Ex Post Evaluation of Economic Impacts of Agricultural Research Programs: A Tour of Good Practice

Mywish Maredia, Derek Byerlee and Jock R. Anderson*

July 31, 2000

Revised version of a paper presented to the Workshop on "The Future of Impact Assessment in CGIAR: Needs, Constraints, and Options", Standing Panel on Impact Assessment (SPIA) of the Technical Advisory Committee, Rome, May 3-6, 2000, FAO, Rome.

---

* Maredia is Adjunct Assistant Professor, Michigan State University, East Lansing, and Byerlee and Anderson are, respectively, Principal Economist and Advisor, Rural Development Department, World Bank, Washington, DC.

Ex Post Evaluation of Economic Impacts of Agricultural Research Programs: A Tour of Good Practice

Mywish Maredia, Derek Byerlee and Jock R. Anderson

1. Introduction

The goal of most public agricultural research organizations is to undertake research and development work that will ultimately improve the productivity and sustainability of the agricultural and food sector. In today's world of scarce public funding and greater accountability, governments, donors and research managers are increasingly demanding assessment of the economic returns to their investments in research.

There is a rich literature of economic impact assessment of agricultural research with evaluation studies that range widely in methods used, and in the breadth and scope of analysis undertaken. The book by Alston, Norton, and Pardey (1995) is an excellent and comprehensive reference on the theory and practice of economic evaluation of research. In this paper, rather than dwell on the many finer and sometimes esoteric points in research methods, we summarize the "state of the art" in ex post economic impact assessment of agricultural research with an emphasis on providing a *practical guide* that can be used by research managers and economists working within a research organization under tight time and resource constraints. The focus is on economic evaluation, recognizing that other papers in this workshop series will provide an overview of social, institutional, and environmental evaluation. Nonetheless, the boundaries in these types of evaluations are blurred and, in this paper, we treat income distributional impacts of research (a social dimension) and touch on economic evaluation of research impacts on natural resources (an environmental dimension).

In a paper of this scope, there will inevitably be biases and omissions. Much of the early part of the paper is focused on conceptual and methodological issues. In later sections of the paper, we turn to implementation issues, especially the implementation of impact assessment as a routine activity. Throughout, our focus is on public research organizations in developing countries, both national and international, and only in one section do we focus on the emphasize the special issues in international research organizations. The underlying premise is that evaluation work in the CGIAR should increasingly be based on evaluation work undertaken in NARSs, with IARCs providing a facilitating and synthetic role.

2. Setting the Stage

2.1 Why do impact studies?

Agricultural research organizations (AROs) world wide undergo various forms of reviews and evaluations, as is appropriate to their role as public institutions. Formal economic impact evaluations, which form an increasingly important element of these evaluations, are carried out for both internal and external clients of the ARO. The first group includes governments and donors, who increasingly require accountability for the funds they invest. Since agricultural research is one of many competing investment alternatives in the portfolio of national governments and international aid agencies, these funding agencies need stronger and clearer evidence of the net social benefits of their investments in research.

The second type of client includes scientists and research managers within the ARO, who require information on economic impacts to provide feedback to their research programs. By generating a better understanding of how technology influences the welfare of agricultural producers and consumers, impact evaluation can improve targeting of research programs and help adjust resource allocation across programs.

Although the same studies can serve both types of clients, there is inevitably a tension between the two. In the current climate of declining funding, more emphasis has been placed on the first objective to demonstrate to external clients the positive payoffs to research investments. However, these studies tend to focus on "winners" and are less useful to research managers in reallocating resources or adjusting programs. Few research organizations evaluate unproductive programs, although such studies may be valuable from the point of view of enhancing research efficiency. Institutionalization of regular evaluation of all applied research should be the long-run goal of AROs.

2.2 Types of impact studies

Economic impact evaluations can be classified into ex ante and ex post evaluations. Ex ante evaluations are undertaken before the project or program is initiated, much as economic analysis is used to assess the value of any investment operation (Belli et al. 2000). Such evaluations are usually undertaken as an aid in priority setting. Ex post evaluations are undertaken after diffusion of a research product has been initiated, to assess actual impacts "on the ground". However, ex post evaluations generate information that is useful for the selection, planning and management of future research programs, such as plausible adoption paths. Although the methods employed are essentially the same, the focus of this paper is on ex post economic evaluation.

Ex post impact studies range in scope and depth of evaluation from simple story-telling and anecdotes, to partial and comprehensive assessment of economic impacts. Evaluations of the "story-telling" types are based on researchers' description and assessment of the importance and value of results from their research (e.g., Anderson, Herdt and Scobie 1988). Often these descriptions are supplemented by back-of-the-envelope calculations of economic benefits. Such evaluations may provide decision makers with pertinent information but their lack of rigor often undermines their credibility, especially in today's climate of clear accountability.

Partial impact assessment studies quantify the application of research results but do not estimate aggregate benefits. Adoption studies, to trace the use of innovations from the research station or on-farm trials through networks of adopters, are the most popular type of partial impact study. Adoption studies may also evaluate *private* benefits in the form of increased farm production and incomes, assess client satisfaction with research results, and seek to understand why a technology is or is not being used. As such they are especially useful in feeding back information to research decision makers—that is, research managers, stakeholder advisory committees, and involved scientists.

More comprehensive impact studies look beyond adoption information to quantify the effectiveness of research in terms of achieving the major objectives of AROs, especially the enhancement of productivity. This is quantified in impact studies by estimating the economic benefits to both producers and consumers produced by research, computing a rate of return (ROR) to the research investments. There are two broad approaches to estimating RORs—the econometric approach that relates productivity changes to investment in research, and the economic surplus method that builds benefits from the bottom up, based on estimated productivity changes at the field level and adoption rates for each technology. Due to data limitations on the econometric approach, the economic surplus approach has been much more widely applied in developing countries, and is the main focus of this paper.

Comprehensive studies may also trace wider economic benefits of research through factor and product markets. Also, AROs are increasingly being asked to address other objectives beyond efficiency, such as equity and poverty alleviation, environmental quality, food safety, and nutrition. The extent that AROs should depart from their traditional efficiency objective is much debated, but given that other papers in this workshop deal with social and environmental impacts of research, we focus here on efficiency and equity objectives.

Undertaking a comprehensive economic impact assessment requires considerable financial resources and specialized skills in economics. Developing-country NARSs, and increasingly the IARCs, are especially short of economists and other social scientists to carry out impact assessment studies. Research managers, therefore, need to weigh the tradeoffs between costs of more comprehensive analyses, and the value of the additional information that they provide. Thus, this paper emphasizes “quick and clean” methods of impact assessment—methods based on careful analyses that can be implemented within the resources of most AROs and that provide timely feedback to decision makers.

2.3 Level of evaluation: System-wide, program or project?

Impact assessments can be carried-out at different levels of aggregation—individual research projects, specific research programs, or the research and technology system as a whole—depending on the objectives of the exercise. There are a number of arguments against conducting impact assessment of individual projects. First, it is often not possible to assess a project’s impact independently of other projects within the research program. Second, the idea of doing impact assessment is usually to assess whether *total* investment in research has paid off in terms of national goals. In any large portfolio of research projects, there are bound to be a wide range of returns, including many failures. The incidence of some “dry holes” should not deter investment since a few profitable strikes more than cover the costs of the unsuccessful ones. Third, the costs of evaluating each project are usually high in relation to the size of the project investments, and selecting single projects for impact evaluation usually leads to selection bias towards known success stories. However, large, discrete projects in some research systems, such as some mega-projects in the IARCs, may warrant individual studies of impact.

The most appropriate level for impact evaluation is at the program level, which includes costs of all successful and unsuccessful projects and may involve evaluation of one or more products of the research program. These evaluations usually use a benefit-cost analysis framework to measure the rate of return or net present value of research program investments. Program-level evaluation should be a major focus of impact assessment studies and should be conducted on a regular basis (e.g., once every five years).

Evaluation at the system level is less frequently performed (Anderson and Herdt 1990). This requires evaluating the agricultural R&D system at an aggregate level (country or state) or by specific sub-sector (e.g., crops versus livestock). System-level evaluation studies are resource- and data-intensive, requiring time-series data and the use of econometric techniques. Results are most useful for those who fund research and less useful for decision making on resource allocation within the research organization.[1]

2.4 Defining the spatial and temporal dimensions of the analysis

Impact assessment studies must necessarily define the time period and spatial dimensions of evaluation. The time period of the analysis depends on the main objective of impact assessment. It is sometimes interesting to analyze impacts over a long period of 20 or more years into the past. However, if impact analysis is to be used in decision making by research funders or implementers, it should focus on technologies released in the most recent 5-10 year period. This inevitably involves projection of the benefit stream into the future, since adoption and diffusion will likely be incomplete. Thus, useful ex post impact studies will normally include some element of ex ante estimation of adoption (Anderson 1997a, b).

The spatial dimension of the analysis also needs to be carefully determined. This involves answering questions such as what is the geographic mandate of the research program being assessed? National research is usually funded according to political boundaries (such as a state or country), although impacts will be influenced by agro-ecological and socio-economic characteristics. Most analyses assess direct impacts within the politically defined constituency of the funding body, which will often be less than total impacts due to spillovers across political boundaries. In the case of internationally funded research, political boundaries are less important, although spillovers to the industrialized world may not be counted as impacts by donors, who may wish to emphasize development imperatives such as poverty alleviation and food security.

3. Econometric Approaches to Impact Assessment

The econometric approach employs a production function, cost function, or a total factor productivity analysis to estimate the change in productivity due to investment in research Econometric studies of research impacts are conducted within the framework of a production function that incorporates conventional inputs (land, labor etc.), non-conventional inputs (education, infrastructure etc.), and the stock of technical knowledge (perhaps represented by investment in research and extension). Research effort is used as one of the arguments in the function and is usually measured by lagged research expenditures. Recent efforts have expanded the specification to include resource quality variables (e.g., soil erosion, nutrient status etc.), and weather variables.[2] The estimated research coefficients (measuring marginal product) are then used to calculate the value of additional output attributable to the lagged research expenditures (holding other inputs constant) and to derive a marginal ROR to the research investment.

There are many technical areas of debate and refinement in the literature on econometric methods, such as the use of a production function, cost function, or total factor productivity

[1] However, attempts have been made to refine the econometric approach so that it could be used for decision-making purposes (e.g., Bredahl and Peterson 1976, Flores-Moya, Evenson and Hayami 1978). The econometric methods provide a potential means of statistically isolating the effects of various research programs (basic vs. applied; breeding versus physiology/pathology/soil science; domestic vs. foreign). The benefits of research can then be imputed to particular research programs to inform resource allocation decisions.

[2] In a more advanced approach, resource quality can be modeled endogenously as a function of past productivity and management practices (e.g., the analysis of soil quality by Lindert 1996).

(TFP) index, functional form, the length and shape of the time-lag, the appropriate method of determining the rate of return from the estimations, and the quality of indices used as the dependent variable (see Alston, Norton and Pardey 1995 for a comprehensive discussion). However, the main constraint on the wider application of econometric approaches in developing countries is data availability and quality.

The econometric approach requires good-quality time-series data, which are difficult to obtain below the national or state level in most developing countries. Therefore, the approach is generally best for ex post evaluations of entire agricultural research systems over a long period (say, 25-30 years), if the quantity and quality of data allow the use of statistical methods. Much of the work in this area in developing countries has been pioneered by Robert Evenson (e.g., various contributions in Evenson and Pray 1991).

The approach is less relevant for individual research organizations, either national or international, since pertinent time-series data are rarely sufficiently long enough or complete enough or available at the needed level of disaggregation to allow useful econometric estimation. Also, because of the long period of analysis, results are dated and therefore less useful to decision makers.

One compromise is to use panel data to capture both cross-sectional and time-series variability. Secondary data of a panel nature are increasingly available for many of the variables at the district level, especially production and input data, and some recent studies have even included district-wise data on resource quality (e.g., Huang and Rozelle (1995) for China, Lindert (1996) for Indonesia and China, and Ali and Byerlee (2000) for Pakistan). Especially for analysis of impacts of natural resources management (NRM), the appropriate level of analysis is at the cropping or farming systems level defined in terms of a relatively homogeneous agro-ecological resource base.[3] For example, Murgai, Ali and Byerlee (2000) find dramatically different patterns of TFP growth across cropping systems, underscoring the importance of aggregation criteria.

