subsection describes four priorities for making economic rate of return assessments more investor-friendly for showing impacts on poverty.

Firstly, more in-depth understanding of adoption can provide valuable information towards making impact analyses multi-dimensional. The minimum requirement for economic rate of return analysis is to estimate the level and rate of adoption/dis-adoption over time. Collecting information on other adoption-related parameters, such as characteristics of the recommendation domain and of adopters and non-adopters, can help enrich the impact analysis and the presentation of economic impact results by stratified groups of adopters (e.g., by level of income, sources of income, geographic location, and other factors.).

Secondly, information on the importance of estimated net benefits in the total household incomes of adopters can be a valuable supplement to conventional economic estimates of cost–benefit parameters in economic rate of return assessments. As evidenced by Davis et al. (2007), drawing on FAO’s rural-income generating activity database, household income data that formed the basis for Chapter 3 of the 2008

World Development Report (World Bank, 2007) appear to be more available in developing countries than was conventionally thought. If they exist and are reliable, national survey data can be tapped to develop proxy models of household income that form the basis for income estimation in 5–10 page reduced modules that, in turn, can be appended to adoption surveys (Tschirley et al., 2000; Walker et al., 2004). Estimates of the relative importance of different sources of earnings in total household income can also provide guidance on prospects for documenting wider impacts on different groups in society.

Given the tendency of households to underreport income in surveys, calculations on relative importance are likely to only be upper-bound estimates. But such estimates can still be informative in an order of magnitude sense and are better than no estimate or no discussion of the contribution that a technology or policy has made to household income.

Thirdly, conversations with users on how the adoption of a technology has affected their lives can be a persuasive source of information on expected poverty impacts. An example related to poverty and gender is given in Box 3.5. Such human interest stories also help to strengthen the links between epIA and public awareness as discussed in the sixth section of Chapter 4 (The communication and dissemination of EpIA results).

Fourthly, analytically reasoned expectations about the impact of a technology or policy on the various dimensions of poverty can also be a welcome extension of economic rate of return assessments. Expectations on the performance of the technology or policy can be assessed against a checklist of poverty indicators taken from the sustainable livelihoods framework discussed in the subsection in Chapter 2 on 'Livelihood approaches in multi-dimensional impact assessments' (see page 20) (Adato and Meinzen-Dick, 2007), or from an empirically derived compilation (Lipton and Longhurst, 1989). For example, some relevant aspects of production-related poverty include country poverty, spatial commodity poverty, poverty of the recommendation domain, poverty and adoption, poverty and gender, and growing season poverty (Walker, 2000). Many developing countries now have geographically detailed poverty maps that build on World Bank-assisted Living Standards Measurement Surveys (LSMS) and draw on census data for extrapolation. Research institutes such as the International Livestock Research Institute (ILRI) have also invested significantly in poverty mapping to bring prospective technological change into focus with regard to one of the major MDGs. Most future epIAs should be poised to describe the geographic incidence of absolute poverty in the recommendation domains of the selected technology vis-à-vis other regions in the same country.

Other dimensions include commodity-consumption poverty, labor poverty (as reflected in employment and wages for agricultural workers), and stochastic (transient) poverty that embraces risk benefits and vulnerability. Decision-makers attached to projects should discuss the impact of the technology or policy being promoted by that project on the various aspects of poverty. It is important to identify which aspects are being made worse and which are experiencing only negligible impact. This information will provide a more nuanced and balanced story about a project's impact on poverty (Walker, 2000).

Again, it is important to recognize that comparative statements about the average levels of poverty intensity relating to a specific aspect of poverty may say little about the total impact on poverty. For example, the number of poor people is often greater in regions with better agricultural production than in more marginal regions, even though the latter are characterized by significantly higher average rates of poverty (World Bank, 2007). Multidimensional impact assessments may find that poorer households are more likely to adopt some technologies, such as tree fallows, than more cash-intensive technologies, such as fertilizer and maize hybrids, compared to richer households (Place et al., 2007). This tendency does not necessarily mean that a technology that requires few purchased inputs will have more of an impact on poverty than a more cash-intensive technology. The greater the diffusion of a less location-specific technology and the more favorable indirect effects there

Box 3.5. The social impacts of expanding the soybean area and varietal change in Southern Nigeria

"Soybean is my husband because it gives me money to take care of my problems and to pay for my children's school fees and hospital bills," a female farmer in Benue State said during a focus group interview after farmers were asked what income benefits they had gained from growing soybeans. She continued, "I plant soybeans to have money. Sometimes I can harvest up to 10 bags or more. Then I sell some and keep some for my daughter who is in college at Yandev. When she comes home we sell some and she uses the money to buy books and pay her school fees. She will get a good husband in town because men nowadays don't want to marry illiterate women… I have also bought many other things that most people would like to have… You see why I say soybean is my husband. I can't abandon it for anything else. How can you leave your husband?"

This is one of several comments gathered by a study on the level of adoption and impact of soybean on farm households in Nigeria's southern Guinea savanna. One of the most interesting facets of the results distinguished four common alternative patterns of gender responsibility:

- Soybean considered as a man's cash crop with women confined to the production of food crops. In some cases, however, soybean was also considered as a woman's crop in households where men were engaged in producing other cash crops such as tobacco, oranges, and rice, or had other off-farm activities
- Soybean cultivated by both men and women, but on separate farms
- Soybean production as a family enterprise with men and women performing different but complementary tasks on the same plot
- Women managing soybean farms without men's assistance (in a few cases), as independent decision-makers and effective farm managers.

The female farmer from Benue quoted above appears to have been in this last group. The comments of three other farmers follow.