Increasingly time-series panel data are becoming available at the household level (Anderson 1985) that could be employed in econometric approaches to impact evaluation. For example, Cassman and Pingali (1995) use panel farm-level data to estimate TFP for intensive irrigated rice systems in the Philippines. As panel data become more widely available, the use of econometric approaches to research evaluation will expand, although the emphasis will probably continue to be on providing historical aggregate returns to investment in research to those who fund research, rather than to inform decisions of research managers in AROs.

4. The Economic Surplus Approach: Elements of Good Practice

Economic surplus has been the most popular and fruitful approach used to assess the consequences of investments in agricultural research since the pioneering work of Griliches (1958). The approach estimates returns on investment (generally, a weighted average rate of return over time) by (a) calculating the change in consumer and producer surpluses that results from technological change brought about through research that causes the industry supply function to shift outward, and (b) using estimated economic surplus together with research costs to compute the net present value (NPV) or internal rate of return (IRR).

[3] Statistical and GIS techniques are becoming available to meaningfully aggregate district data on the basis of dominant farming systems (ICRISAT 1998).

Economic surplus analysis requires information on values of production affected by research, yield increases, reduction of unit costs and/or the maintenance effects of research, adoption rates, research and adoption lags, cost of R&D, and the appropriate discount rate for converting benefits and costs into present values or for comparing them against an IRR. Price elasticities of demand and supply, and the magnitude and nature of the supply shift determine the relative benefits gained by producers and consumers. In practice there is wide variation in the process, and the methods and data used to measure research costs and benefits. The following sections summarize what we judge to be good practice at each step in the process.

4.1 Measuring program costs

Although measuring program costs may appear straightforward, most past studies have underestimated costs for a number of reasons. First, program costs should include all research, development, and technology transfer costs. Research costs include not only direct program costs but also research overheads, such as administration expenditures, depreciation of capital assets, and costs of complementary services such as libraries. There is also the difficulty in knowing how to treat the input of past knowledge, especially basic and strategic research, given that research lags are a critical determinant of the returns on the investment. Usually, impact analysis excludes such sunk costs associated with previous-generation technology development.[4]

One issue that frequently arises at the national level is whether to include development costs incurred by an IARC that had a role upstream in generating the technology outputs. In the past, IARC products have generally been regarded as a free good by NARSs in impact assessment. However, with NARSs increasingly contributing to the budget of IARCs, a trend that is likely to accelerate, good practice would be to include pro-rated IARC costs according to the country's share in the production of a commodity. For evaluation of IARC research, relevant NARSs' costs should almost always be included (see section 8). Finally, technology transfer costs, including the cost of on-farm verification, public expenditures on extension, and expenditures by the private sector (e.g., farmer organizations), should be included, if accountability to society at large is the focus of analysis.

In defining the costs, it is often difficult to appropriately delineate program-specific costs, unless the accounting department of a research institute is very sophisticated. In practice, approximations have to be employed, such as allocation of scientists' time by program and using an institute-wide average cost per scientist (e.g., Traxler and Byerlee 2000). For certain categories of costs there are other definitional and measurement problems that make the cost estimation step less than straightforward. For example, markets for scientific skills are neither efficient nor free of distortions and shadow pricing may be appropriate.

Good practices in estimating program costs are:

1. Include both **research and extension costs** relevant to the development and dissemination of technologies being assessed.
2. Include costs of both **IARCs** and NARSs if they both played a substantive role in the development and transfer of the technology being assessed.

[4] Statistical procedures have been used to isolate the separate effects of the various programs, and the costs and benefits computed accordingly (e. g., Evenson, Waggoner and Ruttan 1979).

3. Include pro -rated **overhead and administrative costs** of the research center.

4.2 Defining the research products to be evaluated

Impact assessment requires the definition of the products of the research organization. Most commonly the product is new or improved knowledge that may be embodied in a production input, such as seed. New knowledge may be provided directly to the user in the form of information (e.g., a crop management recommendation). Most impact studies have emphasized embodied knowledge, but since a considerable share of the output of research organizations is in the form of information, this type of product is treated below.

4.3 Estimating the K-factor

The central piece of information for any research evaluation study using the economic surplus method is the shift in the supply curve that has resulted from research (K-factor). Most commonly K is conceptualized as a vertical (downward) shift in the supply curve. A more conservative assumption is to use a pivotal shift in the supply curve (Akino and Hayami 1975).

A number of options are available for estimating K, depending on the purpose of the analysis, data available, and the overall methodolog ical approach being applied in the study. Some studies have estimated commodity-supply functions directly, with past expenditures on research included as an explanatory variable in an econometric model. Some studies estimate production functions and deduce the value of K from the estimated production function shifter (e.g., Akino and Hayami 1975). These methods, however, require rather extensive time-series data on inputs and outputs. Hence most ex post economic surplus studies deduce the pattern of past supply shifts (K-factor) attributable to research based on $K_t = \alpha_t k_t$: where (a) α_t is the adoption rate of the technology in year t, and (b) k_t is the per-unit cost reduction in year t resulting from the defined or presumed technological change.

4.3.1 Estimating the adoption of program outputs

Both the temporal and spatial dimensions of the adoption parameter are important in determining the estimates of actual shifts in the supply function (the K-factor) over time. Of these the most difficult is to obtain estimates of temporal parameters, such as when technology adoption begins, when it reaches its peak, and when disadoption begins.

The shape of the lag distribution and the lag length in the earlier years is relatively important in determining economic benefits. The S-shaped logistic curve has been widely used, following Griliches (1958), and generally seems an acceptable approximation to the reality of the adoption phase. In empirical work, simpler linear functions (e.g., Edwards and Freebairn 1981), polynomial lags (e.g., Cline 1975, Davis 1979), or trapezoidal lags (e.g., Antony and Anderson 1991, Huffman and Evenson 1992) have also often been used, and some allow the modeling of disadoption as well—a limitation of the logistic model.

Adoption data are provided by farm surveys, agricultural statistics, seed and other input sales, and estimates from research and extension workers. For some technologies, especially those related to livestock research, the tools of GIS are also used to extrapolate the distribution or occurrence of adoption-inducing factors (such as prevalence of pests) in a region (e.g., Kristjanson et al. 1999). In an ex post setting, the data and information on the year in which a technology was released and a few estimates of adoption at selected years in the past can be used to estimate the adoption curve (e.g., fitting a logistic curve) and the appropriate time lag

(CIMMYT 1993). For discrete technologies, such as crop varieties, retrospective questioning of farmers on the year of adoption usually provides acceptable estimates of adoption data. However, for adoption of research products that are released as information on continuous variables (e.g., rates, timing, and levels of inputs), adoption information is best collected over time.

4.3.2 The per unit reduction in costs (k-value)

The size of the research-induced change in the unit cost of production—the k-value—is a crucial determinant of the total benefits from research. The k-value is usually estimated by the yield increments per unit area or animal resulting from a technology. Translating a yield increase into a cost saving must, however, be adjusted to reflect changes in input use and take account of the industry supply elasticity (Alston, Norton and Pardey 1995).[5] At the firm level, a research-induced per-unit cost reduction may be realized as a result of: (a) the cost reduction arising from changes in productivity when input use is held constant at the pre-adoption optimum, and (b) the cost reduction associated with changes in the input mix to optimize input combinations under the new technology (Alston, Norton and Pardey 1995). The measure of research-induced cost savings therefore needs to reflect research-induced changes in the use of inputs and their opportunity costs. At the level of the individual firm, this means that estimated per unit cost savings ought to reflect the comparison of the commodity enterprise budgets between the new and old technologies, in order to take account of both effects.

The results of on-farm trials, especially farmer-managed trials, and other experimental data are essential for estimating the k value for a specific technology since they control for all other variables. Results from one-off experiments over several locations provide estimates of k for specific agro-ecological zones, while trial data over several years can be used in a statistical model to deduce yield-growth rates attributable to research (e.g., Byerlee and Traxler 1995, Jain and Byerlee 1999). In livestock research, simulation models are often used to derive estimates of productivity impacts of a technology (e.g., Kristjanson et al. 1999).[6]

A critical step is to translate increases in experimental yields into industry-level cost savings. Experimental yields are typically higher than farm yields (Dillon and Anderson 1991) and gains in experimental yields often exceed gains experienced on farms. But the sizes of these differences vary by location and technology, and it is difficult to make empirical generalizations. In practice, a conservative approach is to scale down experimental yield gains to better reflect likely on-farm gains (CIMMYT 1988).

Historical data on farm yields over time and in different locations are often more readily available and more complete than experiment data. Due to the multiplicity of factors affecting yield changes, however, these data are not generally a good source for estimating k-values. With a carefully specified statistical model, however, the historical record can be

[5] The value for the industry supply-response elasticity, ε, is also a critical factor in converting an experimental yield into an industry-level, per-unit cost saving. However, when experimental yields are used to deduce values for k and information on the supply elasticity is lacking, it is often pragmatically expedient to use a supply elasticity of 1.0 (Alston, Norton and Pardey 1995).

[6] Of course, simulation models are also used widely in crop research too, e.g., as advocated by Anderson and Dent (1972) and Harrison, Dent and Thornton (1991), and explored more recently, for example, by White and Grace (2000).

informative and useful when combined with, and juxtaposed against, experimental data. Such data issues are cases of more general scaling issues often faced in agricultural research (Anderson 2000).

4.3.3 Deriving the aggregate supply shift (K-factor)

Even with good estimates of the per unit cost reduction (k-value) and the adoption rate of the technology, there are additional methodological problems in deriving the aggregated value of the supply shift due to a research program.

One difficulty relates to the possibility of *aggregation bias* if different program components yield very different time profiles for cost savings (i.e., k-values). In is usually best to disaggregate the measure of k by individual technologies, to account for different time profiles, so long as they can be regarded as independent. For technologies with a high degree of complementarity or substitution, the only practical alternative is to treat them jointly and estimate the k-value using budgets for the alternative packages of technology.

Another conceptual challenge facing impact assessment at the program level is to determine the *counterfactual situation*—what would have been the shift in the supply curve in the absence of research. Without research, some changes in productivity may occur due to spill-ins from other research programs, or alternatively, productivity may decline due to depreciation in the knowledge stock (Anderson 1992).

Large errors, usually over-estimates, are possible from a lack of rigor in thinking about the "with" and especially, the "without" situation, as noted by Hardaker, Anderson and Dillon (1984). One approach is to model research only as an acceleration of adoption, with benefits deriving only from shortening the time lag (Martinez and Sain 1983, Morris, Dubin and Pokhrel 1994). The major challenge is to estimate the without-research adoption curve. The effects of maintenance research on the estimated incremental research benefits also need to be carefully assessed.[7] Thus, in order to ensure that the estimated incremental net benefit represents only the benefits of the research investment under consideration, one needs to carefully construct cogent "without" scenarios.

A final issue is whether to treat the products of research as *discrete* or *continuous* outputs. Generally for economic surplus analysis, technologies are regarded as discrete products and the impact assessment of a program considers all technologies released in a given time period. However, in some situations, and especially for varietal technologies in post-Green Revolution areas, research outputs are almost continually released, adopted and replaced. In these cases, k is better conceived as an annual average rate of change in unit cost, based on, say, the rate of genetic gains in yields achieved through release of new varieties, rather than a one time shift (Byerlee and Traxler 1995).

In sum, good practice for calculating the K-factor includes;

[7] In a recent study by Thirtle and Townsend (2000) the authors show the effect of not accounting for the maintenance-research impact in livestock research evaluation studies. By separating the maintenance effects of animal health research from output increases due to animal improvement research, they estimate a ROR to research investment which are higher than previous estimates. The implication is that all ROR estimates must be severely biased downwards when they are based on an implicit assumption that, with no research, there would be no change in output or productivity.

1. Combine technical, scientific, and economic information from a number of sources. If available, use the results of previous experimental trials to assess changes in cost and yield.
2. Reflect research-induced changes in the use of inputs and their opportunity costs in the measure of cost savings.
3. Disaggregate supply-shift estimates by components of a research program.
4. Carefully construct the "with" and "without" scenarios to ensure that the estimated incremental net benefit truly represents the benefits of the research investment under consideration.
5. Most importantly, make explicit the assumptions applied, so that others can transparently assess the analysis.

4.4 Tradable status, output prices, and partitioning of benefits

Total economic benefits and the partitioning of benefits among producers and consumers requires knowledge of the supply and demand curves for the product. The key issues relate to determining (a) the tradable status of a commodity, a gap in much of the research evaluation literature, and (b) prices appropriate for valuing outputs.

Many analysts have uncritically assumed the extremes of a completely open or closed economy. Improved infrastructure and liberalized markets and trade in recent years mean that many commodities that were non-tradables are now tradable, implying that more, and perhaps most, domestic research benefits are now being captured by producers. In situations of poor infrastructure or long distances between producing and consuming points, however, there is a large "wedge" between import- and export-parity prices so that potentially tradable commodities, even basic grains, are converted into non-tradables within a wide price range (Byerlee and Morris 1993).

Figure 1 summarizes a common market situation for food staples in developing countries. Urban consumers in a port city with demand curve D_u face a supply curve, S^oS^u dominated by imports at price, P_I, while rural interior producers and consumers effectively operate in a closed economy in the price range, P_I +t to P_I, where t is the transport cost from rural areas to the port. S^oS^r and D^u are the supply and demand schedules for the rural interior market, respectively. In this situation, it is possible and even likely that, with a research-induced shift in the supply curve from S^o to S^1, rural consumers register a gain in consumer surplus as they move from R^o to R^1 while urban consumers at U^o experience no such change. Thus where most of the food-purchasing population is located in urban areas, consuming tradable food staples, it is likely that consumer price effects of research will be limited. This characterizes the situation of many countries in Latin America, North Africa, and increasingly much of Asia.[8] On the other hand, where most of the population is located in rural areas, and infrastructure and markets are not well developed, markets for food staples in rural areas are likely to operate in the non-tradable zone in a wide range between P_i and P_i + t in Figure 1. This is the case for much of sub-Saharan Africa. The main point is that the tradable status of a commodity needs to carefully assessed for each situation. Even within a country, a food

---

[8] However, the large size of many Asian countries and the fact that the world market for rice, the main staple, is thin, means that production can influence local consumer prices through effects on world market prices.

staple will often be non-tradable in one area (e.g., rural areas with poor infrastructure) and tradable in another area (e.g., urban areas, perhaps located on the coast).

Another important market condition to consider in estimating economic benefits is possible presence of policy distortions in the form of taxes or subsidies on output, exports or imports, production quotas, and production target prices, that affect output prices. Depending on the size of the output market and the trade status of the country, the presence of market distortions can significantly complicate the analytical model for estimating research benefits. Alston, Norton and Pardey (1995) is an excellent source for convenient formulas that can be used for computing changes in economic surplus under different market conditions and policy interventions. The extent of any distortion, if present, needs to be measured to derive an "economic price" to be used in the estimation of economic benefits. Byerlee and Morris (1993) provide conditions under which border prices or estimated equilibrium prices are appropriate economic prices, in the presence of policy distortions.

Total benefits are relatively insensitive to assumptions about the elasticity of demand and supply. Indeed, if the purpose of the economic analysis is to measure only the total economic surplus and not its distribution between consumers and producers, then a simplified benefit-cost approach gives reasonable estimates of economic benefits. The simplified approach is valid in cases where the country is clearly a "price-taker", and where the intervention being evaluated is not expected to change the country's status from being a net importer to a net exporter of that commodity. The advantage of this approach is that it is simple and presents a more transparent view to non-economists of the data and assumptions used in the analysis.

Key elements of good practice are:

1. Carefully study the market situation of the commodities affected by research, and determine the tradable status appropriate for estimating total surplus.
2. Use economic prices that are appropriately adjusted to reflect policy distortions in the output market. All monetary variables should, of course, be expressed in real terms.
3. Estimates of research benefits should be disaggregated by commodities, production environment or geographical basis if the parameter estimates are different for different components of a research program.

4.5 Calculating the return to the investment

The final step for the analyst is to bring together all the information to calculate the NPV or IRR of research investment. A number of spreadsheet templates have been developed specifically for computation of economic surplus and other derived measures of value—MODEXC originally developed by Lynam and Jones (1984), RE4 developed at ACIAR, and DREAM developed at ISNAR/IFPRI (see Appendix).

There are several options for evaluating the stream of net benefits—NPV, IRR and Benefit-Cost (BC) ratio. The NPV method is preferred where concern over the scale of investment is not an issue. More commonly, the IRR is computed and compared to the opportunity cost of funds, although it does not reveal either the scale of the investment or the value of the programs. The best approach is to use a combination of IRR and NPV calculations to summarize the relevant information on the total returns to research.

This basic economic surplus approach can be extended in many ways. Substitution effects on the production and consumption side can be readily incorporated if a full set of supply and demand elasticities, including cross elasticities is available (Alston, Norton, and Pardey 1995). To incorporate agro-ecological differences and regional impacts, geographical information systems are increasingly being applied to spatially characterize and map research or problem "domains" within inter-regional models (e.g., Pardey and Wood 1994, Hassan 1998, Bigman and Fofack 1999).

Good practice in data analysis and calculating summary measures includes:

1. Creating a spreadsheet template for each of the commodity/products/markets affected by a research program using product- (or zone-) specific parameter values for each disaggregated research component (and respective spreadsheet template).
2. Documenting the logical structure of the model/method used for calculating the benefit and cost streams and clearly laying out the assumptions and sources used to derive model parameters.
3. Calculating both summary measures (NPV and IRR) to indicate the profitability of research investment.
4. Perhaps carrying out appropriate sensitivity analyses by varying some of the critical assumptions made in the model and thus to check the robustness of the results.

5. Beyond Commodity Research: Broadening the Application of Impact Assessment

Most economic impact studies have addressed commodity-specific research and, within that category, there has been an overwhelming emphasis on varietal technologies.[9] For other types of research, especially research on natural resource management (NRM), product-quality, crop and livestock management research, and social sciences, specific conceptual and methodological difficulties arise in applying the economic surplus framework. Additionally, some research by AROs is aimed at modifying institutions, and some generate externalities (positive and negative) which complicate the measurement of economic surplus. This gap is now being filled, especially with impact assessment studies being conducted by CGIAR centers such as CIFOR, ICLARM, ICRAF, IFPRI and ILRI.

5.1 Impacts of management research: Improved information

Many of the products of research are not embodied in tangible inputs such as seed, but are provided as information, sometimes in the form of a recommendation. Adoption of information research products is often difficult to track, complicates attribution to research, and raises thorny problems of valuation.

A good example is crop and resource management research, where little economic impact work has been carried out, although this type of research accounts for about one-half of all crop research in developing countries. Information from crop management experiments is usually summarized in the form of production recommendations, which define rate, timing and methods for using inputs, as well as the conditions under which these recommendations apply. The value of improved input management information depends on the interaction of

[9] This is changing as more impact studies (ex post and ex ante) of livestock and fisheries research are conducted both by NAROs and IARCs (e.g., Thornton and Odero 1998, Dey 1999, Mokoena, Townsend and Kirsten 1999, Mukhebi et al. 1999).

input response with location-specific climatic and field conditions (e.g., integrated pest management or phosphorus maintenance doses conditional on soil-test information) (e.g., Perrin 1976, Blackmer and Morris 1992).

It is especially difficult to measure benefits of improved information from research compared to what would have occurred in the absence of the program. The effects of research must be separated from other sources of information, including farmers' learning-by-doing and private-sector suppliers. Often this is approximated by the yield and cost differences observed between adopters and non-adopters of a management practice—that is, the area between the "with" and "without" adoption curves (e.g., Joshi and Bantilan 1998).

Traxler and Byerlee (1992) provide a five-step framework for assessing benefits in such situations:

1. Identify research areas for which an improved management practice has been supplied as a new recommendation (i.e., new information) issued to farmers.

2. Determine which practices farmers have modified in a manner consistent with the new recommendations.

3. Determine whether a revised recommendation has **caused** the change in farmers' practice.

4. Measure the impact of each research-induced change in crop management practice on economic surplus.

5. Sum the economic surplus across practices and compare the benefits stream to the costs of research and extension.

A refinement of the above approach is to track changes in farmers' subjective beliefs about payoffs to a practice in response to improved information provided by research (Feder and Slade 1984). This is empirically more challenging but has been successfully applied by Pingali and Carlson (1985) in a commercial agricultural setting.

5.2 Social science research

Assessing the impact of stand-alone social science research has a number of parallels with crop management research—i.e., social science research that is intended to be used directly by farmers and policy makers rather than as an input into other research programs. This includes social-science management research that leads to more rapid adoption of technical innovations, marketing research that improves price incentives to farmers, and policy research that leads to an increase in agricultural productivity due to the removal of policy distortions.

Determination of the relative contribution of the social-science research information to changed policy or management practice is clearly central to impact evaluation. In recent years, IFPRI has initiated and commissioned several studies to develop methods for estimating the impacts of social science research in general and food-policy research in particular (e.g., Norton and Alwang 1998, Kilpatrick 1998, Gardner 1999, Garrett 1999). Methods proposed or used for evaluating policy research range from a cross-country econometric approach (e.g., Park 1998), calculating the national income impacts of time-savings resulting from hastened policy decisions (e.g., Ryan 1999a), valuing information and calculating the savings in dead-weight loss of policy distortions,[10] to simple case studies documenting and qualitatively assessing the results, outcomes and impacts of research (e.g., Islam and Garrett 1997) [11]. However, attempts to evaluate research projects in this area indicate that more experience is still required to define good practice in this area (Davis and Lubulwa 1995).

5.3 Impacts on natural resources and the environment

Attention to effects of research on the natural resource base is relevant for both productivity-enhancing research, such as commodity research, to the extent that some of the gains are made at the expense of reductions in the quality of the resource base, as well as to research directed explicitly at enhancing natural resources and environmental quality, or at least reducing their degradation. NRM research outputs are not only often highly location specific but monitoring the multiple bio-physical consequences of NRM research outputs in the field is expensive, and the interplay between the economic and bio-physical domain is strong. Moreover, there are still many conceptual and methodological issues that need to be resolved in impact assessment of NRM research (Izac 1998, Pachico 1998). Assessing the full range of impacts requires that appropriate information sources be tapped at various levels, such as field, watershed, and regions. It also raises technical difficulties of measuring and valuing resource degradation, (e.g., soil loss under alternative management practices), or environmental pollution (e.g., through inappropriate use of agricultural chemicals).[12]

5.3.1 Measuring and valuing changes in resource quality

There appears to be a consensus that estimates of the *quantities* of resource depletion or environmental damage are easier to get (although often costly) than estimates of the *correct prices* to attach to the quantities (Crosson and Anderson 1993). Even in valuing *private* shadow prices for onsite effects of resource quality, there are valuation issues such as whether to use foregone output, replacement costs, user costs, or option and existence value to measure the costs of resource degradation (Harrington, Jones and Winograd 1994). In the simplest cases, such as soil nutrient "mining", cost can be approximated by replacement cost (i.e., the market price of chemical or organic fertilizers needed to fully replenish the nutrient). Likewise attempts have been made to value soil erosion effects by estimating erosion losses and output foregone using simulation models of soil run-off and crop productivity (Steiner et al. 1995). Problems are compounded with externalities where costs of agricultural chemical pollution may include regulatory costs, health-related costs, and unpriced environmental costs (Steiner et. al. 1995). A number of techniques now exist for placing economic values on environmental changes (Winpenny 1991), including health effects of pesticide poisoning (e.g., Rola and Pingali 1993), effects on soil degradation (Brandon 1995, Malik and Faeth 1993) and contingent valuation of water pollution. A final complication with valuing changes in natural resources and environmental quality is the effect of policy distortions, such as

---

[10] This method is listed as one of the evaluation model to be used by ACIAR in their research program evaluations (see, for example, Davis and Lubulwa 1995).

[11] Smith (1998) reports implementation of this approach in at least one research institution in Australia.

[12] When it is relevant to look to the human health as well as the environmental and productivity dimensions of any research-related effects, the complexity of the assessment task is magnified accordingly (Ruttan 1994).

subsidized inputs that foster overuse of chemicals, or institutional policies such as land tenure that affect adoption of improved management practices.

Conceptually, such assessments can and should be included in a broader economic impact analysis of agricultural research. The main problem is that they are data-intensive and require highly developed skills in collecting and statistically manipulating large quantities of data. For these reasons, it is unlikely that valuation of natural resources impacts will become an integral part of research impact studies in developing countries in the near future, although as we argue later, IARCs should be providing leadership in developing cost-effective and robust methods for wider inclusion of these impacts.