"I have achieved a lot with soybean. Any Tiv man would like to build a zinc-[roofed] house in his compound. Three years ago, my house was burnt down by fire during the harmattan season. I lost everything I had… I cultivate more than 100 lines of soybean in three different places. I got enough money from just one harvest of soybean! I put a zinc roof on my house. I also bought a big radio cassette, mattresses, and many other things. Now I can keep my valuables in my zinc house and no fire will destroy them." (Male farmer)

"…In our Tiv culture you need to marry more than one wife to look after your farm work. I married my second wife with the money I got from my soybean farm. Now my wives complain that the farm work is too much for them. Now I need more hands for my farm work. This year, I will sell all my soybean to pay bridewealth for a new wife. It is only soybean that can give you enough money to satisfy your needs…" (Male farmer)

"I can harvest between two to three bags, sometimes four or five. I normally sell part of my soybeans before Christmas and keep some to sell later when I need money or when there is hunger in around June. Last year I sold two bags of soybean. I bought one nice 'wrapper' [cloak] for my mother, one for myself, and clothes for my children. I have also bought one goat, and I now have up to 8–10 goats. (Female farmer in Mbalav-Aliade village)

Source: Sanginga et al. (1999)

are (e.g., lower prices to net food buyers and higher waged employment prospects for agricultural laborers), then the higher is the potential to compensate for the substantially lower poverty impact on adopters.

The variation on this relative versus absolute theme also applies to consumption. The size of poverty impacts depends as much if not more on the importance of commodities benefiting from technology in household budgets than on the relative percentage of benefits captured by the poor (Walker, 2000 as constructed from Pinstrup-Anderson, 1977). Although average poverty rates only tell part of the story, impact practitioners are still encouraged to generate reasoned hypotheses about the impact of a technology or policy on poverty through economic rate of return studies.

Documenting food security and other poverty indicators. Food security is an MDG and is one that agricultural research contributes to through effects on incomes, prices, and consumption effects. The impact of agricultural research on food security is not measured explicitly in most epIAs, but is implicit in documented effects on household or aggregate-level production, consumption and the cost of food. Since 'security' is a state of well-being experienced by individual subjects, defining an absolute and objective indicator of 'food security' can be a challenging task. On the other hand, how a technology affects farmers' perceptions of household food security can be readily elicited from a survey or focus group discussions. Thus, food security is often mentioned in participatory subjective assessments as a major component of poverty (Adato and Meinzen-Dick, 2007). The seasonal duration of access to food has effectively figured as an indicator of poverty in both technology and policy-related epIAs (Barahona and Levy, 2007; Minten and Barrett, 2008). (The former features an innovative combination of participatory research with statistical sampling techniques to generate national statistics.) Food security warrants greater attention in both economic rate of return and multi-dimensional impact epIAs.

There are several alternatives to money metric measures to measure impacts on poverty (Place et al., 2007). Wealth rankings in particular, and subjective poverty assessments in general, are two of the most common in the literature on participatory R&D (Chambers, 1994). Over time, participatory approaches have become more sophisticated in combining qualitative and quantitative data and at least one quite rigorous example of quantifying qualitative information is available (Hargreaves et al., 2007). Contrary to Chamber's warning that participatory approaches should not be made standard and routine so that their local character is preserved, researchers have tried to increase the comparability of results across space and over time and to use statistical principles for applying these methods to improve their capacity for generalizing (Barahona and Levy, 2007; Hargreaves et al., 2007).

The standardization of measures of poverty to improve both internal and external validity is a desirable objective (Kanbur and Shaffer, 2007). However, responding to a 'how much' question in terms of poverty alleviation (comparable to Moyo et al., 2007) still seems only a remote possibility using participatory methods whose strengths are geared towards responding to indicating the incidence of non-poor households adopting, poor households adopting, non-poor households not adopting, and poor households not adopting in a localized context. In a targeted policy sense – which however is not strictly applicable to technological change because the rural and urban poor can benefit from lower food prices without being adopters – non-poor households adopting represents an inclusion error and poor households not adopting is equivalent to an exclusion error. Because technological change has to be accessible to all households in an open economy and because of the potential for indirect poverty benefits from adopting non-poor households (e.g., having larger marketable surpluses and influences on prices for the urban poor), inferences about poverty from the relative size of the inclusion error in a localized setting are not actually that meaningful.

Many dimensions of poverty are far removed from the influence of agricultural research. Ideally, impact assessments would want to identify dimensions that are independent of production and consumption poverty to determine how agricultural research impacts on these dimensions. For example, the analysis of national survey data in Papua New Guinea showed that the incidence, depth, and severity of poverty was lower in female-headed than in male-headed households (Kanbur and Shaffer, 2007). However, village participatory poverty assessments suggested that women as a group were worse off than men as a group. The two dimensions of poverty of their excessive workloads and restricted decision-making authority often disproportionately affect women. The identification of these attributes is valuable, and agricultural research can address both of these aspects of poverty in the form of commodity research on labor-saving rustic processing techniques and improved varieties for the management of women's crops. But only a very small part of agricultural research is positioned to generate favorable consequences on these two aspects. As discussed above, agricultural research is expected to impact favorably on income and consumption poverty, but the potential for positive consequences on other aspects that are not associated with income and consumption poverty need to be evaluated on a case-by-case basis.

Documenting deeper and long-term poverty-related consequences. Multi-dimensional impact assessments with a focus on poverty usually respond to the two broad issues of:

- To what extent have benefits accrued to the poor?
- Can deeper, long-term poverty-related consequences be persuasively documented?

Judging from the results of the five micro-oriented assessments in Adato and Meinzen-Dick (2007), the first question is more readily and rigorously answerable than the second at the micro-level. Estimating the determinants of technology adoption is one of the common points of departure for drawing inferences about the distribution of benefits and poor people's share in the direct benefits. However, the long-term poverty impacts if and when such technologies are scaled up have yet to be documented persuasively. This is the case even for some of the large-scale documented success stories of the productivity-enhancing technologies of the Green Revolution type (Hazell, 2008). Nevertheless, serious attempts have been made to analyze the aggregate adoption and productivity impact of germplasm improvement efforts on major food crops to assess the long-term multi-dimensional impacts on world food production, prices, area planted, consumption, and the incidence of malnutrition among children (Evenson and Rosegrant, 2003). More in-depth analysis through household studies of the long-term poverty-related consequences of research-derived improved technologies and policies should target large-scale success stories in the places where adoption is known to be widespread. If sampled and selected appropriately, results from these in-depth, micro-level studies could then serve as a basis for more credible and useful estimates of large-scale poverty impacts from a particular innovation within the region of adoption. Good estimates of the scope and extent of adoption are, however, fundamental.

Demonstrating that technological change has led to the accumulation of assets and to longer-term impacts related to food security and poverty is a more difficult proposition. Even so, an exceptionally deep impact was uncovered when assessing the impact of new maize hybrids by Bourdillon et al. (2007). The study looked at an area of Zimbabwe where Hoddinott and Kinsey (2001) had documented that young children of households who had lower livestock holdings during a severe drought in the early 1990s suffered adverse nutritional consequences compared to children from richer households. The study showed that the adoption of the maize hybrids increased household income that was subsequently partially invested in livestock, which in turn helped to maintain adequate diets during times of drought. This study's findings illustrate a potentially important point for impact practitioners. In most instances, having to show that the consequences of the technology have a direct role in contributing to asset accumulation and long-term impacts on nutritional status may be difficult. But the less taxing, more realistic, and more conventional approach is to follow the Zimbabwe example in going step-by-step from assessing the impact on production, to assessing household income, and then on to assessing asset accumulation.

In a specific context (when a research intervention does not contribute directly to one of the deeper impacts), practitioners may find no relationship in their datasets between income and asset accumulation and nutritional status[13]. At least in such a case, practitioners have a literature to fall back on to place their findings in context, and their assessments have the added advantage of contributing to that same literature. The hope that technologies are imbued with special indirect effects that can always be rigorously and explicitly documented is unlikely to be realized. Sequential documentation that is built on flowcharting the nature and order of consequences, as described in the second section of Chapter 2 (Documenting consequences along the impact pathway) and the subsection further on in Chapter 2 on 'Outcome mapping …' (see page 16) seems like a more viable alternative than trying to include the use of the technology in every impact-estimating equation. Documenting interesting, important, and relevant effects further along the impact pathway is desirable, and such estimation should draw as much as possible on well-established disciplinary knowledge.

Good practice 3.27. Practitioners should pay more attention to documenting poverty consequences beyond estimates of economic rates of return especially in those contexts where the research intervention's potential to alleviate poverty is more uncertain. The uncertain contexts are especially the transforming economies and urbanized economies – particularly those outside South Asia and sub-Saharan Africa – and to research interventions characterized by small net present value.

Good practice 3.28. Ways to enhance the poverty content of economic rate of return assessments include estimating the income effects for adopting households, the elicitation of human interest stories, and the qualitative assessment of key poverty aspects that condition consequences in the research intervention and the context of interest.

Good practice 3.29. Food security as an indicator of poverty warrants more attention in both Stage I and Stage II epIAs. How much of research intervention's estimated net present value accrues to the poor is a relevant question in both economic rate of return and multi-dimensional impact assessments.

Good practice 3.30. Impact practitioners should exploit the increasing availability of national rural income and household consumption expenditure surveys to provide a basis for documenting effects on income and consumption poverty.

Good practice 3.31. In investigating deeper poverty-related consequences in multi-dimensional impact studies, practitioners should carefully weigh poverty dimensions and focus on those that the research intervention has the best chance of influencing.

Good practice 3.32. The distinction between average poverty incidence rates and the absolute numbers of the poor is an important one in discussing and reporting the results of Stage I and Stage II epIAs. Practitioners should pay as much attention to absolute numbers as to the average poverty rates of beneficiaries. For example, a density measure such as the number of rural poor people per square km, may be more informative than the average incidence of poverty or the head count index.

Key references

Hossain M., Lewis D., Bose M.L., and Chowdhury A. 2007. Rice Research, Technological Progress, and Poverty: The Bangladesh case. • A most comprehensive, yet incisive and well-reasoned, poverty-related multi-dimensional impact assessment that drew on years of field research.

Moyo S., Norton G.W., Alwang J., Rhinehart I., and Deom C.M. 2007. Peanut research and poverty reduction: Impacts of variety improvement to control peanut

13 For example, very small income elasticities in the demand for calories are a common finding in the analysis of household cross-sectional data (Behrman and Deolalikar, 1990).

viruses in Uganda. • Although not an epIA, this research presents a clear example of how absolute poverty impact can be quantified in an economic rate of return assessment.

Place F., Adato M., Hebinck P., and Omosa M. 2007. Impacts of agroforestry-based soil fertility replenishment practices on the poor in western Kenya. • An excellent example of the use of mixed methods in a multi-dimensional impact assessment to analyze poverty-related consequences.

4. Institutionalizing and Managing EpIA

This chapter discusses the institutional aspects and management of ex post impact assessments. The first three subsections focus on conceptual issues related to accountability and epIA in public-sector agricultural research institutes. The first section focuses on the role played by epIA in evaluating research. We underscore the fact that a primary emphasis on accountability should generate as a by-product considerable learning about what has and has not worked. The second section sheds light on another dimension of accountability: the desirability and feasibility of undertaking epIA on newer research initiatives. The third section addresses accountability from the donor perspective.

The first three subsections cover management issues relating to epIA. The fourth section discusses the funding of epIA whilst the fifth section delineates the tradeoffs between external and internal assessments and also examines the organization of epIA in a research institute. The sixth section brings out the complementarities in investing in public awareness and epIA and the importance of disseminating the results of epIAs.

Accountability, learning, and where EpIA fits into research evaluations

The main objective of epIA is to increase accountability. In the CGIAR, donors want to see evidence of impact to justify their funding of the IARCs (Raitzer and Winkel, 2005). Generally, epIAs play a supporting role in accountability to periodic comprehensive program reviews including EPMRs. The availability of several epIAs from the same time period as an EPMR should enhance the prospects for a favorable review.

Learning is a potentially important objective of epIA, and accountability can contribute to learning. But, unlike accountability, there are multiple mechanisms for learning in a research setting, with epIA being only one amongst several such instruments as rapid rural appraisals and diagnostic surveys. However, although the contribution epIA can make to improving accountability is well recognized, its potential to play a positive role in research planning and decision-making has yet to be widely appreciated.

One difficulty associated with using epIA for 'real time' institutional learning is that it mostly involves assembling information on investments made and outputs adopted from a decade or more previously. With fading institutional memories, the relevance of epIAs for informing current and institutional organization and management can be a contentious issue and largely depends on the stability of the mandate of the research institute, the dynamics of change in the external environment, and the generalizability of individual epIA findings to thematic research areas. One way to address the problem of time lags is to start epIAs sooner, although when beginning epIAs in the early stages of adoption, there is a risk of generating imprecise estimates for accountability. This tradeoff was addressed in the subsection in Chapter 3 on 'The uptake of research interventions' (see page 31).