5.3.2 The debate on including resource quality in TFP

There is a parallel literature on the inclusion of resource qualityvariables and environmental externalities in econometric approaches using TFP. Joint outputs or externalities of agricultural production such as soil erosion, depletion of soil fertility, and groundwater aquifer mining are usually not taken into account in TFP measurement yet are clearly central to evaluating research impacts on natural resources (Alston, Anderson and Pardey 1995). To address these problems, several researchers have argued that changes in the quality of the natural resource base (including externalities) should be accounted for in a total social factor productivity indices (TFP) (Antle and McGuckin 1993, Harrington, Jones and Winograd 1994, Alston, Norton and Pardey 1995, Herdt and Steiner 1995, Repetto et al. 1997). A few studies have attempted to empirically measure Total Social Factor Productivity (TSFP), usually using data from experimental plots (e.g., Ehui and Spencer 1993, Whitaker 1993, Barnett, Payne and Steiner 1995).

The inclusion of the costs of environmental regulation and abatement, as well as benefits of improved environmental quality, in TSFP is the subject of considerable debate (Repetto et al. 1997, Gollop and Swinand 1998, Byerlee and Murgai 1999). There are both conceptual and practical problems in using TSFP. On the conceptual side, TFP indices are based on an underlying production function, assuming profit maximizing behavior. Therefore, inclusion of resource degradation is valid only if degradation is on-site and under the possibly dubious assumption that farmers are aware of, and take into account its long-run effects in their input and output choices. The most relevant case for agriculture is the accounting for soil nutrient mining and the on-site cost of erosion, since these are probably internalized by farmers in decision making. If it is a pure externality (off-site cost) that does not affect the system being evaluated (e.g., water quality for consumers), it should not enter the calculations, although private regulatory costs may be included as an input into production.

Deleted: ¶

TSFP also requires that inputs and outputs be valued at long-term *economic* prices. Again the profit maximizing assumptions underlying TFP require that observed input and output choices have to be valued at the prices faced by producers. Therefore, valuing production at the societal shadow prices of goods (e.g., Gollop and Swinand 1998) rather than producer prices is conceptually flawed. In addition, as discussed above, valuing resource quality and particularly externalities is especially challenging.

In sum, the recent emphasis in some of the economics literature on searching for an all-embracing single measure of sustainability in the form of TSFP has not been helpful. Rather, attention should now turn to doing a better job of measuring productivity and trends in resource quality, and relating the two through production function analysis. Inclusion of resource-quality variables in the production function is also an advantage in that it allows valuation of resource degradation through the estimated marginal product.

5.4 Non-research products of research organizations

Another ARO product that is difficult to fit into an "economic surplus" model of impact assessment is the economic evaluation of the performance of an ARO in non-research activities such as training, networking, and advisory services aimed at improving institutional performance in agriculture and over the long term, enhancing the impacts of these institutions on agricultural productivity. Many national and international agricultural research institutes devote sizeable resources to such activities. However, there has been little methodological and practical work in the area of economic impact assessment of these non-research products.

The recent surge in interest in evaluating the "institutional impacts" of CGIAR centers, has led to significant advances in the methods and application of "institutional impact assessment". This includes the impact an agricultural research organization has on capacity building, human-resource development, and performance of other institutions (Goldsmith 1993). A companion paper in this series deals with institutional impacts, and we simply note recent work of the CGIAR in this area, especially by ISNAR (e.g., Horton and Borges-Andrade 1999, Horton and Mackay 1998, Mackay and Debela 1998, Mackay, Horton and Debela 1998) and IFPRI (e.g., Paarlberg 1999, Ryan 1999b). The framework for these studies consist of a variety of methods such as surveys, case studies, and document analysis to understand four dimensions of institutional assessment—the external environment, institutional motivation, institutional capacity, and institutional performance (Mackay et al. 1998). These experiences and the use of case study methodology show that it is possible to discern links between institutional capacity building and management and policy actions. However, none of these efforts has been successful in putting economic values on the outcomes of capacity building, and we do not feel that this is a priority for AROs at this time.

6. Measuring Wider Economic Impacts: Spillovers, General Equilibrium, and Equity Effects

6.1 Incorporating spillovers and spillins

Research often generates four classes of spillovers: (a) interlocational (direct spillovers), (b) interfoci (indirect spillovers), (c) intercommodity, and (d) intersectoral (Evenson 1989). The most relevant for *ex post* impact assessment are the interlocational spillovers, which affect the production technology and prices in non-mandated regions of a research program. Spillover effects are usually measured in terms of the impacts the local technology has on the K-factor of other regions (e.g., Flores-Moya, Evenson and Hayami 1978, Edwards and Freebairn 1984, Davis, Oram and Ryan 1987). Interlocational spillover effects may be international (e.g., Flores-Moya, Evenson and Hayami 1978) or arise within a country between regions when a research program involves multiple institutions or multiple sites within institutions (e.g., Brennan 1999, Jain and Byerlee 1999).

Spillovers from one research program are spillins to another. Interlocational spillins are not conventionally part of ex post impact studies as they are conceptually and methodologically difficult to measure. However, recently there have been a few attempts at measuring research spillins of international crop improvement research (Brennan 1989, Brennan and Fox 1995, Pardey et al. 1996, Maredia and Byerlee 1999).

In order to minimize the potential for double counting, and in acknowledgement of the site-specific nature of many research results, Wood and Pardey (1993) have suggested assuming no within-country spillovers between those locales where a national research program is simultaneously developing new technologies.

But contrary to the perception that biological technology is location specific, spillovers, both direct and indirect, are pervasive within a country and internationally (Byerlee and Traxler 1995, Evenson and Gollin 1994, Brennan 1999). The analyst therefore needs to carefully identify the characteristics and sources of technology to estimate returns to a specific research program. The study by Traxler and Byerlee (2000) show for the case of wheat research in India that simply analyzing productivity gains in the mandate area of the program without tracing the origin of the technology can lead to highly biased estimates of returns to individual programs, with returns overestimated for spillin areas, and underestimated for spillover-producing programs.

With spillins, it is more meaningful to estimate the marginal value added of additional investments in research, rather than the average IRR which might be high even for inefficient programs. Maredia and Byerlee (2000) show that incorporating research spillins in an ex ante assessment of research investments of a crop improvement research program implies that many wheat research programs, especially those for small mandate areas, are inefficient in terms of overinvesting in local technology development rather than testing and releasing technologies from other programs. Thus when evidence of research spillovers and spillins exists, good practice for ex post impact assessment is to conduct a spatially disaggregated analysis that explicitly treats spillovers and spillins (e.g., Jain and Byerlee 1999). The information requirements for such highly disaggregated studies are great, and simplifying assumptions are inevitable. Integrating spillins of research in a conventional ex post impact study adds to the complexity of the model, especially if marginal rather than average RORs are to be estimated, and it may not be practical for all program assessments to include them. Many studies ignore research spillovers also on the ground that, for the purposes of accountability to funders, spillovers add little value to the analysis. However, if ex post impact assessment is to be used as an input in making strategic investment decisions and research priority setting, estimation of research spillovers and especially spillins is valuable and should be incorporated into impact analysis.

6.2 General equilibrium effects of research.

Agricultural research generates several indirect economic effects, including labor-market effects and growth linkages of technical change, that go beyond the standard measures of consumer and producer surpluses discussed above. The effect of technology adoption on demand for labor has potentially major indirect benefits. These effects will also often operate inter-regionally to the extent that inter-regional labor markets exist (David and Otsuka 1994). Total effects depend on labor supply and demand characteristics as well as elasticities of supply and demand for targeted agricultural products but are often limited by the low observed elasticity of total agricultural production (Binswanger and Quizon 1986).

Overall effects of technical change as a result of research ultimately depend on general growth and income effects induced throughout the economy by consumption, input and output linkages. A full accounting of these effects requires a general equilibrium (GE) model. The few studies that have tried to estimate these effects indicate that, in many cases, indirect growth-linkage effects rival or even exceed the effects of conventionally measured economic surplus (Byerlee 1973, Janssen et al. 1990, Renkow 1991, de Janvry et al. 1999). These studies have estimated multiplier effects of direct benefits of research of the order of 1.3-2.0 (Hazell and Haggblade 1993, Delgado, Hopkins and Kelly 1998). Growth linkages are higher where agricultural income is a high proportion of household income, where initial asset distribution is relatively equitable, and where there is underutilized economic capacity (Hazell and Haggblade 1993).

It is not conceptually difficult to formulate such models within a GE framework but their practical application will necessarily be limited by the information, time and analytical capacity required. The incorporation of the various effects may require regional disaggregation of product and factor markets and inter-regional trade flows. For large countries that account for a significant share of international trade in the commodity of interest, such analysis should extend to international markets (Edwards and Freebairn 1981). However, economists have not yet developed GE models that are both practical and detailed enough to provide much guidance for evaluating research impacts.[13] The modeling of the entire economy (or beyond) is an exercise that is likely to be beyond the scope of most research-evaluation and priority-setting studies, and the capacity of most AROs to implement. In addition, there is as yet no evidence that including indirect effects of research would lead to a difference in ranking of research program impacts, or in returns to research relative to other public investments.

6.3 Impacts on income distribution and poverty alleviation

The economic surplus approach can be extended to assess impacts on the distribution of income: (a) between producers and consumers (e.g., Akino and Hayami 1975, Edwards and Freebairn 1981), (b) among different groups of consumers and producers (Scobie and Posada 1978, Hayami and Herdt 1977), and (c) between producers in different regions (Flores-Moya, Evenson and Hayami 1978, Alston, Edwards and Freebairn 1988).

With the emphasis on poverty alleviation as a central objective of many donors and governments, tracing the impacts of research on the distribution of producer and consumer surplus between different income groups is a logical extension of the economic surplus approach. On the consumer side, this is relatively straightforward since consumer expenditure information stratified by income or expenditure group is widely available and benefits can be partitioned between income groups according to their share in consumption. The overall effects, however, depend on substitutions among commodities, requiring a full specification of demand functions (Pinstrup-Andersen, Londoño and Hoover 1976).

Producer surplus can also be partitioned to income groups according to their share of production or proportion of production marketed. Lack of data on sources of producer income by commodity for different income strata is, however, a major constraint to this analysis, although the increasing availability of household survey data for analyzing poverty

[13] In practice and in theory, the distinction between partial equilibrium and GE is not always clear. At one extreme are the detailed economy-wide GE models that allow everything to change. At the other extreme is the typical commodity market model of supply and demand that limits changes to direct effects on producers and indirect effects on consumers through prices.

should help in this regard (Byerlee 2000). More commonly, some proxy, such as farm size, is employed (e.g., Mutangadura and Norton 1999).

Market-price effects of technological change in one region also have important price spillovers for producers in other regions, often with negative equity effects (Coxhead and Warr 1991). The classic study of Scobie and Posada (1978) indicates how increases in productivity of relatively well-endowed rice farmers with irrigation produced negative benefits for upland rice farmers who could not adopt the technology; but many other scenarios are possible (Byerlee 2000). With spatial mapping of poverty variables, an interregional equity dimension can also be incorporated (Bigman and Fofack 1999).

These studies indicate that the poverty incidence of research can be readily incorporated in extensions of the economic surplus models. More refined measures of poverty impacts can be employed by considering effects not only on the number of poor but also on the severity of poverty (Alwang and Siegel 1999). Application of these approaches is, however, constrained by the availability of data on poverty incidence, especially on the producer side. Indirect effects on poverty through labor markets and general growth linkages are also not usually considered, although there are examples of the application of GE models to analyze broader income-distributional impacts of technical change (de Janssen et al. 1990, Janvry et al. 1999).

In sum, good practice is beginning to emerge in this area. Given the strong poverty alleviation focus on the CGIAR, IARCs should strive to be leaders in mainstreaming this type of analysis.

7. Institutionalizing Economic Impact Assessment

7.1 Who should do impact assessment?

Impact assessment can be carried out by analysts within the organization or by independent professional evaluators (including those working under contract to the organization). Impact assessments designed to influence internal decision making are clearly better done internally. If the objective is to provide information to those who fund the research, the tradeoff must be assessed between the credibility that may accompany "independent" external assessments and the cost-effectiveness of more informed and targeted internal assessments (Anderson 1997a,b).

Most AROs, including IARCs, should have "in-house" social science capacity to conduct impact assessment. A few IARCs have formalized this capacity into specialized impact assessment units that report directly to the Director General in order to provide independence. External evaluators may be contracted by the ARO itself, or by the research-funding agency. Many impact assessment studies have also been done "independently" by university researchers. Results of these academically based studies may provide useful information but their lack of formal connection to decision makers in AROs or funding bodies limits their influence.

Whether evaluation is done internally or externally, the "ownership" of impact studies should remain with the ARO whose research programs are being assessed. The involvement of decision makers and researchers of the research institutes is the key to the successful use of the results of impact studies. Not only social scientists but technical scientists should be full participants in impact assessment in order to bring the best information to the analysis and also facilitate timely feedback of results to research program priorities.