The linkages among priority setting, monitoring and evaluation, and epIA are described in Figure 4.1, showing the forward and backward flows of information. Focusing first on the forward linkages (going clockwise), effective priority setting contributes to the formulation of projects that are monitored and evaluated. The related outcome mapping and participatory impact pathway analysis were discussed in Chapter 2 as tools for monitoring and evaluation. Rigorous monitoring and evaluation is not a sufficient or even a necessary condition for making practical impact, but research projects that have been adequately monitored and evaluated tend to have a higher likelihood of having favorable outcomes attributed to them in impact assessments. In turn, the results from epIAs

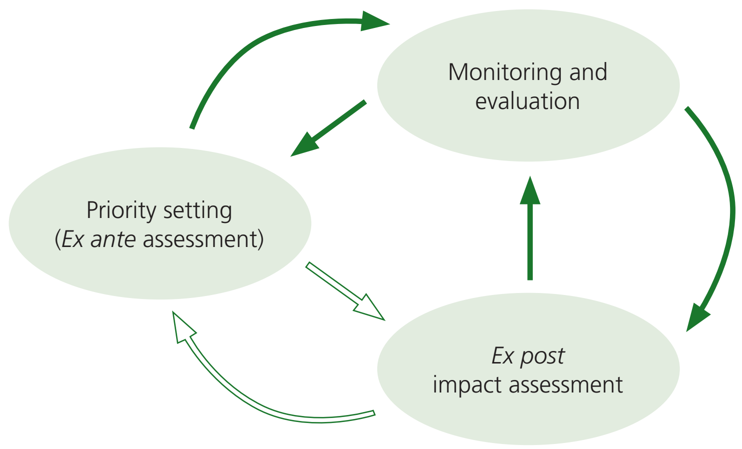

Figure 4.1. Linkages between priority setting, monitoring and evaluation, and epIA.

should feed back to the next round of priority setting. To date, there are few good examples of where this has been done effectively.

The flow of information from epIA to priority setting is one of the weakest links (designated by the longer white arrow in Figure 4.1) in the chain of this triad of activities that define research evaluation. The main reasons for the disarticulation between epIA and priority setting are:

- Insufficient investment in epIA to inform priority setting
- Priority setting being structured in such a way that information from epIA on levels and rates of technology adoption, net benefits per unit adopted, and inferences about probabilities of technological success by programmatic research area become difficult to incorporate into a priority-setting exercise founded typically on specialist scientist opinion. (Most scientists are equally optimistic about the prospects for and the importance of success in their own research areas)
- Time lags may render the value of information less pertinent to today's challenges (as discussed above).

Turning to the counterclockwise arrows representing smaller flows of information in Figure 4.1, the results of monitoring and evaluation should have implications for setting priorities and for formulating new projects, once again assuming that such information can be accommodated in a priority-setting exercise. *Ex ante* studies should inform hypothesis testing whilst formulating epIAs, which in turn should confront the predictions of *ex ante* assessments.

The analytical description in Table 4.1 helps to further clarify the role of learning in *ex ante* and *ex post* impact assessments (Boardman et al., 2001). *Ex ante* assessment contributes information on which to base the decision of whether or not a research project should go ahead. Comparing the results of *ex ante* and *ex post* assessments is useful for learning about the predictive accuracy of *ex ante* impact assessments and particularly the assumptions on which they were based. Such comparisons are rare in the literature (see Box 4.1 on page 62). In agriculture, bovine somatotrophin (rBST) is arguably the technology that has received the most attention during the past 30 years. Scores of ex ante assessments were conducted on rBST in the 1980s. Comparisons of *ex ante* predictions with *ex post* realities show that many studies overestimated the expected adoption of this bovine growth hormone and the size of the effects on milk production (Barham et al., 2002).

In Table 4.1, 'learning about the value of a project' refers to accountability. The *ex post* assessment of impact scores highly for enhancing accountability. But epIA can potentially contribute information on the design of technologies for future similar projects in the same research area and to inform on

Table 4.1. Potential for learning from the impact assessment of a project.
Adapted from: Boardman et al. (2001)

Value	Type of assessment *Ex ante*	*Ex post*	Comparing *ex ante* and *ex post* assessments
Resource allocation decisions for a project	Yes – helps to select best project or make 'go' versus 'no-go' decisions, if accurate	Too late – the project is over	Same as *ex post* analysis (too late)
Learning about actual value of a project (accountability)	Often poor estimates – high uncertainty about benefits and costs	Excellent – although some errors may remain. May have to wait a long time to study	Same as *ex post* analysis (excellent)
Contributing to learning about actual value of similar projects	Unlikely to add much	Very useful – although may be some errors and the need to adjust for uniqueness. May have to wait a long time for project completion	Same as *ex post* analysis (very useful)
Learning about omission, forecasting, measurement, and evaluation errors in cost–benefit analysis of a project	No	No	Yes, provides information about these errors and about the accuracy of cost–benefit analysis for similar projects

decisions about allocating resources in the assessed research area as against assigning the resources to other research areas.

Boardman et al. (2001) highlight the potential of epIA to contribute to learning about the actual value of similar projects: '*Ex post* analyses not only provide information about a particular policy intervention but, more importantly, about similar interventions as well…Furthermore, *ex post* analyses potentially contribute to learning by political and bureaucratic decision makers, as well as policy researchers, about whether particular kinds of projects are worthwhile…The amount of societal learning from...*ex post* analyses depends on the generalizability of a particular project' (Boardman et al., 2001, pp.4-5).

Although the above discussion originates from the perspective of specific government interventions and policies, it also applies to agricultural research. The potential for learning from epIA is also likely to be affected by the sampling strategy in selecting research-derived policy and candidate technologies for analysis. A random selection of projects in a thematic area or a random selection of thematic areas may be more informative for learning than the success story strategy advocated above in Chapter 2 for promoting accountability.

Whether or not researchers and other stakeholders act as a result of what is learned from assessments usually depends on there being incentives in the larger research environment. If the results from past epIAs across research institutes and research areas are not reflected in concomitant changes in budgetary allocations, then what has been learned will not usually be translated into effective action. Such dissonance in epIA results and in budgetary allocations encourage 'free riding' in the sense that institutes that do not carry out epIAs are not penalized and those that do are not rewarded because donors need accountability information for the system as a whole. In effect, some can ride on the coat tails of others as long as a minimum threshold level of epIA is carried out for the system as a whole.

Under conditions where the results of epIA are divorced from budgetary allocations, the incentives for the conduct of epIA will also be dampened. The analysis of reward structures for investing in epIAs is beyond the scope of these guidelines. Suffice it to say that incentives for carrying out epIAs

Box 4.1. ***Ex ante*** **predictions and** ***ex post*** **reality in generating and delivering vaccines against Newcastle disease**

In the 1980s and 1990s, ACIAR invested in several projects to develop vaccines to combat the Newcastle disease virus, a chronic source of mortality amongst village chickens in the developing world. The projects were technically successful, generating several innovative vaccines designed to suit village circumstances. The first *ex ante* assessment on one of the vaccines estimated large net benefits equivalent to an NPV of A$144 million in several countries in Southeast Asia (Johnston and Cumming, 1991).

Two subsequent *ex ante* assessments in 1998 and 2005 showed lower (but still attractive) levels of benefits, since early adoption did not match expectations. A later program review noted that the evidence of adoption was scanty and concluded that the challenges of uptake in smallholder low-input systems may not have been fully appreciated (ACIAR, 2006). The review recommended that ACIAR undertake economic, community, and institutional research to better understand the poor levels of adoption.

The prospects for these vaccines now seem brighter in southern Africa as several governments are supporting their distribution, admittedly still on a small scale, via public-sector extension. In addition, ACIAR-supported projects have subsequently been much more grounded in the reality of village chicken production, with important diagnostic work carried out to adapt the technology to village circumstances in both the research and extension components (Alders et al., 2005). This experience reinforces the point that there are diminishing returns to *ex ante* assessments, and the opportunity costs of not engaging in *ex post* (e.g. early acceptance studies) analysis can be high.

Sources: Johnston and Cumming (1991); Alders et al. (2005); ACIAR (2006)

and the disincentives for not undertaking them are likely to be context specific and depend as much on perceptions as on empirical facts. It is fundamental to these guidelines that economics teaches that an adequate incentive structure is needed to ensure the health of both epIA-related accountability and learning.

Good practice 4.1. EpIA should be viewed as an integral part of evaluating research. The forging and strengthening of links between epIA and priority setting in general, and between epIA and *ex ante* impact assessment in particular, are themselves priorities.

Good practice 4.2. Although accountability is the main objective of epIA, impact practitioners are encouraged to view epIA as a dual-purpose activity that can also contribute to strategic learning, particularly with respect to the value of the research in economic, social, and environmental terms. EpIAs should inform about the implications of the results for kindred technologies and policies in the same research area.

Good practice 4.3. Improving the incentive structure for carrying out epIAs should result in both enhanced accountability and learning. Indeed, research on the structure of incentives for epIA is warranted from the perspectives of both international and national agricultural research institutes.

Key references

Fuglie K.O. 2007. Research Priority Assessment for CIP 2005–2015 Strategic Plan: Projecting Impacts on Poverty, Employment, Health, and the Environment. • ‘Routine’ priority setting that draws on information in past epIAs and casts results in terms of the MDGs.

Kelley T.G., Ryan J.G., and Patel B.K. 1995. Applied participatory priority setting in international agricultural research: Making trade-offs transparent and explicit. • A detailed and thorough priority setting exercise that analyses the interactions between the *ex ante* and *ex post* assessment of impact.

Gestation lags, new initiatives, and epIA

Technological change attributed to agricultural research has been aptly called 'slow magic' (Pardey and Beintema, 2001). It takes time. The development of technology hinges on basic research discoveries and requires investment in applied, and/or adaptive research. In turn, diffusion entails the adoption and disadoption of technologies and usually falls well short of covering 100% of the target area. Successful agricultural research generates both benefits and costs, with success being a product of many factors. These include, at times, good luck in making unexpected and fortunate discoveries. Depending on the context, the costs of transferring the technology are often substantially greater than the costs of the research. The technology transfer costs may continue for many years after adoption has started, particularly for maintenance research (Marasas et al., 2003). However, for most successful research interventions, the benefits will dwarf the costs by several orders of magnitude.

Profiles of technology generation and diffusion typically show a gestation lag of around 10 years from the start of applied research to the release of an intervention for adoption (Alston et al., 1998). Technological advances in the conduct of research, such as marker-assisted selection, participatory plant breeding, and computer modeling, may reduce this gestation lag in the future, but thus far such gains have not translated into significant practical impact in shortening the time it takes to go from the laboratory to the farmer's field.

Because of the long gestation time, the hallmarks of successful agricultural research are the continuity of research mandates and the stability of research teams (Eicher, 2001). In the past 15–20 years, the mandates of agricultural research institutes, particularly those in the CGIAR, have become more dynamic and extensive because few of the 'old' areas have been divested but new areas have been added on. For example, in one IARC, these 'new' areas now represent 30% of total expenditure (Fuglie, 2007). The 'new' areas are not as productivity-oriented as the 'old' areas and it is considerably more difficult to set priorities for and conduct epIAs on them. If programmatic areas change every 4–5 years, then agricultural research is unlikely to pay productivity dividends and epIAs, if conducted, will have limited relevance to either accountability or learning.

Assuming that a new initiative is equivalent to a new research area, then an epIA will not be expected to take place for some time, perhaps for as long as 10 years into the implementation of a program. This is the time it takes for most major field crops to reach the stage of varietal release. It is also the gestation period of the development and testing of past resource management technologies, such as the vertisol technology options or alley farming in the 1980s. Policy-oriented research could well have a shorter gestation period, but the prospects for uptake and influence may not be nearly as bright as for technology-oriented research.