A research organization should at least have a champion who provides leadership in the area of impact assessment and helps integrate the results of such studies into research planning. The impact of "impact assessment" will depend on mechanisms developed within a research system by which results are communicated and considered as input into the decision-making process. This will depend on several factors (IAEG 1999):

- making researchers focus on the practical application of their work as well as its technical merit,
- increasing the interaction between scientists and evaluators at the initiation and development stages of program proposals,
- requiring that all new proposals beyond a threshold investment be accompanied by an ex ante impact assessment, and
- making monitoring and evaluation (M&E) of ongoing research a routine practice.

Byerlee and Alex (1998) provide some guidelines for institutionalizing impact assessment in national research institutes, through the establishment of M&E capacities (Table 1). **Good practice is that once every five years, each applied research program should undergo a comprehensive external review of technical quality, and an analysis of its economic impacts.**

The major problem in M&E systems for research has been the tendency to collect too much information in a highly centralized and bureaucratic manner with little thought about how such information can be incorporated into decision making. To date, there are few examples of good practice in this area but the key elements of successful institutional capacity in impact assessment appear to be:

- Designing a simple data system that collects data on critical variables on a regular basis (see below).
- Establishing a very small unit to *promote* monitoring and evaluation, provide training, develop standards, and arrange evaluation of research programs.
- Ensuring that there are mechanisms for research managers to receive timely information from impact evaluations, and
- Contracting out some M&E activities, especially impact studies, to independent agencies or individuals.

7.2 Institutionalizing data systems for economic impact analysis

Impact assessment is a data-intensive activity that demands time-series industry-level data (size and productivity measures such as production, yield, area, commodity prices, and elasticities), measures of research outcomes (productivity change, technology adoption, change in cost of production, institutional development) and the research program inputs and outputs that lead to these changes. Collecting appropriate data is perhaps the most time-consuming and costly component of conducting impact assessment.

Research programs must balance data needs between the ideal (and costly) and the practical, and must draw data from various sources (Alex 1998). If impact assessment is to become an integral part of the research process, it is important that an appropriate data system be institutionalized within the research system. Institutionalizing a data system also ensures that the information and data generated by research are available in a systematic and timely manner and are retained for future use as the staff and institution change and evolve.

Data generated internally by a research organization include experimental data (both off-farm and on-farm), data from adoption surveys, data from informal surveys and case studies. Management information systems such as the International Crop Information System are increasingly being employed to store and retrieve experimental data in the IARCs and some NARSs. However, comparable data systems for adoption and farm survey data have yet to be developed, and have normally been implemented in an ad hoc manner in each ARO.

A key issue for many AROs is the role of baseline and panel surveys to provide data on adoption and related household variables. Many types of impacts can only be adequately assessed if relevant baseline data exist (especially for disembodied research products), and regular re-surveys, preferably of the same households, are undertaken over time to monitor changes in farmer practices. However, the resources required for comprehensive panel data sets are usually beyond the capacity of most AROs. National agricultural research organizations, for example, cannot conduct national baseline surveys both because of the large sample size required for national coverage, and the fact that data requirements for tracking impacts of all research programs (various crops, livestock, crop protection, natural resources management etc) are prohibitive.

Compromise must therefore be the order of the day. We think that each ARO should have a panel of households that it regularly monitors for changes in key farm practices and productivity indicators related to the most important types of research outputs. The sample should be representative of an important subset of the mandate area of the ARO, but few AROs will have the capacity to monitor more than 100-150 households. Monitoring can be undertaken at intervals of three to five years to track changes in practices, and more detailed data may be collected from a sub-sample of households. Increasingly with emphasis on natural resources management, such monitoring should include field observation of soil and water quality, pest populations, and biodiversity, which adds considerably to time and resources required. While such data sets will be far from comprehensive for the purposes of impact assessment, they do provide a base that allows "informed extrapolation" to a wider domain.

IARCs have been in the forefront in setting up farm-household and farm plot panels (such as the ICRISAT village surveys, the IRRI rice farmer surveys, and CIMMYT's wheat field panel in northwest Mexico, and wheat-rice field panel in the Terai of Nepal). However, even some of these panel data sets have not been sustained over time (e.g., the ICRISAT data sets have not been updated for several years). In addition, to our knowledge, none of these data sets has yet been geo-referenced using global positioning technology to allow future researchers, and future generations of researchers, to accurately return to the same fields and farms in order to up-date the information.

Such data collection systems necessarily require considerable up-front investment to define the key parameters, sample size, frequency of data collection, and benchmark sites. However, once established, the collection of data at regular intervals has been shown to be a relatively low-cost exercise. Collection of such data should become an integral part of farming systems and natural resource management research programs at major research stations in national systems (Byerlee 1991).[14]

In the meantime, district-level secondary data offer the best opportunity to analyze long-term trends in resource quality. For example, in Pakistan, soil and water test data have been collected continuously for over 30 years in response to demands from farmers and extensionists but only recently have been compiled and made available (Ali and Byerlee 2000). Similar data for Indonesia and parts of China have also provided new insights into long-term trends in soil quality (Lindert 1996). Measures of genetic diversity in farmers' fields have recently been developed, which may be proxies for pest and disease losses (Smale et al. 1998). Wider use of simulation models also has potential to infer trends in resource quality.

Good practice in impact assessment suggests increased efforts by IARCs to work with national AROs to establish and maintain relevant farm-household panel data sets for impact assessment.

7.3 Presenting the results of impact assessment

Research impact studies generate much information and are sometimes very technical. Research managers and policy makers also tend to be skeptical of the data and methods used; they often also find the reports difficult to understand, interpret, and apply. Therefore presenting this information in an appropriate format and through appropriate channels is important to enhance communication with scientists, research managers and funders. Preparing policy briefs summarizing the results of impact studies in non-technical terms is good practice for disseminating the results. Relegating the more detailed explanatory material to the technical appendices of impact reports will provide the plausibility-enhancing material in a place where the specialist can readily find it.

Timeliness is also a key to utilization of impact information. Often program managers and staff have mixed feelings about impact evaluation—they recognize the need for more and better evaluation, but their personal experience is that evaluations have seldom produced information when needed for decision making (Horton and Mackay 1999).

7.4 The "economics of impact studies"

A key question for research managers is how much to invest in impact assessment. Cost of impact assessment must include:

- **Direct costs** of evaluation staff, travel, field work, building the data system, data analysis, and overheads.
- **Opportunity cost** of the time of those involved in the evaluation process which is especially high for social scientists, who are typically in short supply in most AROs,

[14] The exact data to be collected will depend on hypotheses on the main factors likely to affect system productivity and sustainability and will necessarily be system specific. In all situations, data would be collected on carefully selected soil physical, chemical and biological parameters, with parameters such as soil organic matter likely to be included in most situations. In some cases, pest populations might be monitored through qualitative or quantitative scoring.

and who must carry out field-based diagnostic studies, farmer participatory work, and ex ante priority setting.

These costs must be weighed against benefits that may be derived from the process itself, as well as the results of the impact assessment. Program scientists often benefit from the rigor of carefully defining the products and beneficiaries of the research program, and estimating model parameters, such as the K-factor. AROs benefit if the impact assessment leads to increased efficiency in resource allocation, or level of funding.

Thus, like any research activity, impact assessment should be subject to the same pressures for justification as other uses of research funds. Some expenditure on research evaluation is surely justified in any system, but there are no clear guidelines on how much to spend. GTZ aims to spend on evaluation (including all self- and independent evaluation activities) up to 5% of a research program's budget on average. INTA in Argentina spends an estimated $2 million on M&E, or 2.5% of its entire budget (Hogg 1994 cited in Horton and Borges-Andrade 1999). One of the summary recommendations of a research evaluation workshop organized in Australia was that 3-4% of the research and development budget could be spent profitably on impact evaluation (Chudleigh and Bond 1993). The trend towards collecting data and measuring past productivity growth in response to increased demand for impact analysis means that the cost of ex post evaluations may fall. But this may be offset by the trend toward evaluating the social and environmental impacts of research, which are likely to be more difficult and costly.

To institutionalize impact assessment, each research institute needs to develop guidelines on budget share for impact assessment, and these costs should be included up front in program budgets. With further developments in assessing impacts of social science research, it may eventually be possible to compute the economic return to impact assessment!

8. Special Considerations for IARCs

IARCs not only have to wrestle with the multitude of methodological and institutionalization challenges discussed above, but have the additional challenge of evaluating products that are of an intermediate nature, and which are further modified by national AROs into final products that may be used across a wide and dispersed geographical area. This raises a number of special considerations on the role for IARCs in economic impact assessment.

8.1 Attributing costs and benefits between IARCs and NARSs

Most IARC research products are the joint effort of the IARC and national partners. Without extremely careful and thus case-study analysis, it is not easy to determine the relative contributions of collaborators in such joint research activities.[15] Thus the assessment of impact must usually consider the *gross effects* of collaboration, representing the combined contributions of the various partners in the collaboration. Similarly, costs should also be computed on a gross basis, combining the IARC and national ARO contributions. Even for relatively finished products from IARCs such as varieties, national partners have been

[15] One can perhaps make a case in, say, crop improvement work, that to the extent that improved access is provided to international collections of germplasm, the *international* dimension of the effort is important—but just about the same case can be made for nearly every aspect of subsequent innovative activity, whether it be selection, testing, or release.

observed to invest at least twice the IARC cost of varietal development in varietal testing and release (Byerlee and Maredia 1999).

Recently there has been some progress in methods of attribution among collaborators, especially for valuing IARC spillover (or spillin) benefits for crop improvement research. For example, Pardey et al. (1996) have shown the consequences of different schemes for partitioning the benefits of varietal improvement research based on various rules for handling identified genetic contributions to varietal pedigrees.

8.2 Roles for IARCs in mainstreaming impact work

Impact evaluation of IARCs, especially for global efforts such as germplasm improvement, can only be effective if it builds on institutionalized efforts in national research systems. The objective should be for IARCs to even tually play a facilitating and synthetic role in routine impact assessment. As procedures become of more uniform quality with respect to data-collection methods and analytic structures (such as have been outlined in our review), we would also expect that IARC analysts would make increasing use of the greater number and more dispersed results reported by NARSs. The recent efforts at IFPRI to undertake a meta-analysis of such findings is an important step in this direction (Alston et al. 2000a,b). Needless to say, the findings of this study highlight the practical difficulties inherent in any such endeavor, because of inconsistencies of method and coverage. In addition, it is recognized that the feasibility of these IARC roles is challenged by pervasive resource constraints in national AROs (NAROs), and persistent difficulties in adequately staffing their social-science units.

Yet another dimension in which we would anticipate leading-edge IARC engagement is in the wider spatial charting of the impact of agricultural research investments. NARSs clearly have little reason to be involved in impact documentation beyond their national borders, yet some of the more investor-impressive impacts stretch over many borders, as a results of intended and unintended spillovers from both national and international research investments. To the extent that the world market for food products is significantly affected through widespread adoption of IARC research products such as higher-yielding cultivars, comprehensive impacts th rough lower world prices should be measured, particularly for their direct benefits to low-income consumers.

In the meantime, mechanisms and roles must be developed to ensure that “the whole” of global impact evaluation efforts is greater than the sum of the parts, in this case the accumulating efforts of impact evaluation in NARSs and IARCs. One of the tasks is for some organization to take on responsibility for keeping track of individual efforts. This should be increasingly easy in a well-connected e-world. Given the charter of FAO, and especially the NARS Secretariat, it is a prime candidate for taking this work on. A Web-site that is regularly updated would seem to be the way to head, irrespective of who steps into the breech. The database assembled by IFPRI would be a natural starting point for launching the global impact portal.

8.3 A role for IARCs in frontier studies of impact assessment

IARCs should play a lead role in demonstrating the cutting edge of good practice in impact assessment, which is not only to be expected of a relatively well-funded agricultural research system (such as that CGIAR even now still is), but must be demanded of it, if it is to deliver on its own mandate to carry out upstream tasks. Apart from mere demonstration of good practice, we also argue that a natural role for IARCs is to conduct frontier investigations of impact that contribute to the understanding of the links between research and economic impacts, as well as provide new methods for undertaking impact analysis. Given the importance of collaboration and learning by doing, such in-depth studies are best conducted through active partnerships with NARSs. This would provide a cost-effective mechanism for capacity-building in NARSs that should well complement the still fledgling efforts of IAEG/SPIA/TAC in fostering such work. Given the methodological challenge of analyzing more fully the consequences of adoption of research innovations on the reduction (or otherwise) of poverty, at least some of the in-depth work should be addressed to this objective within the framework of the internationally-supported agricultural research portfolio. Analogous challenges (so far probably better addressed in recent work in the IARCs) apply to other contemporary research priorities, such as improved management of natural resources.

9. Conclusions

Deleted: ¶
¶

This paper set out to explore the "state of the art" in ex post economic impact assessment of agricultural research, in order to provide a "tour of good practice" that would bring to light the different contexts in which research impact assessment takes place, indicate the most appropriate methods and techniques, and highlight issues related to institutionalization of impact assessment. This should allow AROs to assess their own standing in relation to "good practices" across a range of types of research activities.

The general case for mainstreaming impact evaluation in AROs has been made above, highlighting the complementary objectives of accountability and providing formalized feedback to research decision makers. Agricultural research covers many different areas, and AROs are engaged in many different kinds of activities other than research. Because of this diversity, impact evaluation in AROs is undertaken for many reasons, and to support different types and levels of decision making. Consequently impact assessments differ greatly in their scope, purpose, and method. This diversity makes it difficult to compress the review of the "state of the art" into a strict "code of practice" to guide research managers and policy makers. However, several themes emerge from the review, which can improve impact assessment processes in AROs.

9.1 State of the art and future methodological development

From this review, we conclude that, for relatively standard productivity-enhancing innovations such as improved cultivars, and improved livestock production methods, fairly standard and accepted methods are available, and are being increasingly applied by both IARCs and national AROs. However, we also conclude that there is much room to improve the quality of these applications. Simplistic assumptions about lags, costs, and supply shifts, together with failure to account for spillins, have biased estimated RORs, usually upward. In addition, the emphasis on evaluating individual technologies in an ad hoc manner, rather than research programs on a regular basis has undoubtedly favored the selection of winners. Practical guides to implementation of impact assessment methods and institutionalization of these methods within AROs on a routine basis, are needed to provide a more balanced picture of impacts of agricultural research.

Impact studies have also been narrowly based on selected technologies, especially crop improvement. But the increasing emphasis on natural resource management work in many research systems, and the relatively undeveloped procedures for adequately getting a good handle on the economic impacts of this type of research, means that much more effort is required to develop, test, implement and refine methods. Similarly, economic impacts must be broadened beyond efficiency objectives to give much more attention to distributional consequences of research. Clearly, the IARCs have a lead role to broaden the study of impact assessment in these emerging areas. IARC staff and resources are, however, limited, and the challenge is large, suggesting that deliberate efforts to engage relevant skills from academia and elsewhere will probably be required to ensure reasonable chances of satisfactory progress on the methods front.

9.2 Keys to successful impact evaluation

Define minimum data sets: The full set of data needed for an "ideal" economic impact assessment study includes time-series industry-level data, measures of research outcomes (both spatial and temporal dimensions), and research inputs. In practice, impact analysts have too little time and too few resources for systematic and thorough data collection and analysis. The main challenge for the analyst, therefore, is to find a practical means of doing impact assessment with available time and data. We recognize that trying to match assessment methods within available data often leads to impact assessment exercises that fall short of "professional standards". Nonetheless, we believe that efforts to institutionalize a minimum data set that is regularly collected over time should eventually allow for more comprehensive impact assessment. These data sets should be focused on the K factor—adoption data and data that document changes in per unit costs, especially experimental data. In an imperfect world, regular collation and reporting of adoption data alone can provide many of the benefits of more comprehensive but costly impact assessment.

Combine quantitative and qualitative assessment: Formal methods of economic evaluation are based on quantitative assessment of research impacts, often summarized in a single figure, the IRR. The process of conducting impact studies however, provides much qualitative information that can provide valuable feedback to research managers and scientists. These include qualitative outcomes both at the organizational level (e.g., human resource development, institutional linkages) and at the clientele level (e.g., food security impacts, inter-sectoral spillovers, environmental impacts) that allow a more complete picture of the benefits from investing in research and development. The reporting and utilization of impact information could be greatly enriched by emphasizing these qualitative dimensions.

Decentralize as far as possible, but have a central oversight capability: Traditionally, the administrations of AROs have been highly centralized. This is not conducive for inculcating an "impact culture" in an organization. Decentralization provides an environment that promotes mechanisms to ensure accountability, quality control, and uniform reporting in decentralized units. In a survey of evaluation practices, Horton and Borges-Andrade (1999) found that evaluations tended to be most common in organizations that had decentralized decision making. Doing in-house impact assessment in decentralized units is expedient and contributes to the better understanding of the research impact process. However, the decentralized system, whereby each research unit does its own impact assessment, raises the question of bias and credibility of the impact study results. Thus in order to deal with this problem, a central oversight capability at the institute or a system level should be developed to assess program impact analyses.

Build institutional capacity for ongoing evaluation: A key to successful impact evaluation is to make impact assessment an ongoing activity within the organization and not a one-time ad hoc exercise. To sustain and implement impact assessment work on a regular basis requires institutional changes and capacity building within an ARO, such as: (a) making researchers focus on the value side of their work as well as its technical merit. (b) making M&E of ongoing research a routine practice, requiring research programs to submit progress reports, achievements of milestones and final reports, (c) requiring new proposals above some cut-off budget value to be accompanied by an ex ante impact assessment, (d) developing institutional guidelines that require the establishment of databases for research evaluation studies and results, which are regularly updated (e) regularly conducting workshops to develop staff skills and training for research evaluation, and most importantly, (f) dedicating a share of an organization's budget to impact assessment activity.

In most research organizations, it will be important that at least one identifiable staff member focus on impact assessment and helps to build the institutional capacity in impact evaluation. In large AROs, a small impact assessment unit is appropriate to work towards developing a "results-oriented mentality" in the organization. This does not imply that all impact assessment has to be done by that person or the unit. Other social scientists within the ARO, together with judicious use of outside expertise, may carry out the bulk of the work in a decentralized system. The degree of detail, such as analytical sophistication or commodity coverage, can be increased over time as the procedures become more fully integrated into the decision-making processes of an ARO.

Develop mechanisms to integrate information with decision making: Many organizations conduct impact assessment studies but few have developed formal, institutionalized systems that support decision making and institutional learning (Horton and Borges-Andrade 1999). One of the major reasons for institutionalizing impact assessment capacity is that it becomes part of the decision-making process within an ARO and fosters an economic approach to allocating research resources. This requires that results of impact assessment be made available to research managers and policy-makers in a timely and accessible manner. It is also important that the economic information included in the impact assessment is presented to research funders and managers in a form that assists in improved resource allocation.

9.3 Building capacity for the future

There is a need for research managers and researchers themselves to be trained to recognize the value and the limitations of the results of impact assessment. Top-level managers of an ARO should be sensitized to the value and uses of impact evaluation as a management tool. They should be made aware of the "key factors" discussed above and the changes that are usually needed to make impact assessment successful. ISNAR has made considerable efforts in this direction by conducting workshops and developing training modules and materials tailored to the needs of research managers of AROs in developing countries. Other regional and international organizations should follow in ISNAR's footsteps and support the training activities for ARO researchers and managers.

The reality in many AROs is that there are similar training implications also for the social-science staff who are presumed to be or charged with executing such impact work. This stems from the difficulties many AROs experience in retaining good staff in this area, and the limited experience that even well-trained staff often have with such applied analysis. The training needs are thus likely to be on-going, and the need for up-to-date training materials a continuing one. Perhaps this training issue should figure centrally in the planning of priorities for SPIA, as it contemplates the institutional capacity issue of NARSs in its domain.

References

Akino, M. and Y. Hayami. 1975. "Efficiency and equity in public research: Rice breeding in Japan's economic development", *American Journal of Agricultural Economics* 57(1), 1-10.

Alex, G. 1998. *Assessing Agricultural Research. Towards Consensus on a Framework for Performance and Impact Assessment.* Agricultural Research and Extension Group, Special Report No.6, World Bank, Washington, DC.

Ali, M. and D. Byerlee. 2000. *A Decomposition Analysis of Productivity Growth and Resource Degradation in Pakistan's Punjab*. Policy Research Working Paper, World Bank, Washington, DC, in press.

Alston, J.M., J.R. Anderson, and P.G. Pardey. 1995. "Perceived productivity, foregone future farm fruitfulness, and rural research resource rationalization", In G.H. Peters and D.D. Hedley (eds.), *Agricultural Competitiveness, Market Forces and Policy Choice*. Dartmouth, Aldershot, 639-651.

Alston, J.M., G.W. Edwards, and J.W. Freebairn. 1988. "Market distortions and benefits from research", *American Journal of Agricultural Economics* 70(2), 281-288.

Alston, J.M., C-C Kang, M.C. Marra, P.G. Pardey, and T.J Wyatt. 2000a. *A Meta-Analysis of Rates of Return to Agricultural R&D: Ex Pede Herculem?* IFPRI Research Report No. 113, IFPRI, Washington DC.

Alston, J.M., M.C. Marra, P.G. Pardey, and T.J Wyatt. 2000b. "Research returns redux: A meta-analysis of returns to agricultural R&D", *Australian Journal of Agricultural and Resource Economics* 44(2), 185-215.

Alston, J.M., G.W. Norton, and P.G. Pardey. 1995. *Science Under Scarcity. Principles and Practice for Agricultural Research Evaluation and Priority Setting*. Cornell University Press and ISNAR, Ithaca, republished in 1998 by CAB International, Wallingford.

Alwang, J. and P.B. Siegel. 2000. "Measuring the impacts of agricultural research on poverty reduction: Improving the dialogue between policy makers and research managers", *Food Policy* (in press).

Anderson, J.R. 1985. "Assessing the impact of farming systems research: Framework and problems", *Agricultural Administration* 20(4), 225-235.

Anderson, J.R. 1992. "Measuring the efficacy of international agricultural research", In *CIMMYT 1991 Annual Report - Improving the Productivity of Maize and Wheat in Developing Countries: An Assessment of Impact*, CIMMYT, Mexico, DF, 8-15.

Anderson, J.R. 1997a. "On measuring the impact of natural resources research", In *Report of the Natural Resources Advisers' Conference*. Department For International Development (DFID), Sparsholt, Winchester, 9 July 1997, Vol. 2, (electronic), 222-232.

Anderson, J.R. 1997b. On grappling with the impact of agricultural research, Background paper prepared for the CGIAR discussion of Center impacts, ICW, Washington, DC, October 31, 1997, CGIAR Secretariat, Washington, DC.

Anderson, J.R. 2000. “Response and risk in rural ecosystems: From models and plots to defined universes”, *Agriculture, Ecosystems & Environment* in press.

Anderson, J.R. and J.B. Dent. 1972. “Systems simulation and agricultural research”, *Journal of the Australian Institute of Agricultural Science* 36(4), 264-9.

Anderson, J.R. and R.W. Herdt.1990. “Reflections on impact assessment”, In R.G. Echeverría (ed.), *Methods for Diagnosing Research System Constraints and Assessing the Impact of Agricultural Research, Vol. 1, Diagnosing Agricultural System Constraints*. ISNAR, The Hague, 33-40.

Anderson, J.R., R.W. Herdt, and G.M. Scobie. 1988. *Science and Food: The CGIAR and Its Partners*. World Bank, Washington, DC.

Antle, J.M. and T. McGuckin. 1993. “Technological innovation, agricultural productivity, and environmental quality”, In G.A. Carlson, D. Zilberman, and J.A. Miranowski (eds.), *Agricultural and Environmental Resource Economics*. Oxford University Press, New York.

Antony, G. and J.R. Anderson. 1991. “Modeling technology replacement over time for the ex-ante analysis of agricultural research projects”, *Agricultural Systems* 37(2), 183-192.

Barnett, V., R. Payne, and R. Steiner (eds.). 1995. *Agricultural Sustainability: Economic, Environmental, and Statistical Considerations*. Wiley, Chichester.

Bigman, D. and H. Fofack. 1999. Geographic targeting for poverty alleviation: methodology and applications. Paper presented at the CIAT international workshop on Assessing the Impact of Research on Poverty Alleviation, San Jose, Costa Rica, Sept 14-16.

Binswanger, H.P. and J.B. Quizon. 1986. *What can Agriculture do for the Poorest Rural Groups?* Report No. 57, Research Unit, Agricultural and Rural Development Department, World Bank, Washington, DC.