This 10-year gestation lag is subject to several caveats. Firstly, more fundamental research that is the basis for new areas may be developing so rapidly that technology development exceeds expectations and occurs sooner. Carrying out epIAs sooner than expected would be a bonus for these 'rising stars'. Secondly, the initiative may be newly organized, but the research areas it encompasses may not be new to the CGIAR. This reservation may apply to research areas in newer initiatives such as the Challenge Programs. CGIAR donors have invested in some component research areas in these programs for several years and even decades. The organization and orientation may have changed, but the disaggregated research area may have stayed the same.

In this case of waiting for technologies to mature, it may be in the interest of new initiatives to invest in epIAs that focus on a successful technology undertaken by another program, but that fit squarely within their own area of research. For example, the HarvestPlus Challenge Program could invest in a collaborative epIA with CIMMYT on the impacts of quality protein maize on human nutrition. Such a technology could not be claimed by the new initiative, but the epIA could be substantially more informative for the prospects of its expected potential technologies than the predicted effects from *ex ante* assessments.

Furthermore, 'waiting' 10 years or more before initiating an epIA does not imply that no evaluations of research and uptake need take place. In the interim, other mechanisms, such as annual review and planning meetings, internal CCERs, EPMRs, outcome mapping, and early acceptance studies, can be deployed to monitor and evaluate research so that mid-term corrections can be undertaken.

Good practice 4.4. A reasonable expectation on the desirability of carrying out an epIA in a truly new research area should be about 10 years after the start of the program in a research institute that focuses on applied research. Other monitoring and evaluation tools are more appropriate than epIA during the 'gestation' phase of a new initiative.

Good practice 4.5. New research programs containing research areas previously invested in by the CGIAR are encouraged to conduct epIAs on technologies of earlier programs of the same research genre while they await the generation of mature technologies.

Meeting multiple donor demands for accountability

The demand by donors for results from epIA is a recurring theme in these guidelines. A recent survey conveys the multifaceted nature of donor demands (Raitzer and Winkel, 2005). Their report contains several 'take home' messages:

- Donor representatives want evidence for general accountability to support their budgetary allocations to the CGIAR
- Donor representatives are optimistic about the prospects for future impact from the CGIAR's research areas, irrespective of an area's past impact
- Donor representatives want to see impact on a comprehensive set of consequences that are closely related to the MDGs of food security, poverty alleviation, and environmental sustainability.

Of the above three goals, donors are most interested in the impact on alleviating poverty.

To respond to the needs and demands of donors, these guidelines propose a two-stage sequential approach (the minimalist good practice model) to carrying out epIAs. For technology-oriented research, the first stage (Stage I) should focus on the technologies' impacts on their users. This can then be followed by a second-stage (Stage II) inquiry that documents significant effects further along the impact pathway.

Aspects relating to Stage I and Stage II epIAs warrant more discussion and some reiteration from the perspective of donor demands. The minimalist good-practice model seems deceptively simple, but its 'looks' are deceiving. Moreover, the need for the more comprehensive impact assessments desired by donors requires a skill set that is considerably beyond that needed to carry out a simple good-practice epIA.

Because it is usually difficult to establish the subject matter focus for Stage II epIAs, the most important element in the skill set needed to carry out such assessments is intuition about the feasibility of studying the deeper consequences of a technology. Although EpIAs of agricultural research are not overly complex, researchers with a doctorate in the social sciences or an allied agricultural field will usually find them difficult to carry out because they do not have the required specialist knowledge to assess multiple impacts associated with specific disciplines and sub-disciplines in Stage II epIAs. The donor demand study suggested that donors often want epIAs that stretch the limits of human capability. With limited human capital this can result in impact assessments made up of unconvincing story telling.

The aforementioned review of meta-analyses of the impact evaluation of social programs (Center for Global Development, 2006) serves as a guide for the appropriate focus of epIAs. Although all programs will indirectly contribute to the MDGs, almost none of the reviewed evaluations 'failed' for not addressing the MDGs directly. Only a few of the evaluations were persuasive; most of them lacked internal validity as 'they could not distinguish impacts on the relevant population that were specific to the program from changes attributable to other factors' (ibid. p. 17).

Most of the evaluations were judged to be inadequate because they could not provide persuasive evidence on a highly-focused question. The focus of epIAs on technologies and programs for policy change should be the same as for assessing the impact of social programs. The main questions to ask are whether or not the selected technology or program (1) had the intended impact on the target beneficiary population (Stage I epIA); or (2) had a beneficial effect on one or two relevant, hypothesized, and specific consequences further along the impact pathway (Stage II epIA). As the definition at the beginning of this paper suggests, large unintended positive and negative consequences could also belong to the possible candidate impacts being assessed.

Good practice 4.6. To meet donors' primary demand for general accountability on what has worked, and their multiple secondary demands to see impact related to achieving the MDGs, these guidelines propose a two-stage, sequential approach to epIA. For technology-oriented research, the first stage should focus on the impacts on the users of the technology. The prospective second-stage inquiry should document significant effects further along the impact pathway. For accountability to donors, impact practitioners should focus on the purposive sampling of perceived success stories rather than random sampling. The latter is more apt if the primary purpose is to learn from the level of impact.

Good practice 4.7. The carrying out of epIAs is a dynamic process requiring a versatile skill set, including a good grounding in general agriculture and statistical analysis.

Good practice 4.8. In terms of the MDGs, epIAs on the impacts of agricultural research should be held to the same standards as epIAs in other sectors. Successful epIAs are those that focus on the purpose of the intervention and on a small subset of intended and unintended consequences along the impact pathway. As much as possible, purposes should be crafted in terms of intermediate outcomes that are characterized in the literature by well-established empirical linkages to the achievement of the MDGs.

Building EpIA into project design and funding for EpIA

The building of epIA into agricultural research project design is not a straightforward task, particularly in the context of the CGIAR. For restricted core projects, the earliest time when successes can be documented will usually be 5–10 years after the usual 3-year project ends. Even if success is achieved and the project's technology is adopted, donor representatives may change and sometimes they will have little interest in documenting successes overseen by their predecessors. Thus, at the end of a research project, about the only determination that can be made for certain is whether or not a project has been technically successful and whether or not this technical success has been translated into early acceptance by users. For projects with promising initial rates of adoption, an epIA can be planned when a next phase or variation of the project is subsequently funded. As argued earlier in this chapter, effective epIAs are more likely to take place on projects with very specific purposes and intended consequences.