Blackmer A. and T. Morris. 1992. “Selecting nitrogen fertilization rates for corn: New options”, In *Building bridges: Cooperative Research and Education for Iowa Agriculture*. Proceedings of the Third Annual Conference, Leopold Center for Sustainable Agriculture, Iowa State University, Ames.

Brandon C. 1995. Valuing environmental costs in Pakistan: The economy-wide impact of environmental degradation. Background paper for “Pakistan 2010,” Asia Environmental Division, World Bank, Washington, DC.

Bredahl, M.E. and W.L. Peterson. 1976. “The productivity and allocation of research at U.S. agricultural experiment stations”, *American Journal of Agricultural Economics* 58(4), 684-692.

Brennan, J.P. 1989. “Spillover effects of international agricultural research: CIMMYT-based semi-dwarf wheats in Australia”, *Agricultural Economics* 3(4), 323-332.

Brennan, J.P. 1999. “Efficiency in wheat improvement research: A case study of Australia” In M.K. Maredia and D. Byerlee (eds.), *The Global Wheat Improvement System: Prospects for Enhancing Efficiency in the Presence of Spillovers*, CIMMYT Research Report No. 5, CIMMYT, Mexico, DF.

Brennan, J.P. and P.N. Fox. 1995. *Impact of CIMMYT Wheats in Australia: Evidence of International Spillovers.* Economics Research Report No. 1/95, NSW Agriculture, Wagga Wagga, Australia.

Byerlee, D. 1973. *Indirect Employment and Income Distribution Effects of Agricultural Development Strategies: A Simulation Approach Applied to Nigeria*. African Rural Employment Paper No. 9, Michigan State University, East Lansing.

Byerlee D. 1991. "Monitoring farmers' practices and fields as a continuing activity in on-farm research for evolving cropping systems", *Journal of Asian Farming Systems Research* 1(1), 179-190.

Byerlee, D. 2000. "Targeting poverty alleviation in research priority setting: Concepts and practice for national research organizations", *Food Policy* in press.

Byerlee, D. and G. Alex. 1998. *Strengthening National Agricultural Research Systems – Policy Issues and Good Practice*. World Bank, Washington, DC.

Byerlee, D. and M.K. Maredia. 1999. "Measures of technical efficiencyof national and international wheat research systems", In M.K. Maredia and D. Byerlee (eds.), *The Global Wheat Improvement System: Prospects for Enhancing Efficiency in the Presence of Spillovers.* CIMMYT Research Report No. 5, CIMMYT, Mexico, DF.

Byerlee, D. and M.L. Morris. 1993. "Calculating levels of protection: Is it always appropriate to use world reference prices based on current trading status?", *World Development* 21(5), 805-815.

Byerlee, D. and R. Murgai. 2000. "Sense and sustainability revisited", *Agricultural Economics*, in press.

Byerlee, D. and G. Traxler. 1995. "National and international wheat improvement research in the post-Green Revolution period: Evolution and impacts", *American Journal of Agricultural Economics* 77(2), 268-278.

Cassman, K.G. and P.L. Pingali. 1995. "Extrapolating trends from long-term experiments to farmers' fields: The case of irrigated rice systems in Asia", In V. Barnett., R. Payne and R. Steiner (eds.), *Agricultural Sustainability: Economic, Environmental, and Statistical Considerations.* Wiley, Chichester, 63-84.

Chudleigh, P.D. and K.A. Bond. 1993. "Summary of workshop findings and implications for further application of benefit-cost analyses for the GRDC", In P.D. Chudleigh and K.A. Bond (eds.) *volume 4—Benefit Cost Methods in Research Evaluation: Proceedings of a Workshop held in Canberra, 25-26 May 1993*. Gains for Grain Occasional Paper Series No. 4, GRDC, Canberra.

CIMMYT. 1988. *From Agronomic Data to Farmer Recommendations: An Economics Training Manual*. Completely revised edition, CIMMYT, Mexico, DF.

CIMMYT. 1993. *The Adoption of Agricultural Technology: A Guide for Survey Design.* CIMMYT, Mexico, DF.

Cline, P.L. 1975. Sources of productivity change in United States Agriculture. Unpublished Ph.D. Thesis, Oklahoma State University, Stillwater.

Coxhead, I.A. and P.G. Warr. 1991. "Technical change, land quality, and income distribution: A general equilibrium analysis", *American Journal of Agricultural Economics* 73(2), 345-360.

Crosson, P. and J.R. Anderson. 1993. *Concerns for Sustainability: Integration of Natural resource and Environmental Issues in the Research Agendas of NARS.* ISNAR Research Report No. 4, International Service for National Agricultural Research, The Hague.

David, C. and K. Otsuka (eds.). 1994. *Modern Rice Technology and Income Distribution in Asia*. Lynn Rienner, Boulder.

Davis, J.S. 1979. Stability of the research production coefficient for U.S. Agriculture. Unpublished Ph.D. Thesis, University of Minnesota, St. Paul.

Davis, J.S. and G. Lubulwa. 1995. *An Overview of ACIAR's Economic Evaluation Activities with an Animal Science Program Focus.* Economic Evaluation Unit Working Paper Series No. 15, ACIAR, Canberra.

Davis, J.S., P.A. Oram, and J.G. Ryan. 1987. *Assessment of Agricultural Research Priorities: An International Perspective.* ACIAR and IFPRI, Canberra.

de Janvry, A., G. Graff, E. Sadoulet, and D. Zilberman. 1999. "Agricultural biotechnology and poverty: Can the potential be made a reality?" Paper presented at the conference, "The shape of the coming agricultural biotechnology transformation: Strategic investment and policy approaches from an economic perspective," University of Rome "Tor Vergatta", June 17-19.

Delgado, C., J. Hopkins, and V.A. Kelly. 1998. *Agricultural Growth Linkages in Sub-Saharan Africa.* IFPRI Research Report 107, IFPRI, Washington, DC.

Dey, M.M. 1999. "The impact of genetically improved farmed Nile Tilapia in Asia", *Aquaculture Economics and Management* 3(3), in press.

Dillon, J.L. and J.R. Anderson. 1990. *The Analysis of Response in Crop and Livestock Production*, 3rd edn. Pergamon, Oxford. Dillon, J.L. and Anderson, J.R. (1990), *The Analysis of Response in Crop and Livestock Production*, 3rd edn, Pergamon, Oxford

Edwards, G.W. and J.W. Freebairn. 1981. *Measuring a Country's Gains from Research: Theory and Application to Rural Research in Australia.* A Report to the Commonwealth Council for Rural Research and Extension, AGPS, Canberra.

Ehui, S.K. and D.S.C. Spencer. 1993. "Measuring the sustainability and economic viability of tropical farming systems: A model from Sub-Saharan Africa", *Agricultural Economics* 9(4), 279-296.

Evenson, R.E. 1989. "Spillover benefits of agricultural research: Evidence from U.S. experience", *American Journal of Agricultural Economics* 71(2), 447-452.

Evenson, R.E. and D. Gollin. 1994. *Genetic Resources, International Organizations, and Rice Varietal Improvement.* Economic Growth Center Discussion Paper No. 713. Yale University, New Haven.

Evenson, R.E. and C. Pray (eds). 1991. *Research and Productivity in Asian Agriculture.* Cornell University Press, Ithaca.

Evenson, R.E., P.E. Waggoner, and V.W. Ruttan. 1979. "Economic benefits from research: An example from agriculture", *Science* 205, 1101-1107.

Feder G, and R. Slade.1984. "The acquisition of information and the adoption of new technology", *American Journal of Agricultural Economics* 66(3), 312-320.

Flores-Moya, P., R. E. Evenson, and Y. Hayami.1978. "Social returns to rice research in the Philippines: Domestic benefits and foreign spillovers", *Economic Development a nd Cultural Change* 26(3), 591-607.

Gardner, B.L.1999. *Policy-Related Social Science Research in Agriculture.* IFPRI Impact Assessment Discussion Paper No. 9, IFPRI, Washington, DC

Garrett, J. 1999. *Research that Matters: The impact of IFPRI's Policy Research*. IFPRI, Washington, DC.

Goldsmith, A. 1993. "Institutional development in national agricultural research: Issues for impact assessment", *Public Administration and Development* 13(3), 195-204.

Gollop, F.M. and G.P. Swinand. 1998. "From total factor productivity to total resource productivity: An application to agriculture", *American Journal of Agricultural Economics* 80(3), 577-583.

Griliches, Z. 1958. "Research costs and social returns: Hybrid corn and related innovations", *Journal of Political Economy* 66(5), 419-431.

Hardaker, J.B., J.R. Anderson, and J.L. Dillon. 1984. "Perspectives on assessing the impacts of improved agricultural technologies in developing countries", *Australian Journal of Agricultural Economics* 28(2/3), 87-108.

Harrington, L., P. Jones, and M. Winograd. 1994. Operationalizing sustainability: A total productivity approach. Paper presented at the Land Quality Indicators Conference, CIAT, Cali, June 1994.

Harrison, S.R., J.B. Dent, and P.K. Thornton. 1991. "Assessment of benefits and costs of risk-related research in crop production", In R. C. Muchow and J. A. Bellamy (eds.), *Climatic Risk in Crop Production: Models and Management for the Semiarid Tropics and Subtropics*, CAB International, Wallingford, 491-510.

Hassan, R. (ed.). 1998. *Maize Technology Development and Transfer: A GIS Application for Research Planning in Kenya*. CAB International (in association with CIMMYT and the Kenya Agricultural Research Institute), Wallingford.

Hayami Y. and R.W. Herdt. 1977. "Market price effects of technological change on income distribution in semi-subsistence Agriculture", *American Journal of Agricultural Economics* 59(2), 245-256.

Hazell, P. and S. Haggblade. 1993. "Farm-nonfarm growth linkages and the welfare of the poor". In M. Lipton and J. Van der Gaag (eds.), *Including the Poor.* World Bank, Washington, DC, 190-204.

Herdt, R.W. and R.A. Steiner. 1995. "Agricultural sustainability: Concepts and conundrums", In V. Barnett, R. Payne and R. Steiner (eds.), *Agricultural Sustainability: Economic, Environmental, and Statistical Considerations.* Wiley, Chichester, 3-13.

Hogg, D.R. 1994. “Planificacion, seguimiento y evaluacion en El INTA, Argentina”, In A.R. Novoa and D. Horton (eds.), *Administracion de la investigacion agropecuaria: Experiencia en las americas.* Tercer Mundo Editores, Santa Fe de Bogota, 243-263.

Horton, D., P. Ballantyne, W. Peterson, B. Uribe, D. Gapasin and K. Sheridan. 1993. *Monitoring and Evaluating Agricultural Research: A Sourcebook.* CABI and ISNAR, Wallingford.

Horton, D. and J. Borges-Andrade. 1999. “Evaluation of agricultural research in Latin America”, *Knowledge, Technology and Policy* 11(4), 42-68.

Horton, D. and R. Mackay. 1998. *Assessing the Institutional Impact of Development Cooperation: A Case from Agricultural R&D.* ISNAR Discussion Paper No. 98-2. ISNAR, The Hague.

Horton, D. and R. Mackay. 1999. “Evaluation in developing countries: An introduction”, *Knowledge, Technology and Policy* 11(4), 5-12.

Huang, J. and S. Rozelle. 1995. “Environmental stress and grain yields in China”, *American Journal of Agricultural Economics* 77(4), 853-864.

Huffman, W.E. and R.E. Evenson. 1992. “Contributions of public and private science and technology to US agricultural productivity”, *American Journal of Agricultural Economics* 74(3), 752-756.

IAEG. 1999. “Impact assessment of agricultural research: Context and the state of the art”, Paper presented at the ASARECA/ECART/CTA Workshop on Impact Assessment of Agricultural Research in Eastern and Central Africa, Entebbe, Uganda, 16-19 November 1999.

ICRISAT. 1998. Database Development, Typology Construction and Economic Policy Analysis for Sustainable Rainfed Agriculture. Sustainable Rainfed Agriculture Research and Development Project (Module I). Submitted to the Indian Council of Agricultural Research, July 1998, pp. 142.

Islam, Y. and L. Garrett. 1997. *IFPRI and the Abolition of the Wheat Flour Ration Shops in Pakistan: A Case Study on Policy Making and the Use and Impact of Research.* Impact Assessment Discussion Paper No. 1, IFPRI, Washington, DC.

Izac, A-M.N. 1998. *Assessing the Impact of Research in Natural Resources Management. Synthesis of an International Workshop 27-29 April 1998*, Nairobi, Kenya, ICRAF Nairobi.

Jain, K.B.L. and D. Byerlee. 1999. “Investment efficiency at the national level: Wheat improvement research in India”, In M.K. Maredia and D. Byerlee (eds.) *The Global Wheat Improvement System: Prospects for Enhancing Efficiency in the Presence of Spillovers,* CIMMYT Research Report No. 5, CIMMYT, Mexico, DF, 85-102.

Janssen W., L. Sanint, L. Rivas, and G. Henry. 1990. “CIAT’s commodity portfolio reexamined: Indicators of present and future importance”, In W. Janssen (ed.), *Trends in CIAT Commodities.* International Center for Tropical Agriculture, Cali, 1-39.

Joshi, P.K. and M.C.S. Bantilan. 1998. *Impact Assessment of Crop and Resource Management Technology: A case of Groundnut Production Technology.* ICRISAT Impact Series No. 2, ICRISAT, Patancheru.

Kilpatrick, Jr. H.E. 1998. *Some Useful Methods for Measuring the Benefits.* IFPRI Impact Assessment Discussion Paper No.5, IFPRI, Washington, DC.

Kristjanson, P., J. Rowlands, B. Swallow, R. Krusha, P. de Leeuw, and S. Nagda. 1999. *Using the Economic Surplus Model to Measure Potential Returns to International Livestock Research: The Case of Trypanosomosis Vaccine Research.* ILRI Impact Assessment Series 4, ILRI, Nairobi.

Lindert, P. 1996. Soil degradation and agricultural change in two developing countries. Paper presented at the conference, Global Agricultural Science Policy to the 21st Century, August 26-28, 1996, Melbourne, Australia.

Lynam, J.K. and P.G. Jones. 1984. Benefits of technical change as measured by supply shifts: An integration of theory and practice. Mimeo., CIAT, Cali.

Mackay, R. and S. Debela. 1998. *ISNAR's Achievements, Impacts and Constraints, 1991-1996*. ISNAR Discussion Paper No. 98-3, ISNAR, The Hague.

Mackay, R., D. Horton, and S. Debela. 1998. *Accounting for Organizational Results: An Evaluation of the International Service for National Agricultural Research.* ISNAR Discussion Paper No. 98-4. ISNAR, The Hague.

Mackay, R., S. Debela, T. Smutylo, J, Borges -Andrade, and C. Lusthaus.1998. *ISNAR's Achievements, Impacts and Constraints: An Assessment of Performance and Institutional Impact*. ISNAR, The Hague.

Malik, R.P.S. and P. Faeth. 1993. "Rice-wheat production in Northwest India", In P. Faeth (ed.), *Agricultural Policy and Sustainability: Case Studies from India, Chile, and the United States.* World Resources Institute, Washington, DC.

Maredia, M.K. and D. Byerlee (eds.). 1999. *The Global Wheat Improvement System: Prospects for Enhancing Efficiency in the Presence of Spillovers.* CIMMYT Research Report No. 5, CIMMYT, Mexico, DF.

Maredia, M.K. and D. Byerlee. 2000. "Efficiency of research investments in the presence of international spillovers: Wheat research in developing countries", *Agricultural Economics* 22(1), 1-16.

Martinez, J. and G. Sain. 1983. *The Economic Returns to Institutional Innovations in National Agricultural Research: On-farm Research in IDIAP in Panama.* CIMMYT Economics Program Working Paper No. 2/83, CIMMYT, Mexico, DF.

Mokoena, M.R., R.F. Townsend, and J.F. Kirsten. 1999. "Cattle improvement schemes in South Africa: Measuring the returns to research investments", *Agrekon* 38(1), 78-89.

Morris, M.L., H.J. Dubin, and T. Pokhrel. 1994. "Returns to wheat research in Nepal", *Agricultural Economics* 10(3), 269-282.

Mukhebi, A.W., T. Chamboko, C.J. O'Callaghan, T.F. Peter, R.L. Kruska, G.F. Medley, S.M. Mahan, and B.D. Perry. 1999. "An assessment of the economic impact of heartwater (Cowdria ruminantium infection) and its control in Zimbabwe", *Preventive Veterinary Medicine* 39: 173-189.

Murgai, R. M. Ali, and D. Byerlee. 2000. "Productivity Growth and Sustainability in Post-Green Revolution Agriculture in the Indian and Pakistan Punjabs", *Word Bank Research Observer* (in process).

Mutangadura, G. and G.W. Norton. 1999. "Agricultural research priority setting under multiple objectives: an example from Zimbabwe", *Agricultural Economics* 20(3), 277-86

Norton, G.W. and J. Alwang. 1998. *Policy for Plenty: Measuring the Benefits of Policy-Oriented Social Science Research.* IFPRI Impact Assessment Discussion Paper No.6, IFPRI, Washington, DC.

Paarlberg, R.L. 1999. *External Impact Assessment of IFPRI's 2020 Vision for Food, Agriculture, and the Environment Initiative.* IFPRI Impact Assessment Discussion Paper 10, IFPRI, Washington, DC.

Pachico, D. 1998. Conceptual framework for natural resource management research and basic methodological issues in impact assessment. Paper presented at the international workshop, Assessing the impact of research in natural resource management, 27-29 April 1998, Nairobi, Kenya.

Pardey, P.G. and S. Wood. 1994. "Targeting research by agricultural environments", In J.R. Anderson (ed.) *Agricultural Technology: Current Policy Issues for the International Community.* CAB International, Wallingford, 566-589.

Pardey, P.G., J.M. Alston, J.E. Christian, and S. Fan. 1996. *Hidden Harvest: U.S. Benefits from International Research Aid*. Food Policy Report, IFPRI, Washington, DC.

Park, D. 1998. *A Proposal for Measuring the Benefits of Policy-oriented Social Science Research.* IFPRI Impact Assessment Discussion Paper No. 3, IFPRI, Washington, DC.

Perrin R.K.1976. "The value of information and the value of theoretical models in crop response research", *American Journal of Agricultural Economics* 58(1), 54-61.

Pingali P.L. and G.A. Carlson. 1985. "Human capital, adjustments in subjective probability, and the demand for pest controls", *American Journal of Agricultural Economics* 67(4), 854-861.

Pinstrup-Andersen P., N. R. de Londoño, and E. Hoover. 1976. "The impact of increasing food supply on human nutrition: Implications for commodity priorities in agricultural research", *American Journal of Agricultural Economics* 58(2), 131-142.

Renkow, M. 1991. *Modeling the Aggregate Effects of Technological Change on Income Distribution in Pakistan's Favored and Marginal Production Environments*. CIMMYT Economics Paper No. 4, CIMMYT, Mexico, DF.

Repetto, R., D. Rothman, P. Faeth, and D. Austin. 1997. "Productivity measures miss the value of environmental protection", *Choices* Fourth Quarter, 16-19.

Rola, A.C. and P. Pingali. 1993. *Pesticide, Rice Productivity, and Farmers' Health: An Economic Assessment.*: World Resources Institute and IRRI, Washington, DC.

Ruttan, V.W.1994. "Challenges to agricultural research in the 21st century", In V.W. Ruttan (ed.), *Agriculture, Environment, and Health: Sustainable Development in the 21st Century.* University of Minnesota Press, Minneapolis, 243-257.

Ryan, J.G. 1999a. *Impact of Rice Policy Changes in Viet Nam.* IFPRI Impact Assessment Discussion Paper No. 8, IFPRI, Washington, DC.

Ryan, J.G. 1999b. *Assessing the impact of policy research and capacity building by IFPRI with Malawi.* IFPRI Impact Assessment Discussion Paper No. 11, IFPRI, Washington, DC.

Scobie, G. M. and R. Posada. 1978. "The impact of technological change on income distribution: The case of rice in Colombia," *American Journal of Agricultural Economics* 60 (1), 85-92.

Smale, M., J. Hartell, P.W. Heisey, and B. Senauer. 1998. "The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan", *American Journal of Agricultural Economics* 80(3), 482-493.

Smith, V.H. 1998. *Measuring the Benefits of Social Science Research.* IFPRI Impact Assessment Discussion Paper No. 2. IFPRI, Washington, DC.

Steiner, R.A., L. McLaughlin, P. Faeth, and R.R. Janke. 1995. "Incorporating externality costs into productivity measures: A case study using US agriculture", In Barnett, V., R. Payne, and R. Steiner (eds). *Agricultural Sustainability: Economic, Environmental, and Statistical Considerations.* Chichester, Wiley, 209-230.

Thirtle, C.G. and R. Townsend. 2000. Is Livestock Research Unproductive? Separating Health Maintenance from Improvements, Contributed Paper presented at the Conference of the International Association of Agricultural Economists (IAAE), Berlin, August 2000.

Thornton, P.K. and A.N. Odero. 1998. *Compendium of ILRI Research Impact and Adoption, 1975-98*. ILRI Impact Assessment Series 1, ILRI, Nairobi.

Traxler, G. and D. Byerlee. 1992. "Economic returns to crop management research in a post-green revolution setting", *American Journal of Agricultural Economics* 74(3), 573-582.

Traxler, G. and D. Byerlee. 2000. "Linking technical change to research effort: An examination of aggregation and spillovers effects", *Agricultural Economics* (forthcoming).

Whitaker, M. 1993. "Using long-term experiment data in evaluating productivity and sustainability of alternative crop technologies", In R.C. Wood and J. Dumanski (eds), *Sustainable Land Management for the 21st Century: Proceedings of the International Workshop on Sustainable Land Management for the 21st Century*. University of Lethbridge, Lethbridge, Canada, June 20-26.

White, J.W. and P.R. Grace (eds.). 2000. *Modeling Extremes of Wheat and Maize Crop Performance in the Tropics*. CIMMYT Proceedings, NRG-GIS Series 00-01, CIMMYT, Mexico, DF.

Winpenny, J. 1991. *Values for the Environment: A Guide to Economic Appraisal.* HMSO/ODL, London.

Deleted: ¶

APPENDIX

Guide to Further Sources for Economic Impact Assessment of Agricultural Research

Manual:

Masters, W.A., B. Coulibaly, D. Sanogo, M. Sidibé, and A. Williams. 1996. The Economic Impact of Agricultural Research: A Practical Guide. Department of Agricultural Economics, Purdue University, West Lafayette, IN.

Available by e-mail: Masters@AgEcon.Purdue.edu

Description: Summarizes tools for conducting persuasive impact studies that quantify the economic benefits and costs of research. Three spreadsheet-based computer exercises and an accompanying diskette help apply the methods described.

Software:

Dream© - Dynamic Research Evaluation for Management

Description: Facilitates the application of the economic surplus model under a variety of market situations. Developed by the International Service for National Agricultural Research, this menu-driven computer program is available on the International Food Policy Research Institute Web site (see below under *Web Sites*). [For more information, contact Stanley Wood at s.wood@cgnet.com]

MODEXC—Modelo de Análisis de Excedentes Económicos

Helps calculate the NPV, IRR, and benefit–cost ratio for investments in agricultural research for both ex ante and ex post economic analysis of technical change under different market scenarios. This spreadsheet-based model is available in Spanish on the International Center for Tropical Agriculture Web site (see below under *Web Sites*).

Web Sites:

International Center for Tropical Agriculture: http://www.ciat.cgiar.org/projects/bar_bp1.htm

Features abstracts on the impact of agricultural research, databases, trends, and a download version of MODEXC (see above under *Software*).

International Food Policy Research Institute: http://www.cgiar.org/ifpri/dream.htm

Provides diverse relevant materials and a description and free download of the Dynamic Research Evaluation for Management program (see above under *Software*).

International Service for National Agricultural Research: http://www.cgiar.org/isnar/Fora/Priority/index.htm

Has a web page exclusively on information and discussion forum on priority setting in agricultural research. Describes the process, steps, and methods of priority setting for research.

Table 1: Framework for Impact Assessment in an ARO

	Indicators	Level	Frequency
	Applied research		
Research outcomes (for monitoring and evaluation)	Technology adoption rates, publications Basic and strategic research New research methods, tools, techniques, hypotheses supported or rejected, publications	Project level and program level with aggregation to institutes and system	On completion of a project (for example, after 3 years) Annually at the program level (for example, annual report) More in-depth at fixed intervals (for example, external reviews of programs)
	Applied research		
Research impacts (for evaluation)	Yield and production increases, cost reductions, economic return on research investment, impact on total factor productivity Basic and strategic research Use of new knowledge to increase the efficiency and efficacy of applied research	Program, institute, and system level	Once every five years for each program

Source: Byerlee and Alex (1998)

Figure 1: Effect of Interior Transport Costs on Tradeable Status in Evaluation of Research Benefits for Consumers