In research areas that are funded serially from unrestricted core resources – a declining possibility nowadays – support for epIA could, in principle, be funded from program funds. This scenario applies to plant breeding programs where the main form of epIA is likely to be the evaluation of programs. But administratively, it is probably easier to do this at the level of the CGIAR centers as a whole rather than at the program level, where the competition for the recurrent budgets for operational expenses is usually severe. Also, carrying out epIAs at the center level will result in economies of scale. In the next section, we argue that the locus for conducting epIAs has implications for the credibility of assessments and that this favors taking a center-level approach.

Turning to the issue of the resources needed to conduct epIAs, there are only a few reference points. The General Accounting Office of the United States Government recommends that 3% of total project budgets go to evaluating impact (E. Chelimsky, ex-Director of Evaluation in the U.S. Federal Government, personal communication, 1998). For the IARCs of the CGIAR, a 2.5–3% allocation seems reasonable in both technology-oriented and policy- and management-oriented research institutes. The important caveat here is that the involved scientist(s) are likely to be working on impact evaluation and assessment more broadly. EpIA may be only one part of a broader agenda that includes activities such as *ex ante* impact assessment, priority assessment, establishment and analysis of baseline surveys, and early adoption studies. All of these are building blocks for subsequent epIAs. Expenditure on EpIA, in a strict sense, may comprise only one half to two thirds of the suggested 2.5–3% allocation. This level of effort is comparable with large NARS such as ACIAR and the Brazilian Agricultural Research Corporation (EMBRAPA), both of which devote an estimated 1–1.5% of their total expenditure on epIA (Jeff Davis and Flavio Avila, impact assessment researchers at ACIAR and EMBRAPA, respectively, personal communications, 2008).

The approach recommended in these guidelines means that the amount of funding allocated for carrying out epIAs will depend on having success stories to analyze. If the selection process described in the first section of Chapter 3 (Selecting the research-related technology or policy for epIA) does not result in viable candidate technologies, epIAs will not happen, irrespective of budget being allocated for impact assessment. In this case, the emphasis on impact assessment would shift from accountability to learning as the research institute would be unable to contribute information on ‘what worked’ for the general accountability of the system as a whole.

An important issue in the allocation of resources for epIA in the international and national agricultural research centers is the need to fund epIAs from the centers’ core recurrent resources. As discussed, donors and governments may build impact assessment into specific projects, but because of the long gestation periods, special project funding for generalized epIA is often difficult to procure. Unless funds are set aside for recurrent expenditures on epIA at the level of a research organization or a funding mechanism is available for such work, it may not make sense to establish a scientific presence for carrying out epIAs at a research institute. The scientists who are charged with conducting epIAs will find it difficult to elicit funding for their work. The frequent lack of core resources allocated to epIA also means that funds are often lacking to hire external assessors. This issue needs further investigation to test the hypothesis that the lack of core operational support is a binding constraint on the usefulness, quantity, and quality of epIA.

Good practice 4.9. The time lags in the realization of benefits and donors’ short funding cycles mean that it is often a challenging task to build epIA into project design in agricultural research. Funding for epIA that is not tied to project budgets needs to be reserved and targeted at the level of the research institute and not at the level of the institute’s programs.

Good practice 4.10. Assigning up to 3% of the total annual budget of an agricultural research institute to assessing impact seems like a reasonable amount to allocate. EpIAs should command a sizable share of that amount and should be viewed as a maintenance research activity that warrants recurring expenditure. It should not be subjected solely to the vagaries of special project funding.

Who should conduct EpIAs?

Context looms large in determining the trade-off between external and internal evaluation. In the CGIAR, EPMRs are arguably the main mechanisms for demonstrating accountability. In contrast to EPMRs carried out by outsiders, the vast majority of epIAs conducted in the CGIAR have been carried out using internal human and other resources. With some notable exceptions, epIAs of CGIAR-related research by external consultants are conspicuous by their absence in the peer-reviewed literature. The question to ask is: should external evaluators play a larger role in CGIAR epIAs?

A range of human resources from inside and outside research institutes should be tapped to carry out impact assessments. A balance needs to be struck between credibility and the potential for learning in deciding who should undertake them. Whilst program team members could undertake these assessments to maximize the scope for learning, doing it this way could well lead to a lack of objectivity and to an unbalanced focus on desirable outcomes, particularly where a program leader is forceful and highly committed to demonstrating a technology's success.

At the other extreme, consultants may be employed to conduct the evaluation with only limited inputs available from the scientists involved. Outsider evaluation is presumed to be more 'objective' and insider evaluation is not nearly as credible for convincing the public through press releases (B. Rose, former executive director of the Harvest Plus Secretariat, personal communication, February, 2002). A mixed option could combine the best of both approaches with insider evaluation at the pilot and early acceptance stages, when mid-course corrections can be facilitated, followed by outsider evaluation by consultants in the late adoption stage to provide accountability. It is important that both stages are subjected to peer review.

One alternative is epIA carried out by institute scientists who have not been involved in the project or program under review. A meta-analysis of rates-of-return studies showed that external economic rate of return estimates were significantly higher than insiders' estimates (Alston et al., 2000). The finding that outside consultants tend to be overly optimistic (or that insiders are overly pessimistic) about the impact of the technological change is not surprising. A research intervention is often complex and its effects hard to understand. External assessors largely have to believe what the involved scientists tell them. Favorable assessments enhance the odds of the scientists extending their employment. Another advantage of internal evaluation is that money spent on an outside consultant could go to the operational budget for carrying out epIAs instead.

The best type of outsider evaluation is one that is peer reviewed and subsequently published in the literature. Such evaluations are especially favorable because an institute usually does not have to pay for them and because of the resulting publicity that comes from published results by presumably disinterested specialists. The prototype for this sort of evaluation is Dana Dalyrmple's assessment of the semi-dwarf wheat high yielding varieties in the 1970s (Dalrymple, 1975). The economists at CIMMYT did not have to undertake an explicit economic rate of return analysis as Dalyrmple's estimates thoroughly vouched for widespread impact. Richard Norgaard's epIA on the impact of biological predators on controlling cassava mealybug had the same outcome (Norgaard, 1988).

Interest in academic assessments by outsiders has a downside in that it is limited by novelty. Efforts tend to be concentrated on 'hot' topics, such as the dozens of evaluations on the Puebla Project in Mexico in the late 1960s and early 1970s (e.g., Gladwin, 1976). During the past decade, interest in the impact of biotechnology in developing countries has been intense (Smale et al., 2006). However, because biotechnology has been slow in arriving in farmers' fields, much of the evaluation still focuses on the *ex ante* assessment of impact. Success stories in sub-Saharan Africa also attract outsiders' attention mainly because such stories are perceived as rare. Donors tend to be willing to fund impact assessments of hot topics and areas. In contrast, they are much less willing to fund the impact assessment of more conventional research areas.

Being inside the center but outside the project under assessment implies the establishment of a small, independent evaluation unit, project, or program that reports to the center's Director of Research. This so-called impact czar model has the benefit that program scientists can be involved in evaluations but are insulated from pressures to show very positive results. Staff in the evaluation unit should be respected as scientists and be seen as being impartial to reporting impact across diverse programmatic areas.

Another institutional option is to site epIA in particular, and impact assessment in general, in the social science division or department of an agricultural research institute. This arrangement reinforces the perception that epIA is a research activity and is not being carried out strictly as a compliance audit.

Whatever the organizational structure for epIA, as outlined in the first section of Chapter 3 there should be a fair selection process to ensure that all programs' success stories have a good chance of being chosen for assessment. Typically, such a unit, department, or division will also conduct research on *ex ante* impact assessments and priority setting. Practitioners responsible for carrying out epIAs will, however, have no particular comparative advantage for being involved in monitoring and non-impact evaluation, which should be the responsibility of particular projects and programs.

Regardless of where epIA is located in a center and whether or not it is external or internal, peer review is crucial. Whilst communication in academic journals is desirable, many epIAs are so laden with context that publication is a challenging task particularly for success stories with precedents in conventional research in Asia and Latin America. Having a formal mechanism for peer review at the level of the CGIAR system can be a welcome institutional innovation.

External input is most needed for carrying out the often resource- and skill-intensive Stage II epIAs. Stage I epIAs, as outlined in Chapter 3, are not rapid assessment exercises. A reliable Stage I epIA usually takes months to carry out and requires careful field research. But, assuming that the research institute has invested in social science research, effective Stage I epIA should be achievable with internal resources. Depending on the consequence(s) for analysis, Stage II epIA may need a degree of specialization in one or more disciplines that exceeds the human capital available in even the strongest national and international agricultural research centers.

Good practice 4.11. The carrying out of both internal and external evaluations is important although, in general, epIA conducted by practitioners located inside centers but outside the program have been a particularly viable model at some research institutes.

Good practice 4.12. The external peer review of both internally- and externally-conducted epIA is essential to ensure credibility.

Good practice 4.13. The desirability of employing external evaluation is greater in Stage II than in Stage I epIAs.

Key reference: Shadish W.R., Newman D.L., Scheirer M.A., and Wye C. (Eds) 1995. Guiding Principles for Evaluators. • Ethical standards for the evaluation profession with some discussion of internal versus external evaluations.

The communication and dissemination of EpIA results

A survey of donor demand for epIA suggests several ways to communicate and disseminate the results of these impact assessments (Raitzer and Winkel, 2005). It is recommended that results are cast in easily understood numbers, such as the number of households affected by the technology, number of hectares covered, and the magnitude of net benefits per beneficiary or per hectare adopted. Donors seem to value this type of general information as much as estimates of financial performance.

Donors, policy-makers, and government officials rarely have time to read lengthy research reports, although they say they value the information provided by epIAs. The dissemination of epIA findings in more bite-sized pieces, such as in research briefs, can make the findings more accessible. Research briefs that carry the imprimatur of centralized oversight and advisory bodies, such as those produced by the Standing Panel on Impact Assessment (SPIA) (see examples at http://impact.cgiar.org), may also enhance the credibility of materials that were disseminated originally in a less accessible form by a research institute. Institutional efforts at a system-wide level, which facilitate peer review, can also boost the credibility of epIAs and solve some of the incentive problems relating to epIAs in a complex institutional setting.

The creative packaging of epIA-related briefs may also advance the storytelling potential of successful R&D interventions. This storytelling should report not only the results of epIAs, but also how human lives have been touched by a technology or policy, which scientists were responsible for these developments, and how a technology or policy came about in terms of the processes of science and innovation. The International Crops Research Institute for the Semi-Arid Tropics' (ICRISAT) Food-From-Thought Briefs, published in the 1990s, are a good example of such multi-dimensional, condensed storytelling.

Good public awareness about the CGIAR's work, and especially the work of its centers, complements the carrying out of impact assessments. In the same way, impact assessments complement the work of the CGIAR. Without epIAs, public awareness pays fewer dividends because less textured 'raw material' (results) is available for dissemination. Likewise, raising public awareness will increase the demand for epIA, particularly at the center level.

In terms of accountability, one of the roles of epIA is to generate reliable information that can be used for raising public awareness. Impact practitioners and public awareness specialists need to work together with mutual respect to develop dissemination materials such as press releases, that project the desired message (as per the reliable information) in an accessible and readily understandable form to the intended audience.

Good practice 4.14. Information should be made available to donors on the number of hectares affected by and the number of households benefiting from the technology in addition to estimates of financial parameters such as NPV and internal economic rate of return.

Good practice 4.15. Packaging the findings of epIAs in research briefs will make the findings more widely accessible and available. Emphasizing the diverse aspects of technologies, from the processes of science and innovation to estimating the impact on users and the effects on livelihoods, will help to make a persuasive case for the importance of R&D activities.

Good practice 4.16. Institutional efforts at a system-wide level, which facilitate peer review, can boost the credibility of epIAs and also assist in resolving some of the incentive problems related to epIAs in complex institutional settings.

Good practice 4.17. Investments in epIA and public awareness raising are complementary and synergies can be exploited to improve the performance of both